



**APPEL A PROJETS  
FRANCO-ALLEMAND EN SCIENCES HUMAINES ET SOCIALES  
EDITION 2012  
Ouvert aussi aux chercheurs post-doctorants**

Ouvert conjointement par l'Agence Nationale de la Recherche  
et la Deutsche Forschungsgemeinschaft

adresse de publication sur le site de l'ANR  
<http://www.agence-nationale-recherche.fr/FRALL-2012>

**Date d'ouverture  
17 janvier 2012**

**Date de clôture  
17 avril 2012**

Renseignements à l'ANR : [Marine.moguen-toursel@agencerecherche.fr](mailto:Marine.moguen-toursel@agencerecherche.fr)

Contact à la DFG : [achim.haag@dfg.de](mailto:achim.haag@dfg.de)

Appel à projets franco-allemand  
en sciences humaines et sociales

Ausschreibung eines deutsch-französischen  
Programms in den Geistes- und  
Sozialwissenschaften

Programme non-thématique

Ohne thematische Vorgaben

2012

2012

Titre du projet  
**L'évolution du système prépositionnel du  
français**

Projekttitlel  
**Die Entwicklung des französischen  
Präpositionalsystems**

PRESTO

PRESTO

Noms des deux coordinateurs  
(français et allemand) et institutions de  
rattachement

Namen der beiden Projektleiter  
(auf deutscher und auf französischer Seite)  
und ihrer Institutionen

## SOMMAIRE

### I. PRESENTATION DU PROJET

#### 1. FICHE D'IDENTITE DU PROJET

1.1.	Partenariat et participants .....	4
1.2.	Titre du projet et acronyme .....	5
1.3.	Discipline et domaine de recherche .....	5
1.4.	Durée du projet .....	5
1.5.	Résumé.....	5

#### 2. ETAT DE LA RECHERCHE, TRAVAUX ANTERIEURS

2.1.	Etat de la recherche	
2.1.1.	Les prépositions en français contemporain .....	6
2.1.2.	Les prépositions du français en diachronie .....	7
2.1.3.	La notion de figement .....	7
2.2.	Travaux antérieurs	
2.2.1.	Les travaux de l'équipe allemande .....	8
2.2.2.	Les travaux des équipes françaises .....	9

#### 3. OBJECTIFS ET PROGRAMME DES TRAVAUX

3.1.	Objectifs	
3.1.1.	Objectifs scientifiques .....	10
3.1.2.	Synergies attendues .....	13
3.1.3.	Diffusion des résultats .....	14
3.2.	Programme des travaux, méthodologie, calendrier	
3.2.1.	Tâches et Méthodologie :	
3.2.1.1.	Tâches et méthodologie de l'étude linguistique .....	15
3.2.1.2.	Tâches et méthodologie de la constitution et du traitement des corpus .....	16
3.2.2.	Calendrier et répartition des travaux entre les équipes	
3.2.2.1.	Programme de l'étude linguistique .....	18
3.2.2.2.	Programme et méthodologie de constitution, de traitement et de diffusion de ressources.....	19
3.3.	Bibliographie .....	20

#### 4. MOYENS DEMANDES

4.1.	Demande de moyens faite à la DFG pour la partie allemande .....	24
4.2.	Demande de moyens faite à l'ANR pour la partie française .....	24

#### 5. RECAPITULATIF DE LA DEMANDE FINANCIERE .....

25

### II. ANNEXES

Fiche des coordinateurs et des chercheurs membres du projet .....	28
Annexes allemande et française destinées à illustrer les outils, les méthodes & les objectifs de ce projet : cf. « documents annexes » au « document scientifique »	

## I. PRESENTATION DU PROJET

### 1. FICHE D'IDENTITE DU PROJET

#### 1.1. Partenariat et participants

I.1.1.1. Noms, rattachements, coordonnées et temps d'engagement des coordinateurs de projet, pour la partie allemande et pour la partie française

Nom	Prénom	Rattachements	Coordonnées	Temps d'engagement
BLUMENTHAL	Peter	Universität zu Köln	peter.blumenthal@uni-koeln.de	50%
VIGIER	Denis	Université Lumière Lyon2 & UMR ICAR 5191	Denis.Vigier@univ-lyon2.fr	50%

I.1.1.2. Noms, rattachements, coordonnées et temps d'engagement des principaux autres participants, pour la partie allemande et la partie française

#### Membres de l'équipe allemande

Nom	Prénom	Rattachements	Coordonnées	% Tps
DIWERSY	Sascha	Universität zu Köln	sascha.diwersy@uni-koeln.de	30%
GILLOV	Patrick	Universität zu Köln	pgililov@uni-koeln.de	50%

#### Membres de l'équipe française

NOM	Prénom	Rattachement		% Tps
CHAROLLES	Michel	Paris 3 & UMR LaTTICe	michel.charolles@ens.fr	10 %
DEMONET	Marie-Luce	Université F. Rabelais, Tours & UMR 6576 (CESR)	marie-luce.demonet@univ-tours.fr	10%
FAGARD	Benjamin	CNRS & UMR LaTTICe	benjamin.fagard@ens.fr	20 %
GARDELLE	Laure	ENS & UMR ICAR 5191	laure.gardelle@ens-lyon.fr	10%
GUILLOT	Céline	ENS & UMR ICAR 5191	celine.guillot@ens-lyon.fr	20%
HEIDEN	Serge	ENS & UMR ICAR 5191	slh@ens-lyon.fr	20%
LAVRENTEV	Alexei	CNRS & UMR ICAR 5191	alexei.lavrentev@ens-lyon.fr	20%
LEEMAN	Danielle	CNRS & UMR MODYCO 7114	danielle.leeman@u-paris10.fr	30%
PINCEMIN	Bénédicte	CNRS & UMR ICAR 5191	benedicte.pincemin@ens-lyon.fr	20%
SARDA	Laure	CNRS & UMR LaTTICe	laure.sarda@ens.fr	10 %

I.1.1.3. Autres contributeurs

Plusieurs contributeurs seront régulièrement sollicités, selon leurs champs d'expertise, pour participer aux réunions partielles ou générales du projet à Lyon, Paris ou Cologne. Leurs frais de déplacement ont été intégrés dans le budget.

#### Analyse des prépositions :

Dany Amiot, *Professeur des Universités, Université de Lille 3*

Jacques François, *Professeur des Universités émérite, Université de Caen*

Jürgen Lang, *Professeur des Universités émérite, Université d'Erlangen*

Walter de Mulder, *Professeur des Universités, Université d'Anvers*  
Christoph Schwartz, *Professeur des Universités émérite, Université de Constance*  
Achim Stein, *Professeur Université de Stuttgart, Institut für Linguistik Romanistik.*  
Dejan Stosic, *Maître de Conférences, Université d'Artois*

#### **Approche en diachronie :**

Bernard Combettes, *Professeur des Universités émérite, Université Nancy2*  
Christiane Marchello-Nizia, *Professeur des Universités émérite, ENS Lyon*  
Nathalie Fournier, *Professeur des Universités, Université Lyon 2*  
Sophie Prévost, *Chargée de Recherche CNRS, UMR LaTTICe*  
Sylviane Rémi-Giraud, *Professeur Emérite des Universités, Université Lyon2.*

#### **Etude du figement :**

Gaston Gross, *Professeur des Universités émérite, Université de Villetaneuse (Paris 13)*  
Béatrice Lamiroy, *Professeur des Universités, Université de KU Leuven*

#### **Lexicométrie & textométrie, traitement automatique du langage**

Dominique Labbé, *Professeur des Universités émérite, PACTE (UMR CNRS/IEP/UPMF/UJF 5194) Grenoble*  
Benoît Sagot, *chercheur INRIA, au sein du projet Alpage (INRIA Paris Rocquencourt et Université Paris 7), France. <http://alpage.inria.fr/~sagot>*  
Jean-Marie Viprey, *Professeur des Universités, Université de Franche-Comté*

#### **Humanités numériques**

Martin Holmes, *Ingénieur de recherche, University of Victoria Humanities Computing and Media Centre, Canada*  
F. Neveu, *Professeur des Universités, Université de Paris 4*  
R. Morrissey, *PhD in Romance Languages and Literatures, University of Chicago (directeur de l'ARTFL) <http://rll.uchicago.edu/faculty/morrissey>*  
M. Olsen *ARTFL / University of Chicago (directeur adjoint de l'ARTFL)*

### **1.2. Titre du projet : L'évolution du système prépositionnel du français** **Acronyme : PRESTO**

### **1.3. Discipline et domaine de recherche**

Linguistique française (synchronie et diachronie), statistique textuelle, textométrie, figement, modélisation du changement linguistique.

### **1.4. Durée du projet (24, 30 ou 36 mois)** 36 mois

### **1.5. Résumé**

Ce projet a pour but l'étude diachronique de l'emploi, des valeurs sémantiques et discursives des prépositions françaises *à, en, par, contre, dès, devant, entre, pour, sans, sur, sous, vers, dans*, de l'ancienne langue jusqu'au français contemporain.

Instrumentée, adossée à une approche statistique et distributionnelle, cette étude portera sur les variations du comportement combinatoire des prépositions suivant des critères de dates (évolution diachronique), de genres et d'auteurs. Notre attention se portera tout particulièrement sur la détection des combinaisons hautement spécifiques<sup>1</sup> dans lesquelles entrent ces prépositions. Nous étudierons l'émergence, la réorganisation et le déclin de structures prépositionnelles figées. Nous comptons parvenir à une vision innovatrice du devenir historique sur la base d'un modèle d'analyse fondé sur deux séries de démarches méthodologiques : (1) l'étude de l'interdépendance entre le quantitatif (les emplois) et le qualitatif (les valeurs) ; (2) l'étude des principes qui ont guidé, en diachronie, la recherche de plus-values sémantiques et discursives, véritable moteur du changement linguistique.

---

<sup>1</sup> Cooccurrences dont la fréquence élevée ne saurait être due au hasard et que les calculs probabilistes révèlent comme étant « spécifiques ». Parallèlement, une attention particulière sera portée aux combinaisons particulièrement sous-représentées pour une préposition donnée.

L'un des réquisits – mais aussi l'un des apports majeurs de ce projet - consistera en la constitution d'un corpus comportant des données étiquetées et lemmatisées, représentant toutes les périodes de l'histoire du français ainsi que différents genres discursifs et types de textes. Ce corpus, élaboré conjointement par les équipes française et allemande, s'appuiera pour partie sur les ressources déjà disponibles à Lyon (période médiévale), à Tours (XVI<sup>e</sup> s.), à Nancy (ATILF) et à Cologne (période moderne, XVII<sup>e</sup> s. – XX<sup>e</sup> s.). Une politique active sera menée pour permettre l'ouverture de la plus grande partie possible de ce corpus.

## 2.1. Etat de la recherche

### 2.1.1. Les prépositions en français contemporain

Les travaux publiés durant ces vingt dernières années sur les prépositions (pour une bibliographie complète, voir notamment Vagner 2006, 2007 ; de Mulder & Stosic 2009) ont contribué à en renouveler l'approche sur les plans syntaxique, sémantique et distributionnel.

**2.1.1.1. Sur le plan syntaxique**, un grand nombre des travaux récents (*inter alii*, Melis, 2003 ; Leeman, 2006, 2008) ont montré que les critères traditionnellement avancés pour définir la préposition (elle posséderait un régime de type nominal, obligatoire, unique, postposé, régime sur lequel elle peut exercer des contraintes de nature sémantique, syntaxique et distributionnelle) ne sont pas toujours vérifiés. Nous ne retiendrons ici qu'un point en rapport direct avec le projet : Le Pesant 2006, Leeman 2006, 2008 (*inter alii*) ont montré que la préposition est syntaxiquement définissable par sa capacité à sous-catégoriser ses arguments. Toutes les prépositions ne présentant pas les mêmes latitudes combinatoires, on peut les classer suivant le critère des restrictions de sélection qu'elles opèrent *a priori*<sup>2</sup> sur les catégories de leurs régimes, ce classement pouvant être présenté sous forme de tables syntaxiques (Vagner, *ibid*). L'un des objectifs de ce projet consistera à se doter, pour des tranches temporelles prédéfinies couvrant la période du XVI<sup>e</sup> s. au XX<sup>e</sup> s., de telles tables.

**2.1.1.2. a) Dans le domaine sémantique**, après le tournant cognitiviste<sup>3</sup> pris par les travaux sur les prépositions durant les années quatre-vingts (Vandeloise 1986, Lakoff 1987, Langacker 1987, ...), des approches relevant d'une linguistique plus spécifiquement instructionnelle (Anscombe 1991, Berthonneau & Cadiot 1993, Cadiot 1997, ...), constructiviste (Cadiot & Visetti 2001, Franckel 2002, Franckel & Paillard 2008) ou encore constructionnelle (Amiot & De Mulder (2012)) ont ouvert de nouvelles perspectives et suscité de nouveaux débats. Mais ces approches récentes souffrent en général d'un grave défaut. Les formulations (*instructions, motifs, formes schématiques, ...*) auxquelles elles aboutissent s'avèrent souvent peu contrôlables et/ou trop „puissantes“ (G. Kleiber, 1999 : 47-49).

**b)** Contrastant avec le caractère parfois très spéculatif de certaines de ces études, les travaux de D. Leeman (1995, 1998, 2008) ont montré l'impérieuse nécessité de prendre en compte, lorsqu'on cherche à établir l'identité sémantique d'une préposition (quel que soit le modèle théorique auquel on s'adosse), les contraintes distributionnelles que cette dernière impose en particulier à son régime. Ainsi montre-t-elle par ex., pour la préposition *en*, comment l'examen des possibilités et impossibilités observées pour la combinaison de la préposition avec ses arguments nominaux permet de construire progressivement une hypothèse d'invariant sémantique.

**c)** Il reste que la prise en compte des combinaisons prohibées *versus* autorisées pour une préposition donnée ne dit rien quant à la hiérarchie des *préférences* combinatoires qu'on observe dans les usages, et que seule une approche statistique de vastes corpus pourrait mettre au jour. Or ces préférences combinatoires (qu'elles soient « libres » ou contraintes) permettent de sonder les « *affinités* » profondes<sup>4</sup> existant entre la préposition et ses collocatifs, nous renseignant du même coup sur l'identité sémantique de celle-ci.

### 2.1.2. Les prépositions du français en diachronie

---

<sup>2</sup> *A priori* car, outre la préposition, la « tête externe » (Melis, *ibid*) du GP (verbe, adjectif, ...) peut aussi intervenir dans la restriction de sélection opérée sur le régime de la préposition.

<sup>3</sup> Sur ce point, voir E. Spang-Hanssen (1993: 20-24), C. Vagner (2006: 119-122)

<sup>4</sup> „Profondes“ car inaccessibles à la conscience des locuteurs.

2.1.2.1. Dans les études menées depuis le XIX<sup>e</sup> s. sur les prépositions du français, on peut distinguer deux orientations majeures : les études synchroniques, qui apportent des informations sur l'évolution de manière indirecte (c'est-à-dire par comparaison), et les études réellement diachroniques, qui s'interrogent principalement sur l'évolution des formes analysées. Depuis une vingtaine d'années, on trouve aussi de nombreuses études menées dans le cadre de la théorie de la grammaticalisation, impliquant à la fois une approche morpho-syntaxique et une étude du sens, à l'image des travaux de Kilroe (1989).

#### 2.1.2.2. Etudes synchroniques sur les périodes anciennes

Les études synchroniques sont les plus nombreuses. Elles portent sur différentes périodes du français – ancien, moyen, préclassique, classique..., et sur différentes caractéristiques des prépositions – en sémantique, syntaxe, morpho-syntaxe.

On notera tout d'abord l'existence d'études très fines sur l'emploi des prépositions, généralement basées sur un corpus restreint – chez tel ou tel auteur de langue ancienne. Ces études datent pour la plupart de la fin du XIX<sup>e</sup> s. ou du début du XX<sup>e</sup> s. (cf. Michael, 1904 ; Shears, 1922 ; Bekkers, 1931 ...) Elles constituent, malgré leur ancienneté, un point de départ inestimable pour les études modernes, en raison de leur précision philologique.

On trouve également des études sur le fonctionnement morphosyntaxique des prépositions, en particulier dans le cadre de l'approche générativiste, comme Lebel (2003). Ces études ont par exemple permis de mettre en évidence les phénomènes de recouvrement entre particules, préverbes et prépositions, par ailleurs décrits dans les grammaires : Buridant (2000) pour l'ancien français, Marchello-Nizia (1979/1997) pour le moyen français, Gougenheim (1974) pour le français classique. Ces grammaires proposent généralement une description relativement détaillée d'une petite sélection de prépositions, sans s'attarder sur les prépositions moins fréquentes ou moins grammaticalisées, ni sur les locutions prépositionnelles.

#### 2.1.2.3. Etudes diachroniques

Certains auteurs se sont intéressés à l'évolution proprement dite. Ainsi, Marchello-Nizia (2002) a montré la spécialisation progressive des morphèmes employés comme prépositions en ancien français, et qui perdent progressivement leur capacité à fonctionner comme particules. Le statut morpho-syntaxique des prépositions a aussi été étudié par Amiot & De Mulder (2002), qui s'interrogent sur la directionnalité de l'évolution adverbe/préfixe/préposition, en rapport avec la théorie de la grammaticalisation. Cette question de la grammaticalisation des prépositions est reprise plus généralement par Fagard (2006), qui précise les types de grammaticalisation observés pour les prépositions du français. La question plus spécifique du mode de création des prépositions complexes a été traitée notamment dans Fagard & De Mulder (2007). Shyldkrot & Kemmer (1995) étudient quant à eux l'évolution des emplois des prépositions fonctionnelles *à* et *de*, en rapport avec la transitivité. La transitivité des prépositions est également étudiée, d'un point de vue très différent, par Dufresne, Dupuis & Tremblay (2003). De nombreuses prépositions ont vu leur évolution sémantique décrite précisément : *à* (Kilroe 1989, 1994, Martineau 2000), *contre* (Dendale & De Mulder 1998, De Mulder & Vanderheyden 2001), *sur* (De Mulder & Vanderheyden 2002), *par* (Raithel 1875), *après* (Fagard 2004), *puis*, *depuis* (Fagard 2007), *devant* (Fagard & De Mulder 2010), *côté* (Fagard 2008), *dans* (Fagard & Sarda 2009).

On ne dispose cependant pas d'études statistiques menées sur de vastes corpus en vue de détecter les variations diachroniques de la combinatoire préférentielle des prépositions. Or de telles données permettraient d'étudier l'évolution diachronique de la phraséologie, thème sur lequel on dispose actuellement de très peu d'études. Elles ouvriraient aussi la voie à une réflexion innovante sur les « plus-values sémantiques » que ces évolutions permettent et plus largement sur les raisons qui président à l'émergence de ces plus-values (voir objectifs scientifiques, *infra*).

#### 2.1.3. La notion de figement

La notion de figement, complexe et donc délicate à définir, a donné lieu ces dernières années à de nombreuses publications dans la littérature francophone (*inter alii*, G. Gross 1996, Svensson 2004, Lecolle & Leroy 2006, Bolly 2008, Blumenthal & Mejri 2008, Lamiroy & al. 2010) et anglophone (*inter alii*, Everaert 1995, Hudson 1998, Moon 1998, Hunston & Francis

2000, Wray 2002). Au-delà des variétés d'approche et de conception, un accord règne autour du fait qu'il s'agit d'un phénomène *central* du langage (à la fois sur le plan qualitatif et quantitatif – cf. Dannell 1992, Senellart 1998) et *polyfactoriel*. D'aucuns insistent sur sa dimension psycholinguistique (*inscription mémorielle* : Grunig, 1997) et sociale (importance de *l'institutionnalisation* : Moon, 1998).

Parmi les critères mis en avant par les auteurs pour déterminer si une séquence est ou non figée (et à quel degré), trois jouissent aujourd'hui d'un large consensus dans la littérature sur le sujet : i) le critère sémantique de *non compositionnalité* du sens ii) le critère lexical de *non substituabilité paradigmatique* iii) le critère de *non modifiabilité* des marques morphosyntaxiques. À ces trois critères s'ajoute le caractère foncièrement *scalaire* du figement. D'où la difficulté bien connue à déterminer, face à des séquences non-prototypiques (i.e. qui ne réunissent pas l'ensemble des critères évoqués ci-dessus), si l'on a affaire ou non à une expression figée. Il conviendrait d'ailleurs de parler d'une *double* scalarité car elle se vérifie à la fois en synchronie et en diachronie. En effet, dans un état de langue donné, les propriétés du figement (ni nécessaires ni suffisantes (Klein & Lamiroy 2005)) se trouvent distribuées de manière très inégale selon les expressions considérées, d'où un effet de gradualité en synchronie. En second lieu, le phénomène de figement s'inscrit dans la temporalité : il s'agit d'un processus *diachronique* et *dynamique* (Mejri 2000), là aussi intrinsèquement scalaire. Ainsi observe-t-on que telle ou telle expression aujourd'hui figée était libre dans un état plus ancien de la langue (cf. Fournier, 1998).

Quels sont les mécanismes à l'origine de ce processus temporalisé qu'est le figement ? Parmi eux figurent sans aucun doute les *affinités combinatoires* entre constituants dont les récents travaux sur les collocations (*inter alii*, Butler 2003, Grossmann & Tutin, 2003, Mel'čuk 2003, Croft & Cruse 2004, Blumenthal & Hausmann 2006) ont montré le caractère central à la fois dans l'usage et dans l'évolution de la langue. Dans le présent projet, c'est avant tout une perspective diachronique et processuelle que nous adopterons sur le figement, et cela dans le cadre d'une approche d'abord quantitative et probabiliste. La méthodologie exposée par Blumenthal (2011, cf. *infra*, § 2.1.1.2.) pour les groupes prépositionnels constituera le point de départ de notre travail.

## 2.2. Travaux antérieurs

### 2.2.1. Les travaux de l'équipe allemande

Les travaux de l'équipe allemande portent, d'une part, de façon générale sur la méthodologie de l'analyse statistique et probabiliste des prépositions, de l'autre, sur des problèmes particuliers en diachronie. Ils s'appuient sur des ressources textuelles et logicielles importantes.

**2.2.1.1. P. Blumenthal** (2008) étudie les méthodes statistiques permettant d'évaluer le degré de similarité entre les combinatoires de deux ou plusieurs prépositions dans un corpus. En partant du profil combinatoire (calculs des accompagnateurs spécifiques) de chacune des prépositions à comparer, on parvient à déterminer de manière largement automatique le sous-ensemble conjoint des accompagnateurs ainsi que les sous-ensembles disjoints qui détaillent les éléments du voisinage caractéristiques d'une seule préposition. L'extension de ce procédé à l'ensemble des prépositions dans le cadre d'un calcul par échelonnement multidimensionnel procure une vision d'ensemble des affinités et des divergences distributionnelles entre prépositions et montre la réalité et les limites de leur quasi-synonymie dans des contextes particuliers.

**2.2.1.2.** Sous l'angle diachronique, Blumenthal (2011) applique cette méthode à deux corpus, l'un du XVII<sup>e</sup> s., l'autre de la deuxième moitié du XX<sup>e</sup> s. L'étude vise à appréhender les glissements fréquentiels à l'intérieur du groupe des prépositions les plus importantes, mais s'intéresse aussi à la redistribution partielle des rôles sémantiques et syntaxiques des prépositions. Elle montre en particulier comment, au cours de trois siècles, certains groupes prépositionnels figés disparaissent, alors que d'autres se forment ou changent de valeur sémantique, voire textuelle. Les causes dernières de ces évolutions se situent dans certains cas dans des domaines du système grammatical bien éloignés de celui de prépositions ; ainsi, la restructuration de l'actualisation du nom au XVII<sup>e</sup> s. n'est pas sans effet sur l'emploi des prépositions *en* et *dans*.



Dans le cadre de cette problématique diachronique, Blumenthal (2007) se penche sur une question bien particulière : la genèse et l'expansion de la combinaison *en* + début + espace de temps (par exemple *en début de semaine*) à partir des années 1920. Il s'agit de montrer ce que doit la fortune quantitativement remarquable de cette construction à des données extralinguistiques, spécialement à une nouvelle manière, notée aussi par les sociologues, de planifier le temps de travail.

2.2.1.3. Sur le plan technique, les travaux menés à Cologne s'appuient sur l'implémentation et l'entretien des archives de corpus de romanistique ainsi que sur le développement de diverses applications textométriques. L'ensemble des corpus disponibles est lemmatisé et annoté morphosyntaxiquement. Une partie des échantillons mis en place est également dotée d'annotations au niveau des relations de dépendance syntaxique. Les archives de corpus sont gérées *via* le système d'interrogation IMS Corpus Workbench (CQP/CWB, cf. <http://cwb.sourceforge.net/> ) et sont associés à plusieurs applications écrites en langage Perl et R, par lesquelles sont mis en place les procédés textométriques suivants:

- ✓ calcul de statistiques de co-occurrences à partir de concordances KWIC (de l'anglais: Keywords in Context), au moyen de différentes mesures d'associations (log-likelihood, mutual information et z-score)
- ✓ calcul de différences fréquentielles entre échantillons de corpus
- ✓ calcul de statistiques multivariées (échelonnement multidimensionnel, partitionnement hiérarchique, AFC) basées sur les données de différences fréquentielles et co-occurentielles
- ✓ caractérisation de la répartition interne des lexicogrammes au moyen des indices stéréotypiques

Une partie des procédés mentionnés sont présentés dans Blumenthal, Diwersy & Mielebacher (2005), Blumenthal (2006 ; 2008 ; 2010) ainsi que Diwersy (2007).

### 2.2.2 Les travaux des équipes françaises

L'équipe française regroupe des linguistes et des ingénieurs qui, suivant des configurations diverses, travaillent ensemble depuis plusieurs années sur des projets communs financés. Leurs domaines de spécialités respectifs sont précisés ci-dessous.

Syntaxe et sémantique des prépositions : **D. Vigier**<sup>5</sup>, après avoir soutenu sa thèse (2004) consacrée aux groupes prépositionnels (GP) ayant pour tête la préposition *en*, a continué ses travaux sur la syntaxe et la sémantique des GP en particulier lorsqu'ils occupent une position détachée (2003, 2005, 2008 a,b,c, 2012 a, b ; Charolles & Vigier, 2005 ; Le Draoulec & Vigier, 2009). **D. Lemann** s'intéresse depuis le milieu de la décennie 1980-1990 aux comportements de ce que l'on range classiquement sous le chef des « compléments d'objet indirects » ou des « compléments circonstanciels » : l'étiquetage grammatical traditionnel ne rend pas compte des propriétés effectives supposées justifier l'analyse. Celle-ci doit en fait d'abord s'appuyer sur l'observable et le manipulable, donc sur les formes matérialisées (lesquelles sont soumises aux tests distributionnels et syntaxiques bien connus : suppression, substitution, addition, emphase, interrogation, etc.), dont on tente ensuite une interprétation (Leeman 1999). **B. Fagard** a consacré sa thèse (2006) à l'étude de l'évolution sémantique des prépositions, et a continué depuis à travailler dans ce domaine, en privilégiant les études diachroniques sur corpus, dans le cadre de la théorie de la grammaticalisation. L. Gardelle a étudié la combinatoire de la préposition *of*, plus spécifiquement la concurrence entre *thereof* et *of it*.

Diachronie : Outre B. Fagard, **C. Guillot** et **A. Lavrentev** sont spécialistes de français médiéval et de diachronie du français. Ils développent conjointement la Base de Français Médiéval (<http://bfm.ens-lyon.fr/>) et le corpus de très ancien français CoRPTeF (<http://corptef.ens-lyon.fr/>). Ils participent également au projet ANR / DFG Syntactic Reference Corpus of Medieval French (<https://listes.cru.fr/wiki/srcmf/>). **M.-L. Demonet**, membre de l'IUF et spécialiste du XVI<sup>e</sup> s., a publié sur Rabelais, Montaigne, Pasquier (éditions critiques et électroniques, actes de colloques, monographies). Elle a créé deux sites Internet dont le dernier héberge des textes inédits et originaux de la Renaissance (<http://www.bvh.univ-tours.fr/Epistemon>) et est responsable du projet « Bibliothèques Virtuelles Humanistes ».

<sup>5</sup> <http://icar.univ-lyon2.fr/membres/dvigier/publications.htm>

Cadre de discours : L. Sarda et M. Charolles travaillent depuis plusieurs années sur les adverbiaux – dont les GP - dans une perspective discursive. M. Charolles a consacré de nombreuses publications (articles, directions de numéros) aux *adverbiaux cadratifs*, dont il a montré qu'ils constituaient une nouvelle famille de marques de cohésion discursive, au même titre que les connecteurs et les anaphores. Tous deux ont coordonné un projet ANR blanc (2006-2009, <http://www.lattice.cnrs.fr/Resume-Projet>) consacré aux cadres de discours de sens spatial (*Spatial Frames Adverbials*) et auquel ont participé B. Fagard et D. Vigier. Dans ce projet, L. Sarda a piloté une étude sur corpus des GP introduits par 9 prépositions. Cette étude visait à établir les corrélations entre leurs sens, leurs positions, leurs fonctions à la fois syntaxiques et discursives (Sarda & Charolles eds. à paraître). Par ailleurs, LS s'intéresse au statut argumental *vs* adjectif des GP et propose de saisir cette opposition à différents niveaux d'analyse (cf. Carlier & Sarda 2010).

Outils et Textométrie : S. Heiden et B. Pincemin sont fortement impliqués dans la recherche en textométrie. Ils pilotent le développement de la plateforme logicielle ouverte TXM, initiée dans le projet ANR Textométrie ANR-06-CORP-029 et diffusée librement en ligne (<http://sourceforge.net/projects/textometrie/>). Cette plateforme reprend et enrichit la tradition textométrique (concordances, spécificités, cooccurrences, analyse factorielle des correspondances, etc.) en l'adaptant aux corpus actuels (structurés, annotés). Elle intègre les standards internationaux (XML, TEI), et adopte une architecture modulaire qui lui permet de bénéficier de composants performants (moteur statistique R, moteur d'indexation et de recherche CQP). Elle est déjà le support de collaborations actives entre membres du présent projet, dans le cadre de la *Base de Français Médiéval* notamment.

### 3. OBJECTIFS ET PROGRAMME DES TRAVAUX

#### 3.1. Objectifs

**a) Améliorer nos connaissances sur la combinatoire des prépositions du français, de l'ancienne langue (IX<sup>e</sup> s) jusqu'au français contemporain.**

Pour la période du IX<sup>e</sup> s. au XX<sup>e</sup> s., on mettra au jour par des méthodes statistiques l'évolution en diachronie du « profil combinatoire » (Blumenthal 2008) de chacune des prépositions étudiées, notamment à partir de l'identification de ses accompagnateurs spécifiques<sup>6</sup>. On déterminera par une analyse multidimensionnelle (échelonnement multidimensionnel, analyse factorielle) les ressemblances contextuelles entre les diverses prépositions examinées. Grâce aux annotations introduites dans notre corpus, nous disposerons d'informations relatives à la (morpho)syntaxe [annotations à vérifier manuellement] ainsi qu'au genre discursifs des textes constitutifs du corpus. Ces informations constitueront autant de variables qui permettront de raffiner l'approche du profil combinatoire des prépositions. Les ANNEXES allemande et française jointes à ce projet sont destinées à illustrer la méthode et les outils mobilisés.

**b) Observer et problématiser l'émergence, la réorganisation et le déclin de structures figées**

Le recours aux méthodes statistiques devrait faire émerger des combinaisons significatives parmi lesquelles on trouvera des groupes prépositionnels plus ou moins figés. Un des objectifs du projet consistera à observer ces combinaisons en diachronie fine : émergence et devenir des figements, évolution des valeurs sémantiques des groupes concernés, changement dans leurs voisinages préférés au sein de la phrase, modifications de position, etc. On cherchera à mettre en relation les phénomènes observés avec des restructurations plus profondes du système de la langue, en lien avec la théorie de la grammaticalisation (*inter alii*, Heine, Claudi & Hünnemeyer (1991), Hopper & Traugott (éds) (1993), Heine & Kuteva 2002). On s'intéressera en particulier aux chaînes de grammaticalisation (*inter alii*, Heine, Claudi & Hünnemeyer (*ibid*), Traugott & König (1991)), à la question du figement et à la place du GP dans l'unité phrase.

---

<sup>6</sup> Noms et autres parties du discours les plus fréquents dans son environnement droit et/ou gauche par rapport à la fréquence absolue attendue.

**c) Établir des rapports entre les fluctuations fréquentielles de la combinatoire prépositionnelle, la recherche de plus-values sémantiques ou discursives et les besoins d'expression au sein d'une société**

Le programme esquissé au point précédent, qui rattache la description quantitative à l'analyse qualitative, doit être étendu au-delà de la dimension structurelle profonde ; dans le sillage de Coseriu (1974 : 246), nous nous interrogerons sur les besoins d'expression repérables dans la société d'une certaine époque qui peuvent être à l'origine des mouvements combinatoires et structurels thématiques ci-dessus en a). C'est dans ce contexte que la notion de « plus-value » que peut avoir une certaine combinaison prépositionnelle, surtout figée, prend tout son intérêt. Une argumentation de ce type a été esquissée dans Blumenthal 2012 à propos de quelques combinaisons de *en* (par exemple *en ville*).

**d) Etablir des rapports entre certaines restructurations observées dans le système de la langue et les préoccupations sociétales**

Une autre des originalités du projet sera la tentative de mettre en rapport l'évolution des structures figées de la langue, « faits sociaux » dans les termes de Durkheim, avec d'autres faits sociaux, qui, eux, correspondent à des préoccupations sociétales (cf. *supra* 2.2.1.2). Dans cette perspective, tout dans l'évolution grammaticale de la langue ne s'explique pas forcément par des principes inhérents au fonctionnement du langage, telle la grammaticalisation. Nous essaierons de dépasser l'opposition traditionnelle entre « histoire interne » et « histoire externe » de la langue en explorant une troisième voie, formée par des facteurs tributaires des deux pôles à la fois (cf. illustrations dans Blumenthal 2003, par exemple à propos de la combinaison *en début de semaine*, en lien avec un certain type d'organisation du travail dans les sociétés industrielles).

**e) Etendre le champ d'application de la notion de profil combinatoire (Blumenthal 2006) à des schémas co-occurentiels se manifestant dans des unités textuelles plus vastes.**

Charolles (1997) a montré que les GP adverbiaux – en particulier de sens spatial ou temporel - détachés en tête de phrase sont fréquemment utilisés dans les textes pour structurer les informations.

*[Dans les zones rurales, les personnes de plus de 25 ans sont avantagées par rapport aux jeunes, car elles entrent bien moins fréquemment en chômage]. [Dans les zones urbaines c'est l'inverse : les jeunes entrent relativement moins souvent en chômage ou en sortent plus facilement comme à Lille, Lens et Douai]. Par contre, les rythmes de sortie du chômage ne sont pas aussi différenciés ... (Les marchés locaux du travail dans le Nord-Pas-de-Calais, Economie et statistique, 182, p. 57).*

Le fonctionnement des adverbiaux dans les contextes de ce genre a fait l'objet (sur le français) d'un grand nombre d'études dans la période récente (*inter alii*, Charolles & Péry-Woodley (eds) 2009, Charolles et Prévost (eds) 2003, Vigier & Terran (eds) 2005) et d'un projet ANR Blanc 2006-2009 « Spatial Framing Adverbials » coordonné par l'UMR LaTTiCe. Ces études ont permis de montrer, à la suite de Thompson (1985) puis, notamment, de Virtanen (1992a et 1992b) que seuls les GP antéposés (en tête de phrase et *a fortiori* de paragraphe) jouissent d'un véritable potentiel cadratif, au contraire des GP insérés ou postposés.

Ce modèle de *l'encadrement du discours* apparaît compatible, sur le plan descriptif, avec le modèle de Hoey dans le cadre de la théorie de l'amorçage lexical (*Lexical Priming Theory*) - notamment pour ce qui concerne le niveau co-occurentiel que Hoey qualifie d'associations sémantiques textuelles (text semantic associations, Hoey (2005 : 122sq)).

Nous proposons donc d'élargir le champ d'application de la notion de *profil combinatoire* à des schémas co-occurentiels se manifestant dans des unités textuelles plus vastes. Selon Hoey, la collocation (au sens de Firth) représente un principe structurant du langage en ce sens que l'emploi d'un mot ou d'une unité syntagmatique est sous-tendu par des configurations combinatoires « amorcées » qui reposent sur un savoir expérientiel de chaque locuteur lié à des co-textes linguistiques récurrents dans le cadre de contextes discursifs spécifiques .

Pour la description de ces configurations, Hoey (cf. 2005 : 13) prévoit un modèle distinguant plusieurs types d'amorçage d'un mot ou d'un syntagme, qui concernent d'un côté la combinatoire lexicale et de l'autre, l'interaction avec les relations de cohérence dans le discours (« collocations textuelles » et « associations sémantico-discursives ») ainsi qu'avec des positions spécifiques dans un texte (« colligations discursives »).

Dans le projet soumis, nous nous proposons de répondre aux questions suivantes :

- ✓ Si un premier GP antéposé à potentiel cadratif fort apparaît en tête de phrase, d'autres expressions introductrices de cadres (GP, adverbes, GN, ...) tendent-elles également à apparaître dans cette même position dans la suite du texte ?
- ✓ Les séquences de cadres de discours sont-elles amorcées dans une phrase précédant immédiatement les segments repérés ? (ce qui confirmerait l'idée d'une planification à long terme)
- ✓ Quel rôle le critère du genre discursif et/ou du type de texte joue-t-il sur la fréquence d'apparition des cadres de discours, notamment employés en séquences ?

Notre enquête se limitera à la période allant du XVII<sup>e</sup> s. au XX<sup>e</sup> s., pour des raisons liées au nombre de textes disponibles et à l'évolution du système de la langue<sup>7</sup>.

#### **f) Constituer un corpus permettant une couverture satisfaisante du français écrit du IX<sup>e</sup> au XX<sup>e</sup> s.**

La constitution d'un corpus adéquat est une condition essentielle à la réalisation du projet. Il s'agit d'utiliser au maximum les ressources dont les équipes partenaires disposent déjà en les complétant et en les enrichissant des annotations linguistiques nécessaires.

Sur le plan chronologique, le corpus couvrira la période des premiers textes français (IX<sup>e</sup> s.) jusqu'à la fin du XX<sup>e</sup> s. La partie médiévale (IX<sup>e</sup>-XV<sup>e</sup> s.) sera constituée de textes de la *Base de Français Médiéval* (BFM) et fera aussi appel aux textes du *Nouveau Corpus d'Amsterdam* (NCA). La période du XVI<sup>e</sup> s. intégrera la base *Epistemon*<sup>8</sup> des *Bibliothèques Virtuelles Humanistes* et sera enrichie par les apports de textes de l'ATILF pour cette période. Pour les tranches temporelles ultérieures, on utilisera les textes de l'équipe de Cologne complétés par les apports d'autres bases : UVIC<sup>9</sup>, ARTFL<sup>10</sup>, la base FRANTEXT de l'ATILF et, le cas échéant, par des ressources textuelles librement accessibles en ligne (CNRTL, Gallica, Wikisource, Gutenberg). Pour ce qui est du nombre de mots, on construira *une* seule tranche de 500 000 mots pour les textes les plus anciens (IX<sup>e</sup> – XII<sup>e</sup> s.), qui sont limités en nombre. Pour la période XIII<sup>e</sup> – XV<sup>e</sup> s., on pourra disposer de tranches de 500 000 mots par siècle. Pour la période allant du XVI<sup>e</sup> s. au XX<sup>e</sup> s., chaque tranche de 50 ans sera représentée par 5 millions de mots environ.

Dans la mesure du possible, le corpus sera équilibré sur le plan des genres discursifs et des types textuels représentés. La typologie précise sera établie au début du projet sur la base des descripteurs textuels adoptés dans la BFM (domaine, genre, forme, etc.). Pour la période médiévale et pour des considérations d'équilibrage de corpus, les textes de taille supérieure à 50 000 mots pourront être échantillonnés. En complément des critères typologiques, les textes seront sélectionnés en fonction de la qualité de leur numérisation. On préférera les ressources libres de droits, mais il sera impossible d'exclure complètement les ressources privées (dont une partie importante de textes déjà exploités par les équipes partenaires).

Tous les textes seront balisés d'une façon plus ou moins « profonde » selon les recommandations de la TEI et enrichis linguistiquement par des outils de traitement automatique de la langue (TAL) : tokenisation, étiquetage morphosyntaxique et lemmatisation par le TreeTagger. Les textes pré-étiquetés manuellement (près de 30 textes, soit 500 000 mots

<sup>7</sup> Une telle étude, focalisée sur certains GP seulement (à fonction adverbiale, détachés en tête de phrase) et qui prend en compte les paramètres du genre discursif et du type de texte, nécessite en effet que nous disposions, pour chacune des tranches temporelles délimitées, de plusieurs centaines de textes et de plusieurs millions de mots. Par ailleurs, ce n'est que vers le XV<sup>e</sup> s. que le figement de l'ordre GV (après celui de VO) en phrase déclarative autorise des structures de type XSVO avec en tête (X) la possibilité de voir apparaître un circonstant thématissant, ce qui ne signifie pas que la position initiale ne pouvait pas être occupée par un GP au cours de la période antérieure. (C. Marchello, 1999 : 58 sq ; 2006 : 133).

<sup>8</sup> <http://www.bvh.univ-tours.fr/Epistemon/index.asp> . 44 textes annotés de cette base ont d'ores et déjà été importés sous TXM

<sup>9</sup> ([http://mariage.uvic.ca/toc.htm?sort=date\\_asc](http://mariage.uvic.ca/toc.htm?sort=date_asc))

<sup>10</sup> (<http://www.lib.uchicago.edu/efts/ARTFL/projects/BibBl/>)

pour la période médiévale) serviront de « golden corpus » pour l'apprentissage des outils de TAL.

### **g) Développer des lemmatiseurs du français pour toutes les périodes du corpus**

Le projet développera un ou des lemmatiseurs pour le français, nécessaire(s) aux différentes périodes du corpus. Pour cela, il utilisera les technologies de construction de modèles linguistiques par apprentissage sur corpus étiqueté-lemmatisés et lexiques morphosyntaxiques comme TreeTagger (Schmid, 1995) ou TnT (Brants, 2000)<sup>11</sup>. Ces modèles linguistiques seront optimisés pour la lemmatisation et l'étiquetage morphosyntaxique du corpus PRESTO et seront diffusés en ligne pour pouvoir être utilisés par d'autres projets de recherche, avec les ressources ayant permis de les construire pour que ces projets puissent également améliorer ou adapter les modèles linguistiques.

### **h) Mettre au point une plateforme logicielle unifiée d'analyse linguistique distributionnelle et combinatoire**

L'instrumentation visée se situe à la croisée d'une approche textométrique et de pratiques de linguistique de corpus. Elle mutualise les ressources et savoir-faire des équipes de Cologne et de Lyon.

Les deux équipes partagent déjà plusieurs choix technologiques importants : toutes deux recourent au moteur CQP pour l'indexation des corpus et à la bibliothèque R pour l'analyse statistique. Elles pratiquent les mêmes genres de calculs : concordances, cooccurrences, statistiques multivariées.

La collaboration permettra de consolider et d'unifier l'instrumentation :

- capacité à relever des profils cooccurentiels plus précis (prenant en compte des informations de dépendance syntaxique) ;
- meilleur contrôle interprétatif des cartographies issues d'analyses multidimensionnelles (retour au texte ciblé, choix des algorithmes, exploitation des indicateurs statistiques) ;
- intégration de tous les modules d'analyse à un environnement open-source et multiplateformes (windows, linux, macintosh).

Elle permettra aussi de concevoir et développer de nouvelles fonctionnalités adaptées aux problématiques spécifiques du projet (diachronie, figement).

Ces développements, mis en oeuvre dans le cadre d'une plateforme générale de textométrie diffusée gratuitement et sous licence open-source, bénéficieront à une large communauté d'utilisateurs.

Enfin, la version portail web de la plateforme TXM permettra de diffuser en ligne les corpus produits par le projet.

#### **3.1.2. Synergies attendues**

Ce projet constitue une nouvelle étape dans une collaboration initiée par la construction d'un numéro de revue (*Langue Française*<sup>12</sup>) consacré à la préposition *en*, numéro coordonné par D. Vigier et auquel participe P. Blumenthal.

La synergie visée par les deux équipes est étroitement liée à la complémentarité de leurs domaines de compétence propres et à leur convergence de fond sur les orientations théoriques, méthodologiques et techniques adoptées dans ce projet. L'un des domaines privilégiés de cette complémentarité sera l'étude de la structure des mêmes textes sous les angles des deux méthodes, dont l'une part de notre « cellule de base » (un syntagme prépositionnel avec son profil combinatoire), l'autre, d'une vision distributionnelle de l'ensemble du texte.

**3.1.2.1.** Dans le domaine linguistique, plusieurs membres des équipes de Lyon et de Paris travaillent depuis plusieurs années sur diverses prépositions françaises, en synchronie (Vigier, Sarda) et en diachronie (Fagard). Leurs études se sont systématiquement appuyées sur des corpus, mais sans recourir à des méthodes statistiques ni à des outils d'exploration aussi développés que ceux proposés dans le projet Presto. L'équipe de Cologne a elle aussi travaillé sur les prépositions (Blumenthal 2008 et *à par.*) mais de manière plus sporadique. Elle possède

<sup>11</sup> Helmut Schmid (1995): *Improvements in Part-of-Speech Tagging with an Application to German*. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland. Brants, Thorsten. In T - A Statistical Part-Of-Speech Tagger. 2000. Print.

<sup>12</sup> Projet en cours d'évaluation par le comité scientifique de *Langue Française* à la date de dépôt de ce projet.

en revanche une grande expérience de l'étude des « profils combinatoires » de mots pivots pouvant appartenir aux différentes parties du discours. Ce projet permettra donc aux équipes de mettre en commun leurs expertises respectives et complémentaires.

Par ailleurs, plusieurs membres des deux équipes françaises ont travaillé et publié sur les *cadres de discours* - dont Charolles qui a été à l'initiative des diverses recherches menées depuis dix ans sur cette question. Ce projet va permettre une collaboration à la fois théorique et pratique en vue de déterminer les points d'articulation entre les cadres d'analyse que proposent l'*amorçage sémantique* (*Lexical Priming*, Hoey, 2005) et les configurations combinatoires « amorcées » d'une part<sup>13</sup>, l'*encadrement du discours* (Charolles, 1997) d'autre part. Les groupes prépositionnels (fonctionnellement adverbiaux) détachés en tête de phrase et à portée cadrative étendue constitueront le point d'ancrage de ces échanges et de ces recherches communes.

**3.1.2.2.** Le travail que nous allons mener en commun sur le terrain des statistiques appliquées au corpus en vue d'observations linguistiques nous permettra de rapprocher et faire dialoguer très concrètement linguistique de corpus et textométrie : en quoi des démarches méthodologiques et des indicateurs pratiqués par la linguistique de corpus peuvent-ils enrichir une approche textométrique? et réciproquement, les spécificités de la textométrie (l'articulation constante entre quantitatif et qualitatif, ses procédures traditionnelles comme le calcul des spécificités, les segments répétés, etc.) montrent-elles un apport décisif dans une étude linguistique de corpus?

### **3.1.3. Diffusion des résultats**

#### **3.1.3.1. Ressources**

On distingue la diffusion des corpus de textes de celle des logiciels.

Pour les textes, on distingue :

- ✓ les corpus de textes annotés (lemmatisés, avec structures syntaxiques, etc.)
- ✓ les corpus d'apprentissage pour construire les outils de traitement automatique de la langue (TAL)
- ✓ les lexiques (forme graphique, description morpho-syntaxique, lemme, etc.) pour construire les outils de TAL

Pour les logiciels, on distingue :

- ✓ les logiciels de TAL et les ressources associées (fichiers paramètres, modèles linguistiques, etc.)
- ✓ les logiciels d'analyse de corpus
- ✓ les plateformes de mise en ligne et d'analyse de corpus (hébergées)

La politique de diffusion des corpus (a) sera la suivante<sup>14</sup>.

On distingue 2 modes de diffusion principaux : le téléchargement (corpus ouvert) et l'accès par plateforme ou portail (corpus restreint).

Le principe est de diffuser en téléchargement le maximum de corpus ouverts selon une licence ouverte à définir (de type Creative Commons<sup>15</sup>)<sup>16</sup>, sachant qu'étant données certaines contraintes possibles liées à des partenaires ayant fourni les sources de certains corpus, tout ne sera pas diffusable de façon ouverte. Pour les ressources dans ce dernier cas de figure, une diffusion restreinte par portail, contre inscription obligatoire après acceptation d'une charte d'accès, sera proposée – en s'appuyant sur la charte déjà utilisée pour la diffusion de la BFM et réalisée en association avec le service juridique de l'ENS et la délégation Rhône-Auvergne du CNRS (voir Annexe X). Au sein d'un tel portail, différentes restrictions pourront être appliquées à l'accès aux ressources selon les contraintes propres à chaque ressource. L'aide et le conseil de juristes spécialisés seront nécessaires (cf. budget) pour l'établissement de licences de diffusion, en lien avec le service des affaires juridiques de l'ENS Lyon.

---

<sup>13</sup> Cadre théorique auquel s'adosse nombre de travaux menés à Cologne

<sup>14</sup> Basé sur une description de la politique du projet de la Base de Français Médiéval actuelle.

<sup>15</sup> <http://creativecommons.org>

<sup>16</sup> Tout en garantissant bien sûr la propriété morale et patrimoniale des ressources à chaque producteur du projet.

Si on prend exemple sur l'état actuel de la BFM<sup>17</sup> qui couvre la période du IX<sup>ème</sup> siècle au XV<sup>ème</sup>. Dans le portail de diffusion :

- I 42 textes représentant 1,5 millions de mots sont accessibles pour téléchargement selon une licence Creative Commons de type BY-NC<sup>18</sup> et pour toute analyse de corpus réalisable avec l'aide des outils du portail ;
- II 25 textes représentant 0,9 millions de mots sont accessibles pour toute analyse réalisable avec l'aide des outils du portail (pas de téléchargement) ;
- III 18 textes représentant 0,6 millions de mots sont accessibles pour toute analyse réalisable avec l'aide des outils du portail avec une limitation permanente de la taille des contextes de concordances lors de recherches et l'impossibilité de lire l'édition d'un texte page par page (téléchargement proscrit).

La politique de diffusion des corpus d'apprentissage et lexiques (b) et (c) est de diffuser au maximum en téléchargement selon une licence ouverte à définir (de type Creative Commons). Les corpus d'apprentissage et lexiques ayant des restrictions d'accès (pour des raisons juridiques quelconques : propriété, etc.) ne seront pas diffusés par le projet PRESTO. Si on prend exemple sur le corpus d'apprentissage pour la catégorisation morpho-syntaxique CATTEX de l'ancien français et son lexique (forme, catégorie, lemme), ces ressources seront diffusées par téléchargement selon une licence ouverte de type Creative Commons ou LGPLLR<sup>19</sup>.

La politique de diffusion des logiciels de TAL et ressources associées produits ou améliorés par le projet (d) est de diffuser au maximum en téléchargement selon une licence ouverte à définir (de type GPL V3 ou LGPLLR selon le cas).

La politique de diffusion des logiciels d'analyse de corpus produits ou améliorés par le projet (e) est de diffuser au maximum en téléchargement selon une licence ouverte à définir (de type GPL V3).

Si on prend exemple sur le portail TXM WEB, utilisé pour diffuser des corpus en ligne ou les analyser sur un poste local, le logiciel sera diffusé gratuitement ainsi que ses sources selon une licence (GPL v3).

Enfin, les plateformes d'hébergement et de diffusion du corpus offriront un accès permanent au corpus mis en ligne par PRESTO dans la mesure des disponibilités des infrastructures d'hébergement mobilisés pendant la durée du projet (et au delà de sa terminaison dans la mesure des capacités des institutions tutelles des partenaires du projet).

### 3.1.3.2. Evénements et parutions

Une journée d'étude sera organisée en 2014 à Cologne. Elle ne donnera pas lieu à un appel à communication. Diverses équipes internationales travaillant sur des domaines proches ainsi que les « contributeurs » du projet seront sollicités pour y participer. Cette journée sera l'occasion pour les membres du projet de faire un premier point sur les résultats obtenus en les présentant à la communauté scientifique.

Un colloque international avec appel à communications sera organisé à Lyon en 2015, pour la clôture du projet.

Un ouvrage collectif sera élaboré, rendant compte du travail commun accompli dans le cadre de Presto

Les résultats de l'analyse linguistique seront mis à la disposition du public sur le site web du projet par l'intermédiaire d'une interface d'interrogation en accès libre ainsi que sous forme de sources SQL téléchargeables dotées d'une licence de type Creative Commons.

Les développements logiciels réalisés dans le cadre de la plateforme TXM seront diffusés en open-source avec la licence de logiciel libre GPL v.3. depuis le site de logiciels libres SourceForge<sup>20</sup>.

<sup>17</sup> Qui sera la partie principale du corpus PRESTO pour la période de l'ancien français.

<sup>18</sup> <http://creativecommons.org/licenses/by-nc/2.0/fr>

<sup>19</sup> <http://infolingu.univ-mlv.fr/DonneesLinguistiques/Lexiques-Grammaires/lgpllr.html>

<sup>20</sup> Une minorité d'implémentations d'algorithmes utilisés par le projet pourra faire l'objet d'une restriction de diffusion à l'extérieur du projet. Le choix des algorithmes concernés par ces restrictions fera l'objet d'une décision prise entre les partenaires du projet. Le mode de diffusion de ces implémentations sera alors de la responsabilité du partenaire demandant la restriction.

### 3.2. Programme des travaux, méthodologie, calendrier

Nos choix dans la programmation et la répartition des tâches ont été effectués en fonction des compétences respectives des collaborateurs : dans chaque équipe, l'un des chercheurs possède une double compétence en informatique et linguistique (S. Heiden à Lyon, S. Diwersy à Cologne), alors que les autres chercheurs sont soit linguistes, soit informaticiens.

#### 3.2.1. Tâches et Méthodologie

##### 3.2.1.1. Tâches et méthodologie de l'étude linguistique

**Tâche T1 : Étude pilote de l'évolution des valeurs des prépositions *en* et *dans* du XVI<sup>e</sup> au XXI<sup>e</sup> siècle, sur la base de corpus disponibles dès la première phase du projet**

Étape 1 : Étude du rapport des deux prépositions en français préclassique (profil combinatoire, échelonnement multidimensionnel, etc.)

Étape 2 : Leur situation en français classique

Étape 3 : L'emploi de *dans* et *en* dans *l'Encyclopédie* (seconde moitié du XVIII<sup>e</sup> s.)

Étape 4 : L'évolution jusqu'aux années 1880

Étape 5 : Nouvelles tendances depuis la fin du XIX<sup>e</sup> s.

Étape 6 : Exploitation des résultats sous les angles mentionnés ci-dessus (figement, défigement, grammaticalisation, lexicalisation, rapport avec l'évolution d'autres segments de la grammaire (comme le système des déterminants), rôle cadratif de certains GP (*en GN / dans GN*).)

**Tâche T2 : Application de la méthodologie développée au cours de la phase précédente à l'ancienne langue (des origines à la fin du XVI<sup>e</sup> s.). Élargissement de la perspective à une dizaine de prépositions constituant le noyau du système.**

Étape 1 : Des premiers documents au milieu du XII<sup>e</sup> s.

Étape 2 : Les prépositions dans l'œuvre de Chrétien de Troyes

Étape 3 : Les prépositions dans la prose narrative du Moyen Français

Étape 4 : Les prépositions dans les textes argumentatifs et expositifs du XVI<sup>e</sup> s.

Étape 5 : Élaboration d'une vision d'ensemble : le système prépositionnel face à la restructuration typologique entre l'ancien français et le moyen français et à la différenciation historique des genres textuels.

**Tâche T3 : Retour au français classique et moderne : l'évolution du système prépositionnel au cours des quatre derniers siècles.**

Étape 1 : Description et interprétation historique ; constitution d'environ huit diagrammes MDS (échelonnement multidimensionnel), deux par siècle, représentant les scores de similarité entre les emplois des prépositions dans nos corpus ; interprétation de ces coupes transversales dans une perspective diachronique.

Étape 2 : Analyse discursive, basée sur l'étude de l'amorçage entre mots et syntagmes, de la fonction argumentative et rhétorique de certains groupes prépositionnels, associés de façon statistiquement significative à des fonctions textuelles (entre autres, évocation d'une scène dans le cas des compléments cadratifs, prévisibilité d'une suite textuelle oppositive dans celui d'indications de domaines thématiques (par exemple *en politique*)).

Étape 3 : Élaboration d'un ouvrage collectif rendant compte du travail commun accompli dans le cadre de Presto.

##### 3.2.1.2. Tâches et méthodologie de la constitution, du traitement et de la diffusion de ressources

**Tâche C1 : Collecte et formatage**



Synthèse précise de l'existant, acquisition des compléments, afin d'équilibrer l'échantillonnage (siècles, genres).

Cette tâche comprend l'élaboration de scripts de transformation de sources diverses (dont les formats WPL et SGML utilisés à Cologne) vers un format compatible Unicode et XML-TEI basé sur le format XML-TXM.

### **Tâche C2 : Echanges et accord entre les membres de l'équipe sur le codage et les annotations**

Réflexion et prise de décision sur le codage des :

- ✓ étiquettes morphosyntaxiques (STTS, Multext, CATTEX...)
- ✓ lemmes (conventions de construction)
- ✓ positions des GP au sein de l'unité phrase
- ✓ fonctions syntaxiques des GP
- ✓ positions des GP au sein des unités textuelles
- ✓ genres et types textuels

### **Tâche C3 : segmentation lexicale, annotation morphosyntaxique et lemmatisation**

Le corpus comportera plusieurs couches d'annotation (automatique) :  
Etape 1 : Segmentation en phrases et tokenisation (forme graphique).  
Etape 2 : Etiquetage morphosyntaxique  
Etape 3 : Lemmatisation.

Ces tâches seront réalisées avec l'aide de scripts implémentant ou intégrant divers algorithmes d'étiquetage par apprentissage ou par application de règles, comme ceux développés dans OpenNLP, Unitex ou encore TreeTagger, et s'appuieront sur l'expérience des partenaires dans la construction ou l'intégration de tels outils.

### **Tâche C4 : Annotation syntaxique des corpus**

Le corpus comportera plusieurs couches d'annotation (automatique) :  
Etape 1 : Annotation des relations de dépendance syntaxique.  
Etape 2 : Annotation au niveau des groupes syntaxiques

Ces tâches seront réalisées avec l'aide de scripts implémentant ou intégrant divers algorithmes d'étiquetage par apprentissage ou règles d'annotations comme ceux développés dans les logiciels MATE-Tools ou LinguaStream.

### **Tâche C5 : segmentation textuelle (sections, paragraphes) et codage des positions**

En vue des calculs cooccurrentiels au niveau des relations transphrastiques, les unités textuelles identifiables (sections, paragraphes...) seront délimitées à l'aide de scripts et vérifiées semi-automatiquement à l'aide de vues d'éditions textuelles TXM. La relation entre unités de différents niveaux sera alors annotée automatiquement selon sa position : initiale, médiane ou finale.

#### **Remarque générale relative aux tâches C1 à C5 :**

- ✓ tous les traitements seront réalisés par des scripts (Groovy et XSLT) intégrés à la plateforme TXM
- ✓ l'annotation automatique peut donner des résultats variables en fonction des textes. Il conviendra de vérifier, à chaque étape, si des textes ont fait l'objet d'annotations problématiques, de qualité trop médiocre pour être retenues dans le corpus de travail final. Certains textes pourront alors faire l'objet d'un travail de correction pour parvenir à un niveau de qualité suffisant.

### **Tâche C6 : développement des outils pour l'exploration textométrique des corpus**

La plateforme TXM articule deux technologies clés déjà utilisées par les partenaires Cologne et Lyon :

- ✓ le moteur de recherche CWB-CQP qui gèrera efficacement le corpus annoté et structuré de 52 millions d'occurrences du projet
- ✓ l'environnement de calculs statistiques R qui implémentera et exécutera les modèles statistiques utilisés par le projet.

La tâche C6 consistera à développer cette plateforme pendant toute la durée du projet selon les sous-tâches suivantes :

1. Mise en commun des outils de Cologne et de Lyon : présentation et formation mutuelle. Intégration des outils de Cologne à la plateforme TXM (réécriture de scripts Perl en Groovy).

2. Corpus mutable : outils pour la correction assistée d'annotations en appui des tâches C3 à C5. Cette tâche demande un investissement important en conception logicielle et développement, car c'est toute l'architecture de TXM qui doit être revue pour rendre possible l'évolution dynamique des index. Mais le retour sur investissement est excellent pour l'efficacité des tâches d'annotation de corpus essentielles à ce projet (gain de temps et de cohérence dans les corrections manuelles).

3. Linguistique distributionnelle et profils cooccurentiels :

- extractions par requêtes syntaxiques (en lien avec la tâche de conception de l'annotation syntaxique)
- expression et extractions par zones<sup>21</sup> : possibilité de contextualiser de façon structurée une requête (Pincemin et al. 2006)
- matrices de cooccurrence d'ordre 1 (décompte des contextes de cooccurrence) et d'ordre 2 (association de mots sur la base de leur profil de cooccurrences)

4. Cartographies et interprétation

- pertinence et complémentarité éventuelle des algorithmes (MDS déjà utilisé à Cologne, AFC déjà utilisée à Lyon, etc.)
- aides à l'interprétation de visualisations de projections avec une forte réduction de l'espace des données

5. Figement

- implémentation de calculs textométriques : segments répétés, inventaire distributionnel, détection de motifs...

6. Calculs diachroniques

- implémentation dans TXM de calculs textométriques : spécificités chronologiques, AFC dynamiques...

### **Tâche C7: développement d'une base de données et d'une interface d'interrogation donnant accès aux résultats de l'analyse linguistique**

Pour donner une vue synthétique des résultats de l'analyse linguistique, ceux-ci seront enregistrés dans une base de données relationnelle (de type MySQL) qui sera rendue accessible sur le site web du projet par l'intermédiaire d'une interface d'interrogation. Ceci impliquera les sous-tâches suivantes:

- ✓ Elaboration d'un modèle entité-association selon la grille d'analyse orientant l'étude linguistique (cf. 3.2.1.1)
- ✓ Mise en place d'une base de données de type MySQL selon le modèle élaboré
- ✓ Création d'une interface permettant l'enregistrement des résultats issus de l'analyse linguistique
- ✓ Mise en relation de la base de données avec les outils d'exploration de corpus
- ✓ Création d'une interface d'interrogation accessible sur le site web du projet

---

<sup>21</sup> Analogue aux extractions tabulées avec l'aide de l'expression de labels dans les requêtes TIGERSearch.

- ✓ Création de sources SQL du contenu de la base téléchargeables

### 3.2.2. Calendrier et répartition des travaux entre les équipes

#### 3.2.2.1 Programme de l'étude linguistique

Mois	1 - 6		7 - 12		13 - 18		19 - 24		25 - 30		31 - 36	
T1 : Étude pilote (Mois : 1-15)	■	■	■	■	■							
T2 : Application de la méthodologie développée (Mois : 16-24)						■	■	■				
T3 : Retour au français classique et moderne (Mois 25-36)									■	■	■	■

Tableau 1 : Planification des tâches liées au traitement linguistique

Responsables : Denis Vigier, Benjamin Fagard et Peter Blumenthal

Personnels permanents impliqués dans la réalisation de ce volet du projet:

LYON: D. Vigier, C. Guillot, L. Gardelle, M.-L. Demonet, L. Sarda, B. Fagard, M. Charolles

COLOGNE : P. Blumenthal, S. Diwersy

Personnels non-permanents impliqués:

LYON : 1 postdoctorant

COLOGNE : P. Gililov

Le temps consacré à la réalisation de ce volet : 36 mois

Les livrables : cartographie des latitudes combinatoires des prépositions étudiées (par tranches temporelles) ; vision d'ensemble des besoins d'expression qui ont pu s'articuler à travers l'évolution des groupes prépositionnels (conformément au modèle présenté plus haut en 3.1.1. b) ; ouvrage collectif.

#### 3.2.2.2 Programme et méthodologie de constitution, de traitement et de diffusion de ressources

Mois	1	2	3	4	5-7	8-12	12-24	25	26	27-36
C1 : Collecte et formatage	■	■	■	■	■	■				
C2 : Principes de codage		■	■	■	■					
C3 : Segmentation, annotation morphosyntaxique et lemmatisation			■	■	■	■	■			
C4 : Annotation syntaxique				■	■	■	■	■		
C5 : segmentation textuelle et codage des positions					■	■	■	■	■	
C6 : Mise au point des outils pour l'exploration textométrique	■	■	■	■	■	■	■	■	■	■

des corpus										
C7: développement d'une base de données et d'une interface d'interrogation donnant accès aux résultats de l'analyse linguistique										

Tableau 2 : Planification des tâches liées à la constitution, au traitement et à la diffusion de ressources

Responsables : Serge Heiden et Sascha Diwersy

Personnels permanents impliqués dans la réalisation de ce volet du projet:

LYON: S. Heiden, B. Pincemin, A. Lavrentiev

COLOGNE : S. Diwersy

Personnels non-permanents impliqués:

LYON : Ingénieur Informatique pour les développements informatiques de la plateforme TXM

COLOGNE : 1 doctorant + 1 vacataire

Le temps consacré à la réalisation de ce volet : 36 mois

Les livrables : Corpus exploitables et outils textométriques

A. Corpus de textes français du IX au XX :

1. encodés en XML-TEI TXM
2. catalogués en genre discursif et en type
3. tokenisés, annotés en morphosyntaxe et lemmatisés
4. annotés syntaxiquement
5. annotés en positions pour les groupes prépositionnels

B. Outils textométriques

1. nouveaux modules d'importation de corpus
2. nouveaux modules de segmentation et d'annotation (morphosyntaxe, lemme, syntaxe et unités textuelles)
3. nouveaux modules de correction de l'annotation
4. nouveaux modules d'extraction pour les calculs statistiques distributionnels
5. nouveaux modules de calculs de cooccurrences
6. nouveaux modules de visualisation cartographique
7. nouveaux modules pour l'analyse du figement
8. nouveaux modules pour l'analyse diachronique

C. Base de données contenant les résultats de l'analyse linguistique

1. Sources SQL téléchargeables
2. Interface d'interrogation accessible sur le site web du projet

Les corpus seront livrés en plusieurs phases pour la réalisation des tâches d'analyses linguistiques.

Les nouveaux modules logiciels seront diffusés avec une licence open source

### 3.3. BIBLIOGRAPHIE

- Amiot, D. & W. De Mulder, (2002), De l'adverbe au préfixe en passant par la préposition : un phénomène de grammaticalisation ?, *Linguisticae Investigationes* XXV/2, 247-273.
- Bekkers, F. (1931), *Etude sur l'emploi que Froissart fait de la préposition*. Dissertation, Groningen.
- Blumenthal, P. (2002), Profil combinatoire des noms : synonymie distinctive et analyse contrastive, *Zeitschrift für französische Sprache und Literatur* 112 : 115-138.
- Blumenthal, P. (2003), Der Begriff der externen und internen Sprachgeschichte in der Romanistik, in *Romanische Sprachgeschichte* 1. Berlin :Teilband, 38-45.

- Blumenthal, P. (2006), *Wortprofil im Französischen* (Beihefte zur ZrP 332), Tübingen.
- Blumenthal, P. (2007), Histoire de la montée en puissance de *début*, in: L. Begioni/C. Muller (s.l.d.): *Problèmes de sémantique et de syntaxe - Hommage à André Rousseau*, Lille : Editions du Conseil Scientifique de l'Université Charles-de-Gaulle - Lille 3, (Travaux et recherches), 445-471.
- Blumenthal, P. (2008), Combinatoire des prépositions : approche quantitative, *Langue Française*, 157, 37-51.
- Blumenthal, P. (2009). "Les noms d'émotion : trois systèmes d'ordre", in: I. Novakova & A. Tutin (eds), *Le Lexique des émotions*, Grenoble : ELLUG 2009 (Langues, Gestes, Paroles), 41-64.
- Blumenthal, P. (2011), Le figement : du XVII<sup>e</sup> siècle au français contemporain, in: J.-C. Anscombe/S. Mejri (éds.): *Le figement linguistique: la parole entravée*, Paris: Éditions Champion 2011 (Lexica 18), 283-302.
- Blumenthal, P. (2012, à paraître) : Évolutions de la combinatoire prépositionnelle : le cas de *en*, *Mélanges Olivier Soutet*.
- Blumenthal P., S. Diwersy & J. Mielebacher (2005), Kombinatorische Profile und Profilkontraste: Berechnungsverfahren und Anwendungen, *Zeitschrift für Romanische Philologie* n°121, 1: 49-83.
- Blumenthal, P. & M.-J. Hausmann (2006), *Collocations, corpus, dictionnaires*, *Langue Française*, 150.
- Blumenthal, P. & S. Mejri (2008) (eds), *Les séquences figées : entre langue et discours*, *Zeitschrift für Französische Sprache und Literatur*, Beihefte 36, Franz Steiner Verlag : Stuttgart.
- Bonami, O. (1999), *Les constructions du verbe : le cas des groupes prépositionnels argumentaux*, Thèse de doctorat, Université Paris 7.
- Bolly, C. (2008), *Les unités phraséologiques : un phénomène linguistique complexe ?* Thèse de doctorat, Université Louvain-la-Neuve.
- Buridant, C. (2000), *Grammaire Nouvelle De L'ancien Français*. Paris : Sedes.
- Butler, C. (2003), *Structure and Function: A Guide to Three Major Structural-Functional Theorie*, Amsterdam, Benjamins.
- Cadiot, P. (1997a), Les paramètres de la notion de préposition incolore, *Faits de langues*, 9, 127-134
- Cadiot, P. (1997b), *Les prépositions abstraites en français*, Paris : Armand Colin.
- Carlier A. & Sarda L. (2010). Le complément de la localisation spatiale : entre argument et adjectif, In F. Neveu, V. Muni Toke, T. Klingler, J. Durand, L. Mondada & S. Prévost (eds.) *CMLF 2010 - 2ème Congrès Mondial de Linguistique Française*.
- Cervoni, J. (1991), *La préposition*, Paris-Louvain-la-Neuve, Duculot.
- Charolles, M. (1997), *L'encadrement du discours : univers, champs, domaines et espaces*, *Cahier de Recherche Linguistique, LANDISCO, URA-CNRS 1035 Université Nancy 2*, n° 6, 1-73.
- Charolles, M. & S. Prévost (2003)(eds.), *Adverbiaux et topiques. Travaux de linguistique* 47.
- Charolles, M. & M.-P. Péry-Woodley (2005)(eds.), *Les adverbiaux cadratifs. Langue française* 148.
- Charolles, M., Vigier, D. (2005), Les adverbiaux en position préverbale : portée cadrative et organisation des discours. *Langue Française*, 148, 9-30.
- Coseriu, E. (1974), *Synchronie, Diachronie und Geschichte. Das Problem des Sprachwandels*, München.
- Croft, W. & A. Cruse (2004), *Cognitive Linguistics*, Cambridge, CUP.
- Dannell, K. J., Nothing but phrases. About the distribution of idioms and stock phrases, in Edlund, L.E. & Persson, G. (éds). *Language: the time machine*, Umeå : Umeå University
- Dendale, P. & W. De Mulder, (1998), *Contre et sur : du spatial au métaphorique ou inversement ?*, *Verbum* XX/4, 405-434.
- De Mulder, W. & D. Stosic (2009) *Présentation*, in De Mulder, W. & Stosic, D. (2009) (eds) *Approches récentes de la préposition, Langages*, 173, 3-13.
- De Mulder, W. & A. Vanderheyden (2002), The evolution of French 'sur' : steps toward a diachronic prototype semantics of spatial prepositions, In Hubert Cuyckens (ed.), *Perspectives on prepositions*. Tübingen : Niemeyer, 177-193.
- Denis, P., Sagot B. (2009), « Coupling an annotated corpus and a morphosyntactic lexicon for state-of-the-art POS tagging with less human effort », in *Proceedings of The Pacific Asia Conference on Language, Information and Computation (PACLIC 23)*, Hong Kong, China.
- Dufresne, M., F. Dupuis & M. Tremblay (2003), Les prépositions dans l'histoire du français : transitivité, grammaticalisation et lexicalisation, *Verbum* XXV/4, 551-564.
- Everaert, M. et al.(eds). 1995. *Idioms*. Hillsdale, N.J.: Erlbaum.

- Fagard, B. (2004), *Après : de l'espace au temps, la sémantique en diachronie*. In A. Soares da Silva, A. Torres & M. Gonçalves (eds.), *Linguagem, cultura e cognição*, Braga : Almedina, 231-246.
- Fagard, B. (2006), La grammaticalisation en question – du latin aux langues romanes modernes, *Modèles linguistiques* XXVII-1 / 53, 91-110.
- Fagard, B. (2007), *De puis à depuis : préfixation et évolution sémantique*. In B. Combettes & C. Marchello-Nizia (eds.), *Etudes sur le changement linguistique du français*, 129-143.
- Fagard, B. (2008), 'Côté' dégrammaticalisation – le cas des prépositions. In B. Fagard, S. Prévost, B. Combettes & O. Bertrand (eds.), *Evolutions en français*, Berne : Peter Lang.
- Fagard, B. 2010. *Espace et grammaticalisation - L'évolution sémantique des prépositions dans les langues romanes*. Sarrebruck : Editions Universitaires Européennes.
- Fagard, B. & W. De Mulder (2007), La formation des prépositions complexes : grammaticalisation ou lexicalisation ?, *Langue française* 156, 9-29.
- Fagard, B. & L. Sarda (2009), Etude diachronique de la préposition *dans*, in J. François, É. Gilbert, C. Guimier & M. Krause (eds.), *Autour de la préposition*, Caen : Presses Universitaires de Caen, 225-236.
- Fournier, N. (1998), *Grammaire du français classique*, Paris : Belin
- Franckel J.J. & Paillard D. (2006), De la couleur des prépositions dans leurs emplois fonctionnels, *Modèles Linguistiques*, XXVII, 2, 54, 51-56.
- Franckel J.J. & Paillard D. (2008), *Grammaire des prépositions*, Paris : Ophrys.
- Gougenheim, G. (1974), *Grammaire de la langue française du 16e siècle*. Paris : Picard.
- Gross, G. (1996), *Les expressions figées en français. Noms composés et autres locutions*, Paris : Ophrys.
- Grossmann, F. & A. Tutin (2003) (eds), *Les collocations. Analyse et traitement*, Amsterdam : De Werelt.
- Grunig, B.-N. (1997), La locution comme défi aux théories linguistiques : une solution d'ordre mémoriel?, in Martins-Baltar, M. (éd.), *La locution entre langue et usages*, Paris Fontenay-St Cloud : ENS, 225–240.
- Heine B., U. Claudi & F. Hünnemeyer (1991), *Grammaticalization : A Conceptual Framework*. Chicago : University of Chicago Press.
- Heine, B. & T. Kuteva, (2002), *World Lexicon of Grammaticalization*. Cambridge, Cambridge University Press.
- Hopper P. & E. Traugott (éds) (1993), *Grammaticalization*. Cambridge University Press.
- Hoey, M. (2005), *Lexical Priming*, London, Routledge
- Homles, M. (2008) « Domestic strife in early modern Europe: images and texts in a virtual anthology ». (Conference presentation) with Claire Carlin; Digital Humanities 2008.
- Homles, M. (2010) « Using the Universal Similarity Metric to Map Correspondences between Witnesses ». (Conference presentation) Digital Humanities 2010. Conference, Kings College London..
- Hudson, J. (1998), *Perspectives on fixedness: applied and theoretical*, *Lund Studies in English*, 94. Lund: Lund University Press.
- Hunston, S. & G. Francis, (2000), *Pattern Grammar*, Amsterdam: Benjamins.
- Kilroe, P. (1989), *The grammaticalization of French à*. Doctoral dissertation, University of Texas : Austin.
- Kilroe, P. (1994), The grammaticalization of French à. In William Pagliuca (ed.), *Perspectives on Grammaticalization*, Amsterdam / Philadelphia : J. Benjamins, 49-61.
- Lamiroy, B., (2010), *Les expressions verbales figées de la francophonie. Les variétés de Belgique, de France, du Québec et de Suisse*. Paris : Ophrys.
- Lamiroy, B. & J. Klein (2010), Le problème central du figement est le semi-figement, in *Linx*, 53 (2005), 135–154.
- Lang, J. (1991), *Die französischen Präpositionen. Funktion und Bedeutung*, Heidelberg.
- Lavieu, B. (2006), Distinction entre groupes prépositionnels régis et non régis. *Modèles linguistiques* 53, tome XXVII-1 : 131-144.
- Lebel, M-É. (2003), *Le morphème par en ancien français*. Thèse en linguistique, Université du Québec à Montréal
- Lecolle, M. & S. Leroy (2006) (eds), *Changements linguistiques : figement, lexicalisation, grammaticalisation*, *Cahiers de praxématique*, 46 .
- Le Draoulec, A. & Vigier, D. (2009). *Dans la nuit je rentre chez moi. Interprétations temporelle et spatiale de quelques constructions dans + SNtps*. Actes des cinquièmes Rencontres de

- Sémantique et Pragmatique - « Espace, Temps dans les langues », *Revue de sémantique et de pragmatique*.
- Leeman, D. (1998), *Les circonstants en question(s)*, Paris, Kimé.
- Leeman, D. (1999), L'unité lexicale dans la perspective harrissienne, *LINX* 40 : 117-136.
- Leeman, D. (2006), la préposition française : caractérisation syntaxique de la catégorie, in Leeman, D. (2006) (éd), *la préposition en français*, I, *Modèles linguistiques* 54, tome XXVII, 7-17.
- Leeman, D. (2008), Prépositions du français, état des lieux, in Leeman, D. (2008) (éd), *Enigmatiques prépositions*, *Langue Française*, 157, 5-19.
- Le Pesant, D. (2006), Classification à partir des propriétés syntaxiques, *Modèles linguistiques* 53, tome XXVII-1 : 51-74.
- Marchello-Nizia, C. (1997 [1979]), *Histoire de la langue française aux 14<sup>ème</sup> et 15<sup>ème</sup> siècles*. Paris : Bordas.
- Marchello-Nizia, C. (1999), *Le Français en diachronie*, Paris : Ophrys.
- Marchello-Nizia, C. (2002), Prépositions françaises en diachronie : une catégorie en question, *Linguisticae Investigationes* XXV/2, 205-221.
- Martineau, F. (2000), Les prépositions à et de de l'ancien français au français classique. In Claude Buridant (dir.), *Actes du IXe colloque international sur le moyen français*. Strasbourg : Presses Universitaires de Strasbourg, 165-183.
- Marchello-Nizia, C. (2006), *Grammaticalisation et changement linguistique*. Bruxelles : Duculot / De Boeck
- Mejri, S. (2000), Figement lexical et renouvellement du lexique : quand le processus détermine la dynamique du système, *Le français moderne* LXVIII (1), 39-62.
- Melis, L. (2003), *La préposition en français*, Paris : Ophrys.
- Mel'čuk, I. (2003), Collocations : définition, rôle et utilité, In F.Grossman et A.Tutin (éds), 23-31.
- Michael, F. (1904), *Über den Gebrauch der Präpositionen mit Ausschluss von a, de, en, im altfranzösischen Rolandsliede*. Dissertation, Greifswald.
- Moon, R. (1998), *Fixed Expressions and Idioms in English, a Corpus-Based Approach*, Oxford : Clarendon
- Noailly, M. (2010), L'adjectif, quand il est invariable, sort-il pour autant de sa catégorie ?, in J. Goes & E. Moliné (eds), *L'adjectif hors de sa catégorie*, Actes du sixième colloque franco-roumain, Université d'Artois, 23-25 mai 2007, Arras.
- Raithel, G. (1875), *Über den Gebrauch und die begriffliche Entwicklung der altfranzösischen Präpositionen od, par, en*. Dissertation, Göttingen.
- Sagot B. (2010), « The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French », in *Proceedings of the 7th Language Resources and Evaluation Conference (LREC'10)*, Vallette, Malta.
- Sarda L. & Charolles M. (eds) (soumis) Les adverbiaux prépositionnels : position, fonction et portée. Numéro thématique, *Travaux de Linguistique*.
- Senellart, J. (1998), Reconnaissance automatique des entrées du lexique-grammaire des phrases figées, in Lamiroy, B. (éd), *Le lexique-grammaire*. *Travaux de Linguistique*, 37, numéro spécial. 109-127.
- Shears, F. (1922), *Recherches sur les prépositions dans la prose du moyen français (XIVe et XVe siècles)*. Paris : Champion.
- Shyldkrot, H. B.-Z. & S. Kemmer (1995), La grammaticalisation des prépositions : concurrence et substitutions, *Revue Romane* 30, 205-226.
- Stein, A. (2003), « Étiquetage morphologique et Lemmatisation de textes d'Ancien Français. » in: *Ancien et Moyen Français sur le Web: enjeux méthodologiques et Analyse du Discours*, P. Kunstmann & al. (eds), Ottawa: David Éditions, 273-284.
- Stein, A. and al. (eds) (2006), *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'Ancien Français (1150-1350)*, texte établi par A. Dees (Amsterdam 1987), remanié par A. Stein, P. Kunstmann & M.-D. Glessgen. Stuttgart: Institut für Linguistik/Romanistik.
- Svensson, M.H. (2004), Critères de figement. L'identification des expressions figées en français contemporain. Umeå : Umeå University.
- Thompson, S. (1985), Grammar and written discourse: Initial vs. final purpose clause in English, *Text* 5(1-2), 55-84.

- Traugott E.C. & König E. (1991), The Semantics-Pragmatics of Grammaticalization Revisited, in E.C. Traugott & B. Heine (eds), *Approaches to Grammaticalization*, vol. 1, Amsterdam/Philadelphia, John Benjamins, 189- 218.
- Vaguer, C. (2006), Bibliographie générale des prépositions du français (I et II). *Modèles linguistiques* 54, tome XXVII-II : 171-203.
- Vaguer, C. (2007), Bibliographie générale des prépositions du français (III). *Modèles linguistiques* 55, tome XVIII-I.
- Vaguer, C. (2008), Classement syntaxique des prépositions simples du français, *Langue Française*, 157, 20-36.
- Vigier, D. (2003), Les syntagmes prépositionnels en « en N » détachés en tête de phrase référant à des domaines d'activité, *Linguisticae Investigationes*, 26 (1), 97 - 122, Amsterdam, John Benjamins B.V.
- Vigier, D. (2004), *Les groupes prépositionnels en « En N » : de la phrase au discours*, Thèse de 3è cycle, Université Paris 3.
- Vigier, D. (2005). Les adverbiaux praxéologiques détachés en position initiale et leur portée. *Verbum*, XXVII, 2005, 3, 293-312.
- Vigier, D. (2008a). Les adverbiaux de localisation spéculaire : détachement et portée. In D. Apothéloz, B. Combettes, F. Neveu (eds) *Les linguistiques du détachement. Actes du colloque international de Nancy (7-9 juin 2006)*. Berne : P. Lang, 567-580.
- Vigier, D. (2008b), Contribution à une étude des constructions antéposées du type : *En homme intelligent et humain, il partagea tout de suite l'inquiétude de Marcel* (J. Verne), *Discours*, 2.
- Vigier D. (2008c). La gestion des cadres de discours dans une tâche rédactionnelle en FLE. In H. Hilton (ed). *Acquisition et didactique 1. Actes de l'atelier didactique, colloque international AFLS 2005*. Chambéry: Presses Universitaires de Chambéry, 113-129.
- Vigier, D. (à par., 2012a), « Comportements, déguisements, rôles, ... De quelques emplois intraprédicatifs de en », *Linguisticae Investigationes*, Amsterdam, Philadelphia : John Benjamins Publishing Co.
- Vigier, D. (à par., 2012b), « En attendant : un cas de pragmaticalisation », in *Les adverbiaux prépositionnels : position, fonction et portée*, Charolles, M. & Sarda, L. (éds.), *Travaux de linguistique*.
- Vigier, D. (ed.)(en prép.), *La préposition « en »*, *Langue Française*, paris : Hachette
- Vigier, D., Terran E. (eds) (2005). *Les adverbiaux cadratifs et l'organisation des textes*. *Verbum*, XXVII, 3.
- Virtanen, T. (1992a. ), *Discourse functions of adverbial placement in English*. Åbo: Åbo Akademi University Press.
- Virtanen, T. (1992b), Temporal Adverbials in Text Structuring: On Temporal Text Strategy. In A.-C. Lindberg, N. Enkvist & K. Wikberg (eds.), *Nordic Research on Text and Discourse*. Åbo: Åbo Academic Press, 85-197.
- Wray, A. (2002), *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

#### 4. MOYENS DEMANDES

##### Demande de moyens faite à la DFG pour la partie allemande

#### 4. 1 Basismodul

##### 4.1.1 Personalkosten

##### 4.1.1.1 Nach TV-L vergütetes Personal

##### a) Postdoktorand

1 halbe Stelle Postdoktorand/TVL-E 13 Stufe 4	36 Monate	89.550 €
---	-----------	----------

##### b) Doktorand

1 halbe Stelle Doktorand/TVL-E 13 Stufe 1	36 Monate	84.150 €
---	-----------	----------

##### 4.1.1.2 Hilfskräfte

zwei studentische Hilfskräfte/30 Std./Monat	36 Monate	52.000 €
---	-----------	----------



Die studentischen Hilfskräfte sollen bei der Zusammenstellung und Aufbereitung der Korpora (Lemmatisierung, PoS-Tagging, Parsing), beim Training von Software, bei der Literaturrecherche und -beschaffung, bei der Einrichtung der Datenbank und bei der verlegerischen Vorbereitung der Veröffentlichungen helfen.

Personalkosten Gesamt: 225.700 €

#### 4.1.2 Sachmittel

4.1.2.1 Geräte bis 10.000,- EUR, Software und Verbrauchsmaterial

4.1.2.2 Reisemittel

3 Reisen nach Lyon (3 Personen) à 750 € .p.P./Reise	6.750€
3 Reisen nach Lyon (1 Person) à 750 € .p.P./Reise	2.250€
3 x Teilnahme an Tagung Ausland (4 Personen) à 750 € p.P./Reise	9.000€
4 x Teilnahme an Tagung Inland (4 Personen) à 350 € p.P./Reise	5.600€

Reisekosten Gesamt 23.600 €

4.1.2.3 Mittel für wissenschaftliche Gäste

4.1.2.4 Mittel für Versuchstiere

4.1.2.5 Sonstige Mittel

4.1.2.6 Projektinterne Publikationsmittel

Sammelband bei John Benjamins, Amsterdam/Philadelphia, PA (pauschal)	3.000€
---	--------

#### 4.2 Modul Eigene Stelle

#### 4.3 Modul Vertretung

#### 4.4 Modul Mercator-Fellow

#### 4.5 Modul Projektspezifische Workshops

Durchführung eines eintägigen Workshops in Köln (2015)	16.000€
--	---------

#### 4.6 Modul Öffentlichkeitsarbeit

#### Demande de moyens faite à l'ANR pour la partie française

<b>4.7. Dépenses de personnels à recruter</b>	
1 ingénieur (CDD à temps plein) x 36 mois (Se reporter à la <b>fiche profil</b> en annexe)	126 000
1 postDoc (CDD temps plein) x 18 mois = (Se reporter à la <b>fiche profil</b> en annexe)	65 700
Vacations demandées : 30 heures =	658
<b>Total : 192 358 euros</b>	

#### 4.8. Equipements : Aucun

<b>4.9. Petits matériels, consommables, fonctionnement, etc.</b>	
renouvellement portables = 6 unités	9000
Logiciels coût de licence	1000
Achat ouvrages spécialisés liés à la problématique du projet	1000
Facturations internes :	
Reprographie, PEB, visio-conférences, ...	1000

**Total : 12 000 euros**

**4.10. Missions**

Programmation des réunions = 2 réunions générales par an (soit 6 sur tout le projet, alternativement en France et en Allemagne) + 4 réunions partielles par an (soit 12 réunions partielles dans tout le projet alternativement en France et en Allemagne, avec 3 membres allemands, 3 membres français).

Soit, pour le budget **côté français** :

3 réunions générales **de 2 jours** pour 10 personnes à Cologne x 700€ 21 000

Pour les 3 réunions générales à Lyon, prise en charge côté français du déplacement des 3 membres de Paris + 1 membre Tours= 4 x 3 x 120€ 1 440

6 réunions partielles **de 2 jours** pour 3 pers. à Cologne x 700€ 12 600

**Sous Total : 35 040 euros**

**Déplacements colloques**

1 colloques international de 3 jours par membres (10 x 1000 €) 10 000

Participation à la journée d'étude (Cologne) (10 x 700) 7 000

Invitation experts en cours de projet 4 000

Invitation conférenciers étrangers colloque final 4 000

**Sous Total : 25 000 euros**

**Total : 60 040 euros**

**4.11. Prestations de services**

Facturations externes :

Traduction en anglais de documents et d'articles en vue de publications internationales, construction d'un site Web destiné à valoriser le projet (information, accès à des documents de travail, pre-prints, ...), participation au financement de la publication des Actes du colloque, intervention de juristes pour la question des corpus, formation des participants à certains des outils automatiques utilisés dans le projet, traiteur pour colloque,...

**25 000 euros**

TOTAL 289 398 EUROS

FRAIS DE GESTION (4%) : **11 576 EUROS**

BUDGET GLOBAL : **300 974 EUROS**

**5. RECAPITULATIF DE LA DEMANDE FINANCIERE**

Reprendre ici les montants figurant dans les points 4.1 à 4.11

	FR		ALL
Dépenses de personnel	192 358	Basismodul: Personalkosten Modul Eigene Stelle Modul Vertretung Modul Mercator-Fellow	225.700
Equipements		Basismodul: Wissenschaftliche Geräte	
Frais de missions	60 040	Basismodul : Reisekosten	23.600

Autres dépenses <sup>22</sup>	12000	Basismodul : Sonstige Kosten	
-		Basismodul : Publikationskosten	3.000
Prestations de service	25000	-	
Frais de gestion	11 576	-	
		Modul Projektspezifische Workshops	16.000
		Modul Öffentlichkeitsarbeit	
<b>Total France</b>	<b>€ 300 974</b>	<b>Total Allemagne</b>	<b>€ 268.300</b>
<b>GRAND TOTAL € 569.274</b>			

**6. Autres moyens engagés pour la réalisation du projet**

Aucun

**7. Autres financements éventuels**

Aucun autre financement n'a été sollicité pour le présent projet auprès d'autres organismes de financement. Si tel était le cas, nous nous engageons à en informer sans délai l'ANR et la DFG.

**8. Signatures**



(Denis Vigier)



(Peter Blumenthal)

II. ANNEXES

<sup>22</sup> Regrouper sous cette rubrique les éventuelles « Dépenses sur facturation interne » figurant dans les fiches financières de l'ANR

## II.1. Fiches des coordinateurs et des chercheurs membres du projet

---

### BLUMENTHAL Peter

#### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : BLUMENTHAL

Prénom : Peter

Age : 66

Situation professionnelle actuelle : professeur des universités

Cursus :

1965-69 inscrit en *Romanistik* et linguistique générale aux universités de Bonn et de Cologne

1967/68 études à l'université d'Aix-en-Provence

1969 obtention de la maîtrise (Staatsexamen)

1970 doctorat en *Romanistik* obtenu à Bonn

1970-72 assistant associé à l'université de Toulouse-Le Mirail

1972-78 assistant à l'université de Cologne

1978 thèse d'État obtenue à Cologne

1983-88 titulaire de la chaire de linguistique romane à l'université d'Erlangen-Nuremberg

1988-98 titulaire de la chaire de linguistique romane à l'université de Stuttgart

1990 professeur invité à Aarhus (Danemark)

1991 doyen de la faculté des lettres à Stuttgart

1992 professeur invité à l'université Laval (Québec)

1992-2000 expert élu de la Deutsche Forschungsgemeinschaft (CNRS allemand) pour la linguistique romane

depuis 1998 titulaire de la chaire de linguistique romane à l'université de Cologne

2001 février-mars, professeur associé à Paris 7

2003 nomination en qualité d'expert étranger de l'Institut de linguistique française

2007 mars, professeur invité à Paris 13

2007 nomination en qualité d'expert des sciences humaines et sociales (DGRI-A6) par le Ministère de l'enseignement supérieur et de la recherche.

2010 Directeur de recherche associé au CNRS (3 mois)

#### B/ Autres expériences professionnelles

Directeur d'une revue linguistique et de diverses collections

Directeur de l'école des langues de l'université de Stuttgart (années 1990)

#### C/ Cinq publications en relation avec le projet Presto

- 1) *Wortprofil im Französischen* (Beihefte zur ZrP 332), Tübingen 2006.
- 2) "Histoire de la montée en puissance de *début*", in: L. Begioni/C. Muller (s.l.d.): *Problèmes de sémantique et de syntaxe - Hommage à André Rousseau*, Lille: Editions du Conseil Scientifique de l'Université Charles-de-Gaulle - Lille 3 2007 (Travaux et recherches), 445-471.
- 3) "Combinatoire des prépositions: approche quantitative", in: *Langue française* 157 (2008), 37-51.
- 4) "Histoires de mots : affinités (s)électives", in: J. Durand/B. Habert/B. Laks (éds.): *Congrès Mondial de Linguistique Française - CMLF'08*, Paris: Institut de Linguistique Française 2008.
- 5) "La sémantique cognitive face à l'idiomaticité", in: J. François (s.l.d.): *Grandes voies et chemins de traverse de la sémantique cognitive*, Leuven: Peeters 2010 (Mémoires de la Société de Linguistique de Paris, Nouvelle série, Tome XVIII), 147-162. <http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08338.pdf>

---

## VIGIER Denis

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : VIGIER

Prénom : Denis

Age : 51 ans

Situation professionnelle actuelle : Maître de conférences en langue et linguistique françaises (2007- ...), Université Lyon2 Lumière. Membre du laboratoire ICAR (<http://icar.univ-lyon2.fr/membres/dvigier/>)

### Cursus

2005-... : Maître de Conférences en langue et linguistique françaises à l'Université Lyon 2 Lumière

2004 : Thèse de Sciences du Langage, Université Paris 3 « *Les groupes prépositionnels en « en N » : de la phrase au discours* »

1999-2007 : Professeur Agrégé (PRAG) à l'Université Rennes 2 (Haute Bretagne)

1989-1999 : Professeur dans le secondaire

1989 : Agrégation de Lettres

### B/ Autres expériences professionnelles

Membre depuis 2011 du Comité de Pilotage du Laboratoire d'Excellence ASLAN (<http://imagine.universite-lyon.fr/labex/aslan-etudes-avancees-sur-la-complexite-du-langage-160635.kjsp>)

Membre de comités scientifiques pour divers colloques internationaux

Membre expert nommé par l'Aeres pour l'évaluation de l'EA Crisco (octobre 2010)

Co-rédacteur en chef de la revue discours (<http://discours.revues.org/>) depuis 2009, soutenue financièrement par l'INSHS depuis 2010

### C/ Cinq publications en relation avec le projet Presto

Vigier, D. (à par., 2012), « Comportements, déguisements, rôles, ... De quelques emplois intrapredicatifs de *en* », *Linguisticae Investigationes*, Amsterdam, Philadelphia : John Benjamins Publishing Co.

Vigier, D. (à par., 2012), « *En attendant* : un cas de pragmaticalisation », in *Les adverbiaux prépositionnels : position, fonction et portée*, Charolles, M. & Sarda, L. (éds.), *Travaux de linguistique*.

Vigier, D. (2008). Contribution à une étude des constructions antéposées du type : *En homme intelligent et humain, il partagea tout de suite l'inquiétude de Marcel*. (J. Verne) ». *Discours*, 2. (<http://discours.revues.org/index863.html>)

Vigier, D. (2005). Les adverbiaux praxéologiques détachés en position initiale et leur portée. In Vigier, D., Terran E. (eds), *Les adverbiaux cadratifs et l'organisation des textes*, *Verbum*, XXVII, 2005, 3, 293-312 )

Charolles, M., Vigier, D. (2005). Les adverbiaux en position préverbale : portée cadrative et organisation des discours. *Langue Française*, 148, 9-30.

---

## CHAROLLES Michel

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : CHAROLLES

Prénom : Michel

Age : 65 ans

Situation professionnelle actuelle : PRCE, université de Paris III

Cursus : Doctorat d'Etat en linguistique à l'Université de Besançon en 1982. Assistant à

l'Université de Rennes II de 1983 à 1987. Professeur à l'Université de Nancy II (1987-1997), puis à l'université de Paris III.

### **B/ Autres expériences professionnelles**

M. Charolles a participé à de nombreux projets sur la référence, les adverbiaux et la structuration des discours. Il a notamment coordonné un projet PIR Cognisciences (1991-1993) sur l'anaphore et un projet ANR Blanc sur les adverbiaux (2006-2009). A Paris III, il a présidé la Commission de Spécialistes de 7<sup>ème</sup> section (2007-2008) et dirigé l'Ecole Doctorale 268 « Langues et Langage » (2007-2008). Depuis 2008, il est directeur de l'UMR 8094 LATTICE (« Langues, Textes, Traitements Informatiques et Cognition ») dont l'établissement support est l'ENS Paris, avec Paris III comme rattachement secondaire.

### **C/ Cinq publications en relation avec le projet Presto**

- CHAROLLES M. & PREVOST S., eds, 2003. Adverbiaux et topiques. *Travaux de linguistique*, 47.
- CHAROLLES M. & PERY-WOODLEY M-P., eds., 2005, "Les adverbiaux cadratifs", *Langue Française*. 148
- CHAROLLES M., 2005, "Framing Adverbials and their Role in Discourse Cohesion: From Connection to Forward Labelling", in M.Aurnague, M.Bras, A Le Draoulec & L.Vieu, *SEM-05 Proceedings*, Biarritz, 13-30.
- CHAROLLES M., 2006, "Un jour (one day) in narratives", in I.Korzen & L.Lundquist (eds.), *Comparing Anaphors. Between Sentences, Texts and Languages*, Copenhagen, Copenhagen Studies in Language, 34, Copenhagen, Samfundslitteratur Press: 11-26.
- CHAROLLES M. & LAMIROY B., 2007, «Du lexique à la grammaire : seulement, simplement, uniquement», *Cahiers de lexicologie*, 90 : 93-116.

---

## **DEMONET Marie-Luce**

### **A/ Nom, prénom, âge, situation actuelle, cursus**

Nom : DEMONET

Prénom : Marie-Luce

Age : 60

Situation professionnelle actuelle : Professeur de littérature française, Université de Tours

Cursus :

Spécialiste des rapports entre littérature, langues et théories sémiotiques à la Renaissance (*Les Voix du signe*, 1992), M.-L. Demonet a enseigné dans les universités d'Aix-en-Provence, Clermont-Ferrand, Poitiers et Tours (2001). Elle a publié plusieurs études sur Rabelais et une transcription de ses œuvres en ligne (1995), parallèlement à ses travaux sur Montaigne (*Les Essais*, 1985, *Montaigne et le Nouveau Monde*, ed., 1992, *Montaigne et la question de l'homme*, éd. 1999, *A plaisir. Sémiotique et scepticisme chez Montaigne*, 2002, *L'écriture du scepticisme chez Montaigne*, co-éd. 2003). Elle étudie les transformations du vocabulaire intellectuel, juridique et social et les questions de théorie et d'esthétique littéraires : sur le statut de la fiction et sur la conception des signes à la Renaissance. Elle a été directrice du Centre d'Etudes Supérieures de la Renaissance de Tours (2003-2007), et elle est responsable du programme des « Bibliothèques Virtuelles Humanistes » (BVH) depuis 2003, pour la publication en ligne d'ouvrages du XVI<sup>e</sup> siècle et leur traitement par de nouveaux outils d'encodage et d'indexation (<http://www.bvh.univ-tours.fr>). En décembre 2010, les BVH ont obtenu une bourse Google pour les traitements linguistiques des bases en français de la Renaissance (régularisation et requêtes). M.-L. Demonet est membre de l'Institut Universitaire de France depuis 2005 (statut renouvelé en 2010).

### **B/ Cinq publications en relation avec le projet Presto**

1. 1995, *Electro-chroniques de Rabelais*, CD-ROM. Sélection des textes romanesques et parabelaisiens, et établissement des liens en hypertexte (avec E. Brunet), Paris,

éditions Les Temps qui Courent. Version en ligne (par Etienne Brunet) : <http://134.59.31.3/Rabelais.html>. En cours de migration sur le site des BVH.

2. 2004, « La Renaissance numérique: les Bibliothèques Virtuelles Humanistes », Euroconférence *Digital technology and philological disciplines*, éd. J.-L. Lebrave and A. Bozzi, Castelvechio Pascoli, *Linguistica computazionale*, XX-XXI, 2004, p. 175-185.
3. 2008, Nicole Dufournaud, avec la collaboration de Marie-Luce Demonet et Toshinori Uetani, *Manuel d'encodage XML-TEI, Renaissance et temps modernes*, sur [bvh.univ-tours.fr](http://www.bvh.univ-tours.fr). (révision et version 2 en juillet 2009) <http://www.bvh.univ-tours.fr/XML-TEI/index.asp>
4. 2008, Marie-Hélène Lay, Marie-Luce Demonet, « Some elements of reflection about sustainability and shareability of the Humanist Virtual Library (Bibliothèques Virtuelles Humanistes, BVH): a synthesis and experiment feed-back », communication acceptée à l'atelier « Sustainability of Language Resources and Tools for Natural Language Processing », *LREC Conference, 6th Language Resources and Evaluation Conference*, Marrakech. <http://www.lrec-conf.org/proceedings/lrec2008/>
5. 2009, communication au congrès des « Digital Humanities », Université du Maryland, « The 'Bibliothèques Virtuelles Humanistes' (Virtual Humanistic Libraries in Tours): a Collection, or a Corpus ? », [http://mith2.umd.edu/dh09/index.html%3Fpage\\_id=99.html](http://mith2.umd.edu/dh09/index.html%3Fpage_id=99.html).

---

## DIWERSY, Sascha

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : DIWERSY

Prénom : Sascha

Age : 41

Situation professionnelle actuelle : assistant de recherches / enseignant chercheur

Cursus :

1991-93 : inscrit en Linguistique Appliquée/Traductologie (français, russe) à l'Université du Pays de la Sarre, Sarrebruck

1994/95 : assistant-enseignant d'allemand au Lycée Kléber à Strasbourg  
1996-99 : inscrit en Linguistique Appliquée/Traductologie (français, russe) à l'Université du Pays de la Sarre, Sarrebruck

2000 : maîtrise obtenue à l'Université du Pays de la Sarre, Sarrebruck

2002- : assistant de recherches à l'Université de Cologne

2008 : doctorat en Romanistik obtenu à l'Université de Cologne

2009- : enseignant-chercheur au Centre de recherche interdisciplinaire sur la France et la Francophonie (CIFRA) de l'Université de Cologne

### B/ Autres expériences professionnelles

1999-2000 : Chargé de cours, Fremdspracheninstitut Saarbrücken

### C/ Cinq publications en relation avec le projet Presto

Diwersy S. (à par., 2012) "La francophonie multivariée. Ou : comment mesurer les français en Afrique ?", in: Blumenthal, P., Pfänder, S. (eds.): *Convergences, divergences et question de la norme en Afrique francophone* (= Le français en Afrique 27

Diwersy S., Evert S., Neumann S. (à par., 2012) "A semi-supervised multivariate approach to the study of language variation", in: Szmrecsanyi, B., Waelchli, B. (eds.) (2012): *Aggregating dialectology and typology: linguistic variation in text and speech, within and across languages*, Berlin e.a.: de Gruyter.

Diwersy S., Eufe R. (2011) "Die karolingische correctio und der lateinische Wortschatz – Kommunikationsverben in Heiligenviten als Fallbeispiel", in: Dessì Schmid, S., Hafner, J.,

- Heinemann, S. (eds.): *Koineisierung und Standardisierung in der Romania*, Heidelberg: Winter, p. 111-131.
- Diwersy S. (2007) *Kookurrenz, Kontrast, Profil: Korpusinduzierte Studien zur lexikalisch-syntaktischen Kombinatorik französischer Substantive (mit ergänzenden Betrachtungen zum Deutschen)*, thèse de doctorat non publiée, Université de Cologne, 2007).
- Diwersy S., Blumenthal P., Mielebacher J. (2005) "Kombinatorische Profile und Profilkontraste: Berechnungsverfahren und Anwendungen", in : *Zeitschrift für Romanische Philologie* 121, 1 49-83.
- 

## FAGARD Benjamin

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : FAGARD

Prénom : Benjamin

Age : 34 ans

Situation professionnelle actuelle : Chargé de Recherche au CNRS (laboratoire Lattice, CNRS & ENS)

Cursus : Thèse en linguistique théorique à l'Université Paris 7 et Université Rome 3 (co-tutelle) (2006). DEA de linguistique théorique à l'Université Paris 7 (2001). Agrégation de grammaire (2000). Etudes de lettres et linguistique des langues classiques (latin et grec) aux Universités de Paris 4 et Paris 10 (1996-2000).

### B/ Autres expériences professionnelles

Enseignement en tant que titulaire (moniteur-allocataire ou ATER, période 2001-2007) :

- Université Paris 3 (ancien français)
- Université Paris 7 (linguistique théorique, syntaxe, morphologie, diachronie, ancien français, littérature médiévale)
- Université de Valenciennes (ancien français, littérature médiévale, théorie littéraire, critique littéraire)

Enseignement en tant que chargé de cours (période 2007-2012) :

- Université de Rennes 2 (grammaire du 16<sup>ème</sup> et du 18<sup>ème</sup> pour l'agrégation)
- Université de Cayenne (ancien français pour le CAPES)
- Université Catholique de Louvain (la grammaticalisation, séminaire de Master de Liesbeth Degand)
- Université de Neuchâtel (grammaticalisation et langues romanes, séminaire de Master de Gilles Corminboeuf)
- Université Paris 3 (écrit et oral en linguistique française, séminaire de thèse de Florence Lefeuvre)
- Université Paris 7 (grammaticalisation et langues romanes, séminaire de thèse de Sophie Prévost)
- CIEP (linguistique et typologie)

### C/ Cinq publications en relation avec le projet Presto

Fagard, B. 2010. *Espace et grammaticalisation - L'évolution sémantique des prépositions dans les langues romanes*. Sarrebruck : Editions Universitaires Européennes.

Fagard, B. & De Mulder, W. 2010. *Devant* : Evolution sémantique d'une pré-position en français, *CMLF2010 Pro-ceedings*. DOI: 10.1051 / cmlf / 2010185

Fagard B. 2009. *Vers*, du français classique au français moderne ou contemporain : l'aboutissement d'une grammaticalisation ?, *Revue québécoise de linguistique* 33 / 1, 5-29.

Fagard, B., 2006. La grammaticalisation en question – du latin aux langues romanes modernes, *Modèles linguistiques* XXVII-1 / 53, 91-110.

Prévost, S. & Fagard, B., 2007. Grammaticalisation et lexicalisation : la formation d'expressions complexes, *Langue française* 156, 3-8.

---



## GARDELLE Laure

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : Gardelle

Prénom : Laure

Age : 32 ans

Situation professionnelle actuelle : Maître de conférences en linguistique anglaise – ENS Lyon.

Recherche : membre du laboratoire ICAR.

Cursus :

- 1996-1998 DEUG LLCE anglais, Université Lyon III, mention très bien
- 1998-1999 Licence LLCE anglais, Lyon III, mention très bien
- 1999-2000 - Maîtrise LLCE anglais, Lyon III, mention très bien  
- Lectrice en Angleterre (University of Leeds)
- 2000-2001 - Agrégation d'anglais option linguistique (reçue 3<sup>ème</sup>)  
- Capes d'anglais (reçue 5<sup>ème</sup>)
- 2001-2002 - DEA de linguistique, Lyon III : « Les animaux et le genre en anglais moderne : le système des pronoms personnels ». Mention très bien.  
- Stage d'agrégation (lycée Jean-Paul Sartre, Bron)
- 2002-2006 Thèse de linguistique, sous la direction du professeur P. Cotte : « Le genre en anglais moderne (XVI<sup>e</sup> siècle à nos jours) : le système des pronoms ». Soutenue en l'université Paris IV-Sorbonne le 1<sup>er</sup> décembre 2006. Mention très honorable avec félicitations du jury à l'unanimité. (1200 pages, dont 750 de rédaction ; corpus diachronique de 5200 occurrences)
- 2002-2007 Professeur agrégé d'anglais (PRAG), Université Lyon III
- 2007-2009 Maître de conférences en linguistique anglaise, Université Lyon III
- 2009- Maître de conférences en linguistique anglaise, ENS Lyon : linguistique et phonologie de l'anglais (licence, master 1, master 2, agrégation)

### C/ Cinq publications en relation avec le projet Presto

Gardelle, L. (2011) "Whoop her up, hit it, go it alone: the role of the personal pronoun in the fossilization process". Yearbook of Phraseology 2011, pp. 163-178.

---

## GUILLOT Céline

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : GUILLOT

Prénom : Céline

Age : 42 ans

Situation professionnelle actuelle : Maître de conférences à l'ENS de Lyon (7<sup>e</sup> section), membre junior de l'Institut universitaire de France

Cursus :

- septembre 2005 – février 2006 : détachement au CNRS sur un poste de CR2 dans l'UMR ICAR
- septembre 2004 – août 2005 : congé parental
- septembre 1999 – août 2004 / ATER à l'ENS Lettres et sciences humaines, UMR ICAR
- 1993 – 1994 : Scolarité à l'Ecole nationale du Patrimoine
- 1989 – 1992 : Scolarité à l'Ecole nationale des Chartes

### B/ Autres expériences professionnelles

- 1998-1999 : Conseillère pour les archives à la Direction Régionale des Affaires Culturelles de Rhône-Alpes
- 1994-1998 : Conservateur adjoint aux Archives Départementales de la Loire

## C/ Cinq publications en relation avec le projet Presto

(2004) « *Ceste parole et ceste aventure dans la Queste del Saint Graal, marques de structuration discursive et transitions narratives* », *L'Information grammaticale* 103, 29-36.

(2007) avec Heiden Serge et Lavrentiev Alexei. « Typologie des textes et des phénomènes linguistiques pour l'analyse du changement linguistique avec la Base de Français Médiéval », *Linx*, numéro spécial, 125 - 139.

(2007) avec Lavrentiev Alexei et Marchello-Nizia Christiane. « Les corpus de français médiéval : état des lieux et perspectives », *Revue française de linguistique appliquée* 121, 125-128.

(2008) avec Heiden Serge, Lavrentiev A. et Marchello-Nizia Christiane. « Constitution et exploitation des corpus d'ancien français et de moyen français », *Corpus* 7, 5-23.

(2012). « Le pronom anaphorique *cil* de l'ancien français : continuité ou discontinuité topicale ? ». In : Denizot C. & Dupraz E. (éd.). *Anaphore et anaphoriques : variété des langues, variété des emplois*, Mont-Saint-Aignan : Publications des universités de Rouen et du Havre (Cahiers de l'ERAC), 97-115.

---

## HEIDEN Serge

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : HEIDEN

Prénom : Serge

Age : 46

Situation professionnelle actuelle, depuis janvier 2005 : Ingénieur de Recherche en Informatique Linguistique à l'ENS de Lyon

Cursus : Docteur en informatique, Université Paris 6

1993-1994 Attaché d'Enseignement et de Recherche en Linguistique Informatique à l'ENS de Fontenay/Saint-Cloud, Fontenay-aux-roses

1994-2005 : Ingénieur d'Études en Informatique Linguistique à l'ENS de Fontenay/Saint-Cloud, Saint-Cloud ; puis ENS-LSH, Lyon

### B/ Autres expériences professionnelles

- ✓ encadrement du projet ANR 2007-2010 Textométrie : Fédération des recherches et développements en textométrie autour de la création d'une plateforme logicielle ouverte
- ✓ depuis janvier 2007 : co-responsable de l'équipe S3CoDi, une des 3 équipes de l'UMR5191 ICAR - <http://icar.univ-lyon2.fr/Equipe3/index.htm>
- ✓ depuis mai 2010 : membre élu du Comité national du CNRS section 34
- ✓ depuis mars 2010 : membre élu du conseil scientifique de l'ENS de Lyon
- ✓ organisation des JADT 2008 (Journées internationales d'Analyse statistique des Données Textuelles) - <http://jadt2008.ens-lyon.fr>

## C/ Cinq publications en relation avec le projet Presto

- ✓ **Heiden, S.** (2010). The TXM Platform : Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme. In Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation (PACLIC 24). 4-7 November 2010, Sendai.

- ✓ Guillot, C., Heiden, S., Lavrentiev, A. (2005). Typologie des textes et des phénomènes linguistiques pour l'analyse du changement linguistique avec la Base de Français Médiéval. LINX, vol.(spécial – 2ième semestre) : 125-139.
- ✓ Heiden, S. (2004). Interface hypertextuelle à un espace de cooccurrences : implémentation dans Weblex. In « Le poids des mots », Purnelle, G., Fairon, C. and Dister, A. editors, Proc. of JADT'04 (7th Journées internationales d'Analyse Statistique des données Textuelles), pp. 577-588.
- ✓ Heiden, S., Lavrentiev, A. (2004). Ressources électroniques pour l'étude des textes médiévaux : approches et outils. *Revue Française de Linguistique Appliquée*, vol.(9) : 91-102.
- ✓ Prévost S., Heiden S. (2002). Etiquetage d'un corpus hétérogène de français médiéval : enjeux et modalités. In Pusch C. D. and Raible W., editors, *Romanistische Korpuslinguistik : Korpora und gesprochene Sprache, Romance Corpus Linguistics : Corpora and Spoken Language*, pp. 127-136.

## LAVRENTIEV Alexei

### A/ Nom, prénom, âge, situation actuelle, cursus

Nom : Lavrentiev

Prénom : Alexei

Age : 39 ans

Situation professionnelle actuelle : Ingénieur de recherche CNRS, UMR ICAR

Cursus :

- 2009 : Doctorat en sciences du langage, Thèse *Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIIIe au XVe siècle*, ENS LSH, Lyon ;
- 1999 : Candidat ès sciences philologiques (équivalent du doctorat du 3e cycle), thèse *Analyse typologique de la catégorie des cas en russe*, Université d'État de Tomsk (Russie) ;
- 1994 : D.E.A. de linguistique française, Université Paris 3 ;
- 1993 : Diplôme d'études supérieures (équivalent maîtrise) en langue et littérature russe, Université d'état de Novossibirsk.

### B/ Autres expériences professionnelles

1998-2004 : Chercheur à l'Institut de philologie de la branche sibérienne de l'Académie des sciences de Russie ;

2001-2002 : Fulbright visiting fellow, Princeton University, participation au projet *Charrette* ;

2004-2008: Attaché temporaire d'enseignement de recherche, ENS LSH, Lyon, gestionnaire de la Base de Français Médiéval ;

2008-2009 : Ingénieur de recherche contractuel, projets ANR CoRPTeF et ANR-DFG SRCMF.

### C/ Cinq publications en relation avec le projet Presto

Lavrentiev, A (2011) « Syntactic reference corpus of medieval French », Zakharov, V. (éd.), *Proceedings of of the International Conference "Corpus Linguistics - 2011"*, St. Petersburg, 27-29 juin 2011, St. Petersburg, SPBU, p. 41-45.

Lavrentiev, A. (2010) « La 'phrase' en français médiéval : une réalité ou une reconstruction artificielle ? », Neveu, F., Muni Toke, V., Durand, J., Klinger, T., Mondada, L., Prévost, S. (éds), *Actes du 2e Congrès Mondial de Linguistique Française*, La Nouvelle Orléans, 12-15 juillet 2010, p. 277-289.

Guillot, C., S. Heiden, A. Lavrentiev & Ch. Marchello-Nizia (2008) « Constitution et exploitation des corpus d'ancien et de moyen français », *Corpus*, n° 7, p. 5 – 23.

Pincemin, B., C. Guillot, S. Heiden, A. Lavrentiev & Ch. Marchello-Nizia, (2008) « Usages linguistiques de la textométrie. Analyse qualitative de la consultation de la Base de Français Médiéval via le logiciel Weblex », *Syntaxe & Sémantique*, n° 9, p. 87 – 110.

Lavrentiev, A (2007) « Typologie textuelle pour l'étude linguistique de manuscrits français médiévaux », in Lavrentiev A. (dir.), *Systèmes graphiques de manuscrits médiévaux et*

### LEEMAN Danielle

#### **A/ Nom, prénom, âge, situation actuelle, cursus**

Leeman, Danielle, 64 ans, professeur des universités depuis 2001 à Paris-Nanterre. Enseignements de syntaxe, sémantique, linguistique générale aux divers niveaux de la Licence (Lettres classiques et modernes) et du Master professionnel de FLE.

Directeur de l'École doctorale 139 *Connaissance, Langage, Modélisation* entre novembre 2003 et février 2009 (évaluation « A » en 2008).

Vice Présidente déléguée à l'International entre 2009 et 2011, et à nouveau à partir de mars 2012 au sein de l'Équipe nouvelle issue des élections présidentielles de l'université en 2012.

#### **B/ Autres expériences professionnelles**

Direction des revues internationales *Langue française* et *Langages* jusqu'en décembre 2009.

Deux mandats (4 ans) successifs, comme PR, au CNU (démission en décembre 2009 pour raisons personnelles).

Expertises pour le FNRS belge et pour le FNRS suisse (en moyenne deux par an depuis une dizaine d'années).

Expertises pour l'ANR (une en 2009 et une en 2010).

Membre du Comité d'évaluation AERES pour une UMR (2010) et trois EA (2011).

Présidente du Comité d'évaluation AERES pour trois ED du PRES de Lyon (2010).

Nommée par le Ministère de l'Éducation nationale comme membre de la Commission de suivi des enseignements de grammaire pour l'École élémentaire et le collège (novembre 2011-...).

#### **C/ Cinq publications en relation avec le projet Presto**

1. Leeman, D. dir. (2007) *Modèles linguistiques* 55 : *De la préposition à la locution prépositive.*
2. Leeman, D. dir. (2008) *L'Information grammaticale* 117 : *Locutions du français, traduction et traduction automatique.*
3. Leeman, D. dir. (2008) *Langue française* 157 : *Enigmatiques prépositions.*
4. Leeman, D. dir. (2010) *Le Français moderne* LXXVIII-1: *Sur sous et sur et leurs locutions.*
5. Leeman, D. (2011) « Complémentation verbale et préposition », in D. Van Raemdonck (dir.) *Modèles syntaxiques*, Berne, Peter Lang, coll. « Gramm'r » : 133-148.

---

### PINCEMIN Bénédicte

#### **A/ Nom, prénom, âge, situation actuelle, cursus**

Nom : Pincemin

Prénom : Bénédicte

Age : 42 ans

Situation professionnelle actuelle : Chargée de recherche CNRS

Cursus :

1993 : Ingénieur de l'École Centrale Paris.

1999 : Thèse de Doctorat en Linguistique, Paris IV Sorbonne, dir. François Rastier, dans le cadre d'un contrat avec EDF. Soutenue le 6 avril 1999, mention très honorable et félicitations.

Intitulé : Diffusion ciblée automatique d'informations : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents.

2000 : Admission au concours de recrutement de chargés de recherche 2e classe CNRS, section 34 (Représentations – langages – communications). 2004 : promotion au grade de chargée de recherche 1ère classe. 2000-2005 : affectation au LLI (Paris 13 Villetaneuse) ; depuis 2006 : affectation au laboratoire ICAR (ENS de Lyon).

#### **B/ Autres expériences professionnelles**

Gestionnaire de la liste électronique Sémantique des Textes (depuis 1995) (plus de 350 abonnés dans une vingtaine de pays) : <https://listes.cru.fr/sympa/info/sdt>

Co-éditeur du corpus XML-TEI Droits de l'Homme (2005, 2006) <http://www-ldi.univ-paris13.fr/>

Projet ANR Fédération des recherches et développements en textométrie autour de la création d'une plateforme logicielle ouverte (ANR-06-CORP-029) : <http://textometrie.ens-lyon.fr/> :  
Coordination de l'équipe Conception et Application.

Membre du comité scientifique + formateur à l'Ecole thématique CNRS MISAT (Méthodes Méthodes Informatiques et Statistiques en Analyse de Textes) (annuelle depuis 2009).

Membre du comité scientifique des Journées internationales d'Analyse statistique des Données Textuelles, co-organisatrice avec Serge Heiden de l'édition 2008 de la conférence à Lyon.

### **C/ Cinq publications en relation avec le projet Presto**

Heiden Serge, Magué Jean-Philippe, Pincemin Bénédicte (2010) - « TXM : Une plateforme logicielle open-source pour la textométrie -conception et développement », in Sergio Bolasco, Isabella Chiari, Luca Giuliano (eds), *Statistical Analysis of Textual Data -Proceedings of 10th International Conference JADT 2010*, Edizioni Universitarie di Lettere Economia Diritto, Rome, 9-11 juin 2010. [En ligne sur le site LED electronic archive : <http://www.ledonline.it>]

Pincemin Bénédicte, Heiden Serge, Lay Marie-Hélène, Leblanc Jean-Marc, Viprey Jean-Marie (2010) - « Fonctionnalités textométriques : Proposition de typologie selon un point de vue utilisateur », in Sergio Bolasco, Isabella Chiari, Luca Giuliano (eds), *Statistical Analysis of Textual Data -Proceedings of 10th International Conference JADT 2010*, Edizioni Universitarie di Lettere Economia Diritto, Rome, 9-11 juin 2010. [En ligne sur le site LED electronic archive : <http://www.ledonline.it>]

Pincemin Bénédicte, Guillot Céline, Heiden Serge, Lavrentiev Alexei, Marchello-Nizia Christiane (2008) - « Usages linguistiques de la textométrie. Analyse qualitative de la consultation de la Base de Français Médiéval via le logiciel Weblex. », Mathieu Valette (éd.), *Textes, documents numériques, corpus. Pour une science des textes instrumentée, Syntaxe & Sémantique*, 9, pp. 87-110.

Pincemin Bénédicte, Issac Fabrice, Chanove Marc, Mathieu-Colas Michel (2006) – « Concordanciers : thème et variations », *Actes des 8es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Jean-Marie Viprey et al. (éds), Besançon : Presses Universitaires de Franche-Comté, ISBN 2.84867130.0, vol. II, pp. 773-784.

[En ligne sur le site Lexicometrica : <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2006/PDF/II-069.pdf>]

Pincemin Bénédicte (2004) - « Lexicométrie sur corpus étiquetés », *Actes des 7es Journées internationales d'analyse statistique des données textuelles (JADT 2004)*, Gérald Purnelle & al. (éds), Presse universitaires de Louvain, Louvain-la-Neuve (Belgium), 10-12 mars 2004, vol. II, pp. 865-873.

[En ligne sur le site Lexicometrica :

[http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT\\_084.pdf](http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_084.pdf)]

---

## **SARDA Laure**

### **A/ Nom, prénom, âge, situation actuelle, cursus**

Nom : Sarda

Prénom : Laure

Age : 40

Situation professionnelle actuelle : Chargée de recherche au CNRS (1<sup>er</sup> classe)

Cursus :

2007-2010 Co-rédactrice en chef de la revue discours (<http://discours.revues.org/>)

2003 Mutation dans le laboratoire Lattice (CNRS/ENS, Paris)

2000 Séjour de recherché à l'université d'Austin, USA. (Collaboration avec Knud Lambrecht).

1999- Chargée de recherché au CNRS (à l'ERSS, UMR 5610, Toulouse).

1999 Thèse de doctorat en linguistique générale et théorique (avec les félicitations)

Université Toulouse le Mirail

- 1994 DEA en linguistique générale et théorique (avec les félicitations)  
 Université Toulouse le Mirail
- 1993 Maîtrise en linguistique générale et théorique (avec les félicitations)  
 Université Toulouse le Mirail
- 1992 Licence en linguistique générale et théorique  
 Université Toulouse le Mirail
- 1991 DEUG en Lettres Modernes  
 Université de Perpignan

### **B/ Autres expériences professionnelles**

Enseignement de syntaxe et sémantique en deug et licence de Sciences du Langage à l'université de Toulouse-le-Mirail, entre 1994 et 2002.

### **C/ Cinq publications en relation avec le projet Presto**

- Sarda L. & Charolles M. (eds) (soumis) Les adverbiaux prépositionnels : position, fonction et portée. Numéro thématique, *Travaux de Linguistique*.
- Carter-Thomas, S. Sarda, L. (2011). Identifying the semantic roles of postverbal prepositional phrases: the participant / circumstance interface, *La linguistique systématique fonctionnelle et la langue française* (Ed D. Banks).
- Carlier A. & Sarda L. (2010). Le complément de la localisation spatiale : entre argument et adjoint, In Franck Neveu, Valelia Muni Toke, Thomas Klingler, Jacques Durand, Lorenz Mondada et Sophie Prévost (Eds.) *CMLF 2010 - 2ème Congrès Mondial de Linguistique Française*, Publié en ligne : 12 Juillet 2010, CMLF 2010, 138.
- Sarda Laure (2010). "Les adverbiaux prépositionnels en *dans* : exploration en corpus de la notion de contenance ". *Corela*, Numéros spéciaux, Espace, Préposition, Cognition-Hommage à Claude Vandeloise. <http://edel.univ-poitiers.fr/corela/document.php?id=2645>
- Sarda, L., Carter-Thomas, S. (2010). Rôle informationnel et textuel des adverbiaux détachés : exemples anglais et français en *sur* et *on*. In *Directions actuelles en linguistique du texte*. L. Florea, C. Papahagi, L. Pop, A. Curea (eds) Actes du colloque international *Le texte : modèles, méthodes, perspectives* Casa Cărții de Știință, Cluj-Napoca : 255-268.