

*Viprey Jean-Marie*  
*Né le 22 Octobre 1952*  
*Professeur des Universités, 1<sup>ère</sup> Classe*  
*Université de Franche-Comté*  
*Section n° 9 du CNU*

## 0. Déroulement de carrière, diplômes et concours

**2002 Habilité à Diriger les Recherches** (sous la direction du Pr. Claude Condé). Jury : Président : Pr. Georges Molinié (Paris-IV Sorbonne) - Membres : Pr. Jean-Michel Adam (Lausanne), Pr. Claude Blum (Paris-IV Sorbonne), Pr. Claude Condé (Besançon), Pr. André Salem (Paris-III Sorbonne Nouvelle), Pr. Max Silberstein (Besançon)

**Qualifié aux fonctions de Professeur des Universités (7<sup>ème</sup> et 9<sup>ème</sup> sections) en 2003**

**1996 Doctorat de 3<sup>ème</sup> cycle** sous la direction du Pr Jean Peytard (Linguistique, Sémiotique, Didactique) Jury : Président : Pr. Jean-Philippe Massonnie (Besançon) - Membres : Pr. Jean-Michel Adam (Lausanne), Pr. Etienne Brunet (Nice), Pr. Claude Condé (Besançon), Pr. Pierre Malandain (Lille-III), Pr. Jean Peytard (Besançon)

1990 Agrégation (Externe) de Lettres Modernes - **1er Admissible, 4<sup>ème</sup> Admis**

1989 CAPES (Externe) de Lettres Modernes - **1er Admis**

### Carrière professionnelle

**2006 Promotion à la 1<sup>ère</sup> classe du corps des Professeurs d'Université**

**2004 Titulaire de la Prime d'Encadrement Doctoral et de Recherche (2005-2008)**

**2003 Professeur** de Langue et Littérature françaises à l'Université de Franche-Comté.

**2000-2003 Maître de Conférences** de Langue et Littérature françaises, à l'Université de Franche-Comté.

**1992-1999 PRAG** (Lettres Modernes) chargé de l'enseignement de la linguistique française et de la stylistique littéraire à l'Université de Franche-Comté.

**1989-1992 Professeur certifié** puis **agrégé** dans l'enseignement secondaire de 1989 à 1992.

**1976-1989** Première carrière dans les Hôpitaux Publics.

## 1. Production scientifique

### Généralités

Inscrit sur le volet *Linguistique française* de la 9<sup>ème</sup> section du CNU, qualifié aux fonctions de Professeur en 7<sup>ème</sup>, je centre mes travaux sur l'étude en corpus numérisés des productions langagières écrites, en français moderne et contemporain. Ces travaux convoquent les théories, concepts et méthodes de l'analyse du discours, de la linguistique textuelle et de corpus, de la philologie numérique, et de la statistique lexicale.

Après avoir travaillé principalement sur des productions littéraires des 19<sup>ème</sup> et 20<sup>ème</sup> siècles (Baudelaire, le roman de Balzac à Zola, Céline, Claudel), je me suis également tourné vers d'autres sphères du discours en rapport avec les intérêts du master et de la formation doctorale où j'interviens à titre principal. Il s'agit avant tout du discours de presse francophone, incluant des secteurs de la presse historique (régionale, coloniale), et la « petite presse » contemporaine (bulletins régionaux et locaux), et du discours politique postcolonial africain francophone, inclus les discours d'opposants en situation de contrainte (aveux, dénonciations).

Après avoir croisé les statistiques lexicales lors de mes travaux de doctorat (1992-96), j'ai été amené à m'intéresser scientifiquement et méthodologiquement à l'ensemble de la « chaîne » de traitement et d'analyse qui mène de la collecte et de l'établissement des « données » textuelles, jusqu'à leur exploitation disciplinaire et à leur interprétation interdisciplinaire.

J'ai formulé en ce sens le programme du pôle *Archive, Bases, Corpus* de la MSH-E de Franche-Comté, et construit autour de chacun des points de ce programme un réseau de coopérations qui seront détaillées plus loin.

J'ai été amené par un état de l'art déficient, clivé entre les réalisations du TAL peu orientées vers la réalité plénière des textes, et une statistique lexicale peu soucieuse de rigueur linguistique, à développer des outils logiciels propres, qui ont depuis l'origine été en dialogue avec les autres développements francophones du domaine (aujourd'hui auto-désigné sous le label *Textométrie* – projet ANR financé 2006-09).

### Réalisations

Environnement AstaDiag (Astartex-DiaTag) [http://laseldi.univ-fcomte.fr/document/viprey/page\\_JMV.htm](http://laseldi.univ-fcomte.fr/document/viprey/page_JMV.htm)

► **DiaTag** est un ensemble modulaire destiné à répondre aux besoins immédiats et approfondis d'élaboration et de préparation de bases textuelles et d'extraction de corpus textuels en français des 19-20<sup>èmes</sup> siècles. Révisé pour être compatible avec les prescriptions de la TEI et les contraintes de l'interopérabilité, il intègre des modules de correction d'OCR, d'extraction de sous-corpus, d'annotation linguistique poussée avec levée d'ambiguïté semi-automatique. Il recourt à des dictionnaires de corpus, à des scripts de levée d'ambiguïté, et prête une attention particulière aux convivialités de décision et à leur ergonomisation. **DiaTag** fait l'objet d'études spécifiques, co-financées par la Région Franche-Comté, en vue de tester la meilleure faisabilité de l'acquisition de grandes bases numériques plein texte à partir des collections de presse du 19<sup>ème</sup> siècle ou antérieures sans saisie clavier intégrale.

► **Astartex** est un environnement entièrement original, modulaire (de l'utilisation en 1<sup>er</sup> cycle à l'hypertexte avancé), à forte consistance statistique multidimensionnelle (avec recours à une heuristique à base de cartographie statistique), et fondé sur les outils d'un *dictionnaire dynamique multimodal* (contextes « bruts », histogrammes, cartes, sensibilité au clic vers diverses fonctions de saisie et de relance...).

Ces deux développements, intégrés dans le logiciel AstaDiag, font l'objet d'un enrichissement coopératif de ressources et sont associés à la construction de la plate-forme de fonctionnalités *Textométrie* en cours de développement dans le cadre d'un projet ANR. Ils font l'objet de perspectives d'intégration plus poussée avec les logiciels SATO (Daoust-UQAM : thèse Daoust, Ecole d'Eté MCD, visés ci-dessous) et Hyperbase (Brunet-UNSA-CNRS : délégation Viprey auprès de l'UMR BCL-CNRS).

## **2. Réseaux de coopération**

**Responsable de l'Ecole thématique « Méthodes Informatiques et Statistiques en Analyse de Textes » (MISAT) du CNRS, ouverte en 2009 (1<sup>ère</sup> session en juin 2009 à Besançon, 2<sup>ème</sup> session à Nice en 2010).**

Directeur de la collection *Lettres numériques* aux Editions scientifiques Honoré Champion Paris

Evaluateur au CRSH du Canada (5 expertises en 2006-2009)

Membre du Comité de programme des *Journées Internationales d'Analyse des Données Textuelles* (JADT) depuis 2004, organisateur de la session 2006 des *JADT* (à l'Université de Franche-Comté à Besançon) 19-21 Avril 2006.

Membre du Comité scientifique des *Journées Internationales (Workshop) NooJ* (resp. Pr Max Silberztein) depuis 2002, organisateur du Workshop 2005 (à l'Université de Franche-Comté à Besançon) Mai 2005.

Partenaire à titre principal (4 partenaires) du projet ANR *Textométrie* 2006-2009 (200 000 €)

Porteur du projet ANR *Presse Quotidienne Régionale sous la Troisième République* intégrant 2 UMR CNRS (ICAR Lyon2-ENSLSH et BCL-Nice).

Délégation au CNRS (2 années de 2007 à 2009).

### **Coopérations internationales établies (principales) :**

**UQAM** : Centre ATO, chaire de sociologie (et de 2002 à 2008 Chaire du Canada en Mondialisation, Citoyenneté, Démocratie – MCD – Pr Jules Duchastel) à l'UQAM. Voir thèse Daoust ci-dessous.

**Université de Lausanne** : Formation Doctorale IRIS IV des Universités lémaniques (Pr Jean-Michel Adam et Ute Heidmann : écoles d'été en 2003 et 2004).

**Université d'Oxford** : Pr Lou Burnard. Expérimentation des contraintes XML-TEI sur le corpus de presse historique du *Petit Comtois* (numérisation plein texte de collections en état médiocre de conservation)

**Université La Sapienza à Rome** : Pr Sergio Bolasco. Echanges de doctorants autour de la coopération

TALTAC :AstaDiag.

## **3. Animation scientifique**

### **3a. Responsabilités**

**Coordonnateur du pôle *Archive Bases Corpus* de la Maison des Sciences de l'Homme et de l'Environnement de Franche-Comté Claude-Nicolas Ledoux (depuis 2002)**

Directeur-adjoint de l'EA 3187 *Archives, Textes, Sciences des Textes* (ATST) depuis 2003

**Le pôle *Archive Bases Corpus* de la MSH-E** est l'un des trois programmes constituant autour desquels a été acquise la reconnaissance de la MSH-E dans le Réseau national des MSH, puis comme UMS, puis comme USR par le CNRS. Reconnu pour son originalité dans ce contexte, il rassemble les chercheurs en sciences humaines (historiens, psychologues, linguistes, littéraires) de l'établissement soucieux de donner à leurs recherches une consistance empirique forte, à base discursive et textuelle, et/ou de contribuer au développement logistique, méthodologique et épistémologique appelé et induit par de telles pratiques. C'est dans ce cadre qu'ont été renforcés les partenariats extérieurs et étrangers comme l'intégration aux réseaux JADT, Textométrie(ANR), AtoNet(CRSH Canada), au

consortium TEI, l'essor du réseau de développement et de recherche NooJ, la construction du réseau de recherche *Discours d'Afrique*, les partenariats éditoriaux comme celui avec l'éditeur Champion-Electronique.

**Au sein de l'EA 3187**, j'assume la responsabilité de l'axe de recherche « **Bases et ressources numériques** », qui est l'une des activités les plus typiques et anciennement implantées de ce laboratoire dédié aux **fonds d'archives d'auteurs** et de presse littéraire (Claudel, Barbey, Bloy, et pour les contemporains Claude-Louis Combet). Je suis chercheur **associé au Laseldi (EA 2181)** où fonctionne un groupe de recherche en analyse sémio-discursive à forte consistance de corpus textuels. Je partage mes directions de thèses et d'HDR entre ces 2 EA.

Ma délégation au CNRS (2007-9) m'a aussi permis d'approfondir les liens « naturels » et historiques avec le pôle « statistique lexicale et logométrie » de l'UMR Bases Corpus Langages, et d'envisager la mutualisation du développement logiciel d'Hyperbase avec l'apport d'AstaDiag.

### **3b. Relations au monde socio-économique**

Responsable **du contrat de recherche Représentations de l'Europe en Région Franche-Comté**. Appel à projets sur concours, du SGAR de Franche-Comté hiver-printemps 2008 (10 000 €)

Responsable du contrat de recherche *Etude sémio-lexicale de la mutualisation*. Appel à projets sur concours de la Ville de Paris - Fraternelle de recherche de la *Maison des métallos* (24 000 €)

Responsable du *Fonds d'Archive Numérique* (FANUM) soutenu pour diverses actions de recherche à dimension patrimoniale par la Ville de Besançon, la Région Franche-Comté (55 000 € sur 2006-2009)

Responsable du partenariat avec la Bibliothèque d'Etude et de Conservation de Besançon en vue de la numérisation plein texte de sa collection du *Petit Comtois*, principal quotidien régional de 1893 à 1944. Avec la Ville de Besançon pour sa mise en ligne.

## **4. Encadrement scientifique**

### **Thèses et HDR soutenues**

J'ai encadré l'**Habilitation à Diriger les Recherches de M. Philippe Schepens**, spécialiste d'Analyse du Discours, membre du directoire du LASELDI (EA 2181), soutenue en novembre 2008 sous le titre *Sujet, Parole, Discours*, devant les Pr Jacqueline Authier-Revuz, Sonia Branca, Jean-Michel Adam, Jacques Guilhaumou. M. Schepens a **été qualifié, et nommé Professeur des Universités en juin 2009**.

J'ai encadré l'**Habilitation à Diriger les Recherches de M. Alpha Ousmane Barry**, spécialiste d'Analyse du Discours, responsable du réseau international de recherche « Discours d'Afrique » (voir sur ce site), soutenue en juin 2010 devant les Pr Jules Duchastel, M. Ngalasso, Jean-Michel Adam, Marc Bohomme, Bernard Lamizet.

J'ai encadré la thèse de Mlle Virginie Lethier, qui a bénéficié d'une Allocation de la Région Franche-Comté et porte sur l'établissement et l'analyse d'une base textuelle de presse historique (*Presse Quotidienne Régionale sous la Troisième République*). Cette thèse, qui a déjà donné lieu à d'importantes publications, est elle aussi au centre de coopérations nationales et internationales de haut niveau. La base et les travaux qu'elle suscite sont également l'objet de recherches et d'applications dans le cadre du consortium TEI. Le jury : Jean-Michel Adam (Lausanne), Sonia Branca (Paris-3), Michael Palmer (Paris-3), Pierre Fiala (Paris-12), Jacques Guilhaumou (CNRS) a accordé à Mlle Lethier la mention la plus élevée. Mlle Lethier a été qualifiée et **nommée Maître de Conférences en septembre 2010**.

J'ai amené à soutenance, avec la meilleure mention, Mme Magali Bigey, sur une thématique d'analyse du discours littéraire qui a rassemblé un jury interdisciplinaire (littérature, linguistique, sociologie, info-com). Jury : Bruno Péquignot (Paris-3), Michel Bernard (Paris-3), Bernard Lamizet (Lyon-2), Pierre Fiala (Paris-12). Mme Bigey a été qualifiée par le CNU en 7<sup>ème</sup> et 71<sup>ème</sup> sections. **Mlle Bigey est qualifiée aux fonctions de MCF en 7<sup>ème</sup> section.**

### **Thèses encadrées actuellement**

J'encadre à titre principal 6 thèses de doctorat, inscrites entre 2004 et 2008 (mon HDR remonte à 2003).

La plus ancienne de ces thèses est celle de M. **François Daoust**, dont la soutenance est programmée **fin 2010**, et qui aura donc duré 6 années. Il s'agit d'une thèse exceptionnelle à bien des égards. M. Daoust est ingénieur au Centre ATO (Analyse de Textes par Ordinateur) de l'UQAM, et créateur du logiciel SATO, l'un des environnements de référence du domaine. Son doctorat est destiné à produire la synthèse de ses travaux au regard des plus récents développements de la philologie numérique et de la statistique textuelle. Cet encadrement et cette soutenance sont le support de relations de travail en réseau avec les chercheurs du domaine au Québec (Pr Duchastel, Pr Marcoux) et avec le Pr Lou Burnard (Oxford), l'un des responsables fondateurs de la TEI (Text Encoding Initiative), et avec le Pr Jean-Michel Adam (Lausanne) au titre des applications de linguistique textuelle. **JURY** : Yves Marcoux (Univ. de Montréal), Jules Duchastel (UQAM), André Salem (Paris-3), Lou Burnard (Oxford)

Deux thèses plus récemment inscrites portent également sur des corpus de presse, dans les univers historico-culturels des deux doctorantes concernées (Mlle Pham Thi-Thân - Indochine coloniale, Mlle Icheboudene - Algérie contemporaine).

La thèse de Mme Lalaoui est sa deuxième thèse, en littérature, après une première thèse en Sciences du Langage. Mme Lalaoui est Professeur à l'Université d'Oran et tête de pont de nombreuses coopérations scientifiques et pédagogiques avec l'Université de Franche-Comté. Elle vient en soutenance **fin 2010**.

#### **JURYS D'HDR (autres que sous ma direction)**

2007 (Rapporteur) Mme MAGRI-MOURGUES Véronique *Etude lexicométrique du genre 'récit de voyage' dans la littérature française*. Jury : Georges Molinié (Paris-4), Sylvie Mellet, Anna Jaubert dir. (Nice), Jean-Michel Adam (Lausanne)

2010 (Rapporteur) M. ABLALI Driss. Jury : Louis Panier (Lyon-2), Jean-Michel Adam (Lausanne), Dominique Ducard (Paris-12), J.-François Jeandillou (Paris-X), J.-François Bordron (Limoges)

#### **JURYS DE THESES DE 3<sup>ème</sup> CYCLE (autres que sous ma direction)**

2008 (Rapporteur) Co-tutelle ENS-Univ. d'Etat de Saint-Petersbourg. CHEPIGA Valentina *L'attribution d'un style d'"auteur" : le cas GARY / AJAR. Méthodologies contrastées d'analyse (génétique, linguistique et statistiques)* Jury : Irène Fenoglio dir (ITEM-ENS), André Derval (IMEC), Mikail Marushenko, Alexandre Zelenschikov (Etat St Petersburg), Raimond Piotrowski (Herzen St Petersburg)

2008 TZANAVARI Mirsini *Etude lexicale et lexicométrique de l'œuvre romanesque d'Hervé Bazin*. Jury : Michel Bernard (Paris-3), Claude Condé dir. (Besançon), Elisabeth DEMIROUGLOU (Thessalonique)

2008 MESFAR Slim *Analyse morpho-syntaxique automatique et reconnaissance des entités nommées en arabe standard*. Jury : Abdelmajid Ben Hamadou (Sfax), Max Silberztein dir. (Besançon), Denis Maurel (Tours), Odile Piton (Paris-1), Paul Sabatier (Marseille), Jean-Patrick Guillaume (Paris-3)

2007 MATAR Soad *A l'ombre du 11 septembre 2001 : mise en mots du terrorisme et "Islamalgame" dans la presse écrite en France* Jury : Marc Bonhomme (Berne), Jean-Blaise Grize (Neuchatel), Andrée Chauvin-Vileno dir. (Besançon), Alain Rabatel (Lyon-2)

2005 (Rapporteur) LEBLANC Jean-Marc (Paris XII) *Les vœux des Présidents de la Vème République : Analyse et expérimentation lexicométriques sur un genre discursif* Jury : André Salem (Paris 3), Etienne Brunet (Nice), Dominique Maingueneau (Paris 12), Pierre Fiala dir. (Paris 12)

2004 ZELLAGUI Katia (Besançon, Allocataire de Recherche de la Région Franche-Comté) *Principes et application aux textes littéraires de l'étiquetage linguistique (Du côté de chez Swann, de Proust)* Jury : Michel Bernard (Paris 3), Isabelle Turcan (Lyon 2), Claude Condé (Besançon), Max Silberztein dir. (Besançon)