

Curriculum Vitæ  
**Abdel-Malek BOUALEM**  
- Mai 2015 -

---

**Coordonnées professionnelles :**

France Télécom Orange Labs  
OLNC/OLPS/OPENSERV/CONTENT/FAST  
2, avenue Pierre Marzin - 22307 Lannion  
Tél bureau: +33.02.96.07.29.83  
Tél mobile: +33.06.85.71.40.63  
Email: malek.boualem@orange.com

**Informations personnelles :**

54 ans  
Marié, 3 enfants  
Nationalité Française

---

**Profil global de Abdel-Malek BOUALEM :**

- Lauréat du Prix CNRS/ANVIE de la valorisation de la recherche scientifique, 1996.
- Co-auteur de l'éditeur multilingue MtScript largement diffusé dans le monde, 1995-1997 : <http://aune.lpl.univ-aix.fr/projects/multext/mtscript/>
- Discours à l'Assemblée Nationale Française sur les Technologies de l'Information et de la Communication dans le bassin méditerranéen, dans le cadre du projet de l'Union pour la Méditerranée, octobre 2008.
- Membre du Conseil d'Administration de l'ATALA entre 2003 et 2009.
- Membre de nombreux comités scientifiques de revues et de conférences internationales.
- Expert évaluateur de projets soumis à des programmes nationaux, européens et canadiens.
- Expert en traitement automatique des langues naturelles, en traduction automatique et en multilinguisme en informatique.
- Expert Senior au centre de recherche de France Telecom Orange Labs (filiale Recherche).
- Responsable de plusieurs projets de recherche au sein de France Telecom Orange Labs.
  
- Encadrement de thèse, de 3 postdocs, d'ingénieurs et de stagiaires à Orange Labs.
- Membre de jurys de thèses de doctorat en informatique en France et en Algérie.
- Membre du jury de l'examen national du BTS "Comptabilité et Gestion", Académie de Nice (Alpes Maritimes, Var et Corse), 1993-1994.
- Enseignement supérieur durant 12 ans (1988 à 2000) : informatique, programmation, statistiques, systèmes d'information et bases de données, TALN et Traduction Automatique.
- Qualifié aux fonctions de Maître de Conférences en 27ème section (informatique), 1998.
  
- Diplôme d'ingénieur en informatique 1986.
- Diplôme de Master (DEA) en informatique, université de Nice, 1987.
- Diplôme de doctorat en informatique préparée à l'INRIA et soutenue à Nice en 1993.

---

**Remarques :**

1. Le CV est dans sa version détaillée.  
La table des matières permet une lecture ciblée.
2. La liste des publications se trouve à la fin du document.

## Contenu du curriculum vitae détaillé

|  |    |
|--|----|
| I. Partie 1 : Période antérieure à France Télécom Orange Labs .....                                | 4  |
| 1. Distinctions et participations honorifiques .....   | 5  |
| Prix CNRS/ANVIE de la valorisation de la recherche scientifique.....                               | 5  |
| Expert évaluateur de projets européens .....   | 5  |
| Participation aux enseignements du DEA IST (1996 à 2000) .....                                     | 5  |
| Conférencier invité dans des écoles d'été .....  | 5  |
| Qualification aux fonctions de Maître de Conférences.....  | 5  |
| 2. Diplômes .....  | 5  |
| Liste des diplômes.....  | 5  |
| Thèse de doctorat .....  | 5  |
| Jury de thèse de doctorat .....  | 6  |
| 3. Activités de recherche .....  | 6  |
| Domaine de recherche.....  | 6  |
| Travaux de recherche à l'INRIA.....  | 6  |
| Travaux de recherche au CNRS.....  | 7  |
| Travaux de recherche au CRL (USA).....   | 7  |
| Animation de séminaires.....   | 7  |
| Participation à des groupes de recherche et de réflexion européens et internationaux.....          | 8  |
| Activités d'encadrement et de formation à la recherche .....                                       | 8  |
| Autres activités d'encadrement.....  | 8  |
| Participation à des workshops et salons informatiques.....   | 9  |
| 4. Expérience dans l'industrie .....   | 9  |
| 5. Activités dans l'enseignement supérieur et diffusion de supports pédagogiques.....              | 9  |
| Etablissements d'enseignement supérieur .....  | 9  |
| Public .....   | 9  |
| Thèmes enseignés .....   | 10 |
| Conception et diffusion de supports pédagogiques sur le World Wide Web. ....                       | 10 |
| Participation aux charges administratives et collectives.....                                      | 10 |
| Participations honorifiques relatives aux activités pédagogiques.....                              | 10 |
| II. Partie 2 : Période au sein de France Télécom Orange Labs .....                                 | 11 |
| 1. Activités de chef de projets de R&D .....   | 12 |
| PRP TRANSAT .....  | 12 |
| OMTS (Orange Mobile Translation Service) : piste de transfert A-TTM des résultats de TRANSAT ..... | 12 |
| Projet collaboratif EURESCOM P923.....   | 12 |
| Projet collaboratif EURESCOM P1104.....  | 13 |
| Programme collaboratif national TECHNOLANGUE.....  | 13 |
| Pilotage du projet de développement Suite logicielle TiLT 2010 .....                               | 13 |
| Pilotage du projet de développement TiLT industrialisation des composants.....                     | 13 |
| Pilotage du projet de recherche NLP & Semantics 2011 .....   | 13 |
| Pilotage du projet de recherche S&LP 2012 .....  | 13 |
| Pilotage du projet de recherche S&LP 2013 .....  | 14 |
| Pilotage du projet d'anticipation TiLT_Products.....   | 14 |
| Pilotage du projet d'anticipation Composants techniques TiLT et Reperio .....                      | 14 |
| 2. Activités de responsable de lots dans des projets ou des pôles de recherche .....               | 14 |
| Pilotage de lot (Work Package) du projet de recherche QUAERO/MSSE.....                             | 14 |
| Pilotage de lots et de tâches des Pôles et Macro-pôles de Recherche.....                           | 14 |
| Pilotage de tâche du projet PAVIS .....  | 15 |
| Pilotage du lot OMTS du projet NEWCOM.....   | 15 |
| Pilotage de tâche du projet REPERIO/QUAERO sur la traduction.....                                  | 15 |

|   |    |
|---|----|
| Pilotage d'un lot du projet SERACH-05 sur un prototype de moteur arabe .....                | 15 |
| Pilotage de tâche du projet STILTON .....   | 16 |
| Pilotage de lot du projet TILT.....   | 16 |
| 3. Reconnaissance scientifique nationale et internationale.....                             | 16 |
| Membre du Conseil d'Administration de l'ATALA.....  | 16 |
| Reconnaissance en qualité d'expert sur les NTIC en langue arabe.....                        | 17 |
| Reconnaissance en qualité d'expert en recherche à France Télécom.....                       | 17 |
| Membre du réseau d'excellence FLReNet.....  | 17 |
| Review de livres.....   | 17 |
| Expert évaluateur du programme Technolanguage .....   | 17 |
| Expertise d'un projet de traduction pour l'INRIA .....                                      | 17 |
| Expertise de projets de recherche pour le compte du CRSNG au Canada .....                   | 17 |
| Participation aux jurys de thèse et présidence de sessions de conférences.....              | 17 |
| Participation à des comités scientifiques de revues et de conférences .....                 | 18 |
| Conférences invitées ... quelques exemples.....   | 18 |
| Organisation de manifestations scientifiques.....   | 18 |
| 4. Activités de recherche.....  | 19 |
| Traduction automatique .....  | 19 |
| Traitement automatique de la langue arabe .....   | 19 |
| Internationalisation et localisation des applications et des services .....                 | 20 |
| Encadrement de la recherche .....   | 20 |
| 5. Activités de Conseil au sein du groupe ORANGE .....                                      | 20 |
| Conseil pour le projet JEI (Jordan Education Initiative).....                               | 20 |
| Conseil dans le cadre de la mise en place du laboratoire ILAB du Caire.....                 | 20 |
| Conseils pour les unités d'Affaires de France Télécom .....                                 | 20 |
| Projet de localisation des services du portail Voila en différentes langues .....           | 21 |
| Etudes des opportunités des projets collaboratifs nationaux ou européens .....              | 21 |
| 6. Diffusion des résultats des projets au sein de France Télécom.....                       | 21 |
| 7. Activités de formation interne et externe à FTR&D.....                                   | 21 |
| Réunion d'Animation Technique à France Télécom.....   | 21 |
| Enseignement universitaire à l'ENSTB.....   | 21 |
| Enseignement à l'université de Marne La Vallée.....   | 21 |
| 8. Activités de valorisation des travaux de R&D au sein et en dehors de France Télécom..... | 22 |
| Déploiement de l'abrégeur sur le portail intranet France .....                              | 22 |
| Déploiement de logiciels sur la plateforme du bureau-du-futur .....                         | 22 |
| Protection des technologies.....  | 22 |
| Construction d'une offre globale en TALN.....   | 22 |
| Conseil et aide à la mise en place d'une plateforme de démonstration.....                   | 22 |
| Activité de veille et étude de l'état du marché et de la concurrence .....                  | 22 |
| Valorisation de la R&D auprès des industriels .....   | 22 |
| Valorisation de l'abrégeur auprès de THALES.....  | 23 |
| Valorisation des technologies linguistiques auprès de Wanadoo Jordanie.....                 | 23 |
| 9. Représentation de France Télécom dans des manifestations internationales .....           | 23 |
| Membre observateur dans des groupes de normalisation .....                                  | 23 |
| Atelier Semantic Web Technologies à Luxembourg .....  | 23 |
| Séminaire du Centre Français du Commerce Extérieur.....                                     | 23 |
| Autres manifestations.....  | 23 |
| 10. Publications scientifiques.....   | 24 |
| Publications internationales .....  | 24 |
| Publications internes de rapports techniques .....  | 27 |
| Publications COMET à France Télécom .....   | 27 |
| Publications techniques sur le Web .....  | 27 |

## **I. Partie 1 : Période antérieure à France Télécom Orange Labs**

**Cette partie du curriculum vitæ correspondant  
à la période antérieure à l'intégration de  
France Télécom Orange Labs**

**(avant août 1999)**

## 1. Distinctions et participations honorifiques

### Prix CNRS/ANVIE de la valorisation de la recherche scientifique

Le logiciel **MtScript**, produit de plusieurs années de recherche que j'ai menée à l'INRIA et au CNRS, a obtenu le prix CNRS/ANVIE 1996 de la valorisation de la recherche scientifique. Cette distinction récompense une action de valorisation de la recherche dans le secteur socio-économique. MtScript est un logiciel d'édition de documents multilingues intégrant des normes standards pour l'échange d'informations à travers Internet et sur les différentes plates formes. Il a été réalisé dans le cadre du projet européen MULTEXT et a rencontré un succès considérable auprès de la communauté scientifique mondiale. MtScript.1.1 (pour Solaris et Linux) peut être téléchargé gratuitement à l'URL : <http://www.lpl.univ-aix.fr/projects/multext/MtScript/>

### Expert évaluateur de projets européens

Expert évaluateur de projets européens, DGXIII de la Commission Européenne, Programme "Telematics Applications Programme", 1998-1999.

### Participation aux enseignements du DEA IST (1996 à 2000)

Participation aux enseignements du DEA des industries des langues, organisé par les universités Marne La Vallée, Paris VII, Genève, INSTN, CNAM et Aupelf-Uref. Les étudiants sont des ingénieurs en informatique ou des linguistes (thème du cours : traduction automatique, théorie et réalisation pratique de mini-systèmes de traduction), 1996 à 2000.

### Conférencier invité dans des écoles d'été

- Summer School in Language Engineering, New Mexico State University, USA, June 1999.
- Ecole d'été sur la Communication-Homme-Machine, laboratoire CLIPS, avec le soutien de l'IMAG et de l'IESA, Alpe d'Huez, septembre 1998.

### Qualification aux fonctions de Maître de Conférences

- Qualifié aux fonctions de Maître de Conférences en 27ème section (informatique), 1998.

## 2. Diplômes

### Liste des diplômes

- Doctorat en sciences (informatique): Université de Nice Sophia Antipolis, INRIA, 1993.
- DEA (Master) en informatique: Université de Nice Sophia Antipolis, INRIA, 1987.
- Diplôme d'ingénieur en informatique: INES d'informatique, Sétif, Algérie, 1986.

### Thèse de doctorat

- **Université, laboratoire et équipe d'accueil:** Université de Nice Sophia Antipolis, centre de recherche INRIA (Institut National de Recherche en Informatique et en Automatique), projet Malin (Multi-alphabétisme et multilinguisme en informatique), Sophia Antipolis, France.
- **Titre de la thèse:** ML-TASC: système de conversion de formalismes de langages techniques et scientifiques dans un environnement à syntaxe contrôlée et à contexte limité (traduction automatique multilingue).
- **Liste étendue des mots-clés:** traitement automatique des langues naturelles, traduction automatique multilingue, morphologie, lexiques, syntaxe, sémantique, langage pivot, analyse, transfert et génération de textes, documents multilingues, automates à états finis.

## **Jury de thèse de doctorat**

- P.Bernhard, Président, Ancien Directeur de l'INRIA Sophia Antipolis, Professeur à l'ESSI.
- J-M.Pierrel, Rapporteur, Professeur à l'Université de Nancy.
- J-L.Petit, Rapporteur, Professeur à l'Université d'Aix-Marseille, Directeur du LERSI.
- M.Mlouka, Directeur, Directeur de recherche à l'INRIA Sophia Antipolis.
- H.Zinglé, Examineur, Professeur à l'Université de Nice, Directeur du Laboratoire LILLA.

## **3. Activités de recherche**

### **Domaines de recherche**

Mes travaux de recherche se situent dans le domaine du "multilinguisme, traitement automatique des langues naturelles, traduction automatique et localisation des applications informatiques". J'ai entrepris ces travaux en 1986 à l'INRIA, je les ai ensuite poursuivis au CNRS et à l'université de Provence, puis à l'université du Nouveau Mexique aux USA et je poursuis ces activités au centre de recherche France Télécom Orange Labs. Ces travaux s'appuient sur une évolution permanente des techniques de traitement de l'information dans les différentes langues et s'inspirent des travaux menés dans les différents pays (USA, France, Allemagne, Canada, Japon, etc.).

### **Travaux de recherche à l'INRIA**

L'un des aspects fondamentaux du multilinguisme en informatique est "le traitement des textes et des caractères". Lors de mon DEA (Master), en m'appuyant sur quelques travaux réalisés dans ce domaine (J.D.Becker, Winsoft, Universal Word), j'ai travaillé sur la conception et la réalisation d'un environnement d'édition de textes multilingues. Les problèmes rencontrés (et qui sont résolus de nos jours) se situaient au niveau du codage des caractères autres que les caractères latins (arabe, hébreu, russe, chinois, etc.) et leur gestion à l'écran (notamment dans le cas d'écriture de la droite vers la gauche). J'ai participé à la mise au point d'un éditeur de textes multilingues (TE: Text Editor) qui a été par la suite transféré vers l'industrie (voir GF2 plus bas). Parmi les modules mis en place:

- Analyseur de contexte pour l'identification des langues et l'application de règles spécifiques à chacune des langues (sens d'écriture, etc.) pour la saisie de textes multilingues.
- Programme de saisie phonétique des idéogrammes chinois : basé sur la norme phonétique "Pinyin" et la norme de codage des idéogrammes "GB-2312-80".

Ensuite, dans le cadre de ma thèse de doctorat, je me suis intéressé à un domaine fortement lié au multilinguisme : "la traduction automatique et la traduction assistée par ordinateur". J'ai ainsi travaillé sur la validation des modèles de traduction:

- modèle à langage pivot (cf. GETA, B.Vauquois, C.Boitet, systèmes Ariane, Titus, etc.),
- modèle à transfert (cf. travaux du projet Eurotra en particulier, L.Danlos).

Une étude du modèle bilingue (cf. premières versions de Systran, Salat, etc.) m'a conduit à comprendre les limites de celui-ci. Ainsi, j'ai travaillé sur un modèle "multilingue" à structure intermédiaire abstraite associant les modèles à pivot et à transfert. La validation a été faite du français vers l'arabe (langues représentatives de deux familles différentes de langues: indo-européennes et sémitiques). Bien entendu, le domaine d'application choisi (notices techniques) m'a conduit à utiliser un formalisme de représentation du langage à contexte limité et à syntaxe contrôlée. Pour la mise en place des lexiques, je me suis inspiré du modèle développé au LADL (M.Gross) et à EUROTRA que j'ai argumenté de traits sémantiques (sémantique lexicale, cf. travaux du LIM Marseille). Pour la modélisation des grammaires utilisées

(contexte libre, N.Chomsky), j'ai utilisé des automates à états finis. Enfin, pour la validation du modèle "multilingue", j'ai présenté une simulation de traduction vers le chinois à partir de la structure intermédiaire issue de l'analyse de textes sources en français (système ML-TASC: Sun, Unix/Solaris, C, X11).

## **Travaux de recherche au CNRS**

De 1995 à 1998, j'ai rejoint le laboratoire Parole et Langage du CNRS ESA 6057, où j'ai participé au sein de l'équipe "Traitement automatique des langues" à l'étude, la conception et la réalisation d'outils pour le traitement linguistique des documents multilingues. Une étude des avancées technologiques dans ce domaine m'a permis d'identifier des méthodes de codage des caractères et des documents, ainsi que des techniques de saisie et de translittération des caractères et qui étaient en phase de devenir des normes internationales (Unicode/ISO10646, SGML/HTML, TEI, etc.). J'ai étudié les travaux menés aux USA (laboratoire CRL, D.E.Knuth, P.MacKay, Accent, etc.), en Israël (D.Berry, U.Habusha sur l'écriture bi-directionnelle), au Canada (Labonté, etc.) et en Europe (GUTenberg, Omega, etc.) et j'ai travaillé sur la conception et la réalisation d'un éditeur multilingue (beaucoup plus efficace que celui réalisé à l'INRIA). Cet éditeur, appelé **MtScript**, présente un certain nombre de caractéristiques dont :

- compatibilité avec les normes de codages des caractères et des documents,
- utilisable pour le transfert de textes multilingues sur Internet (codage MIME 8 bits),
- mixage de textes à écriture en sens opposés, etc.
- Environnement portable: SUN, UNIX, Tcl/Tk, C, SGML, ISO-8859-\*, Unicode.

MtScript a été bien accueilli et largement utilisé au sein de la communauté scientifique mondiale et a reçu le prix CNRS/ANVIE 96 de la valorisation de la recherche scientifique (URL: <http://www.lpl.univ-aix.fr/projects/multext/MtScript/>).

## **Travaux de recherche au CRL (USA)**

Entre octobre 1998 et juillet 1999, j'ai intégré le laboratoire CRL (Computing Research Laboratory) de l'université du Nouveau Mexique aux USA où j'ai dirigé des travaux de recherche et de développement d'un environnement d'édition de documents multilingues basé sur Unicode. Cet environnement est couplé à un système de traitement automatique des langues naturelles, incluant un traducteur automatique multilingue. Les travaux concernaient également des aspects assez peu explorés : écriture bi-directionnelle, justification des textes multilingues, identification des langues dans un texte multilingue, codage des documents multilingues (HTML, XML) pour le Web.

Par ailleurs, j'ai également participé au développement d'un composant à base de transducteurs à états finis pour l'analyse morphologique et syntaxique.

## **Animation de séminaires**

1994 A.M.Boualem, "Analyse morphologique", Séminaire du centre de recherche XEROX, Meylan, Grenoble, Septembre 1994.

1995 à 1998 Animation de séminaires du Laboratoire Parole & Langage, Université de Provence, Edition de textes multilingues, Traitement automatique du langage naturel, Traduction automatique.

## **Participation à des groupes de recherche et de réflexion européens et internationaux**

- 1997 Participation au séminaire de réflexion sur l'avenir de l'ingénierie linguistique en Europe, EUROMAP, Paris La défense, octobre 1997. Ce séminaire réunissait des représentants de la Commission Européenne, du Gouvernement et des centres de recherche européens.
- 96/98 Membre de plusieurs groupes de travail et listes de distribution regroupant des chercheurs dans les domaines du traitement automatique des langues et du multilinguisme: GdR-PRC I3 (langage et parole), TC46SC2 (translitération), VICNET (multilinguisme), READER (multilinguisme), UNICODE (Unicode), etc.

## **Activités d'encadrement et de formation à la recherche**

- 1998-1999 Responsable d'une petite équipe de R&D constituée d'un ingénieur et deux étudiants de 3e cycle pour la mise en place d'un environnement d'édition de textes multilingues (Java, Unicode), CRL Computing Research Laboratory, New Mexico State University (USA).
- 1996-1999 Encadrement des projets des étudiants en DEA de génie linguistique de l'Aupelf pour la mise en place de mini-systèmes de traduction assistée par ordinateur.
- 1996-1997 Responsable de 2 stages d'initiation à la recherche (2 étudiants ingénieurs) organisé par l'IUSPIM (Institut universitaire des sciences pour l'ingénieur de Marseille): synthèse bibliographique sur l'écriture bi-directionnelle et rédaction d'articles de recherche conformément aux conditions de publication des revues.
- 1995-1997 Suivi des travaux de recherche et applications informatiques des étudiants en thèse, en DEA et en Maîtrise de phonétique, Institut de phonétique, Université de Provence.
- 1997-1998 Suivi des projets des étudiants de l'UV INFZ15 (Informatique, internet et communication) pour la réalisation de pages Web, Université de Provence (URL: <http://www.up.univ-mrs.fr/~wcilsh/>).
- 1989-1990 Encadrement de stage de DEA en informatique "Structure de description des idéogrammes chinois", INRIA Sophia Antipolis, Université de Nice.
- 1990-1991 Encadrement de stage de BTS en informatique "Application des bases de données client-serveur", CNAM, Sophia Antipolis.

## **Autres activités d'encadrement**

- 1997-1998 Encadrement et suivi des tuteurs (15 par semestre) chargés des TP d'informatique, CILSH, Université de Provence, Aix-en-Provence.
- 1993-1995 Définition et suivi des stages en entreprises "Systèmes d'information et bases de données" et organisation des soutenances, IPAG Nice.

1991-1995 Définition et suivi des stages en entreprises “Gestion, Analyse de marché” et organisation des soutenances, ESM/EST/EITC, Nice, Sophia Antipolis.

### **Participation à des workshops et salons informatiques**

1989 Salon du SICOB: présentation du terminal multilingue de l’INRIA, Paris, 1989.

1991 Salon Eurospore: présentation de l’interface multilingue Multipolis GF2/INRIA, Hyères, 1991.

## **4. Expérience dans l’industrie**

### **1990/1991 et 1993**

Société GF2 Informatique / INRIA, Saint-Laurent-du-Var (06), France (société spécialisée en développement d’applications multilingues, localisation de logiciels). J’ai travaillé en qualité d’ingénieur d’études sur le développement d’un logiciel de traitement de textes multilingues (MULTIPOLIS V.1.0). J’étais également chargé de la recherche et la prospection des normes de codages des caractères et des documents. Environnement: IBM-PC, UNIX, MS-DOS, C.

### **1994**

Groupe ASCOM/Monetel, Granges-Valence (07), France (groupe spécialisé dans les télécommunications). Dans le cadre d’un marché pour la Chine, j’ai effectué une mission d’expertise, où j’ai piloté un projet de localisation en chinois de l’interface de visualisation des publiphones. J’ai réalisé l’interface multilingue et formé des techniciens à sa maintenance. Environnement: HP9000, C, UNIX, MOTIF.

### **1998**

Groupe ASCOM/Monetel. Dans le cadre d’un marché pour l’Egypte, j’ai effectué une nouvelle mission d’expertise pour la localisation en langue arabe de l’interface de visualisation des publiphones.

## **5. Activités dans l’enseignement supérieur et diffusion de supports pédagogiques**

Mon expérience dans l’enseignement supérieur est conséquente. Elle s’étend sur une période de 12 ans (1988 à 2000). Cette activité ne sera pas développée dans ce document. Néanmoins, je me limite à en donner les grandes lignes.

### **Etablissements d’enseignement supérieur**

- Université de Provence, Aix-Marseille-1.
- Université de Marne La Vallée.
- Centre multimédia de l’université de Provence.
- IPAG (Ecole d’Administration et de gestion), grande école de commerce.
- ESM (Ecole Supérieure de Management), grande école de commerce.

### **Public**

- DEA Parole et Langage.
- DEA de génie linguistique.
- DEA informatique.
- DEUG et licence dans les filières des sciences humaines.

- Enseignants et chercheurs de l'université de Provence et du CNRS.
- Gestionnaires des bibliothèques de la région PACA.
- Elèves des écoles de commerce.

### **Thèmes enseignés**

- Traitement automatique du langage naturel.
- Traduction automatique.
- Multilinguisme en informatique.
- Nouvelles technologies de l'information et de la communication.
- Internet (initiation, applications, recherche d'information).
- Programmation.
- Systèmes d'information et bases de données.
- Codage des documents électroniques.
- Probabilités et statistiques.
- Analyse des données et techniques quantitatives de gestion.

### **Conception et diffusion de supports pédagogiques sur le World Wide Web.**

Informatique, Internet et Communication

URL: <http://www.up.univ-mrs.fr/~wcilsh/>

*ensuite accès aux cours en ligne, puis accès à INF Z15*

Initiation à l'informatique

URL: <http://www.up.univ-mrs.fr/~wcilsh/>

*ensuite accès aux cours en ligne, puis accès à INF Z10*

### **Participation aux charges administratives et collectives**

- Participation à l'organisation de séminaires.
- Participation à la gestion du parc informatique et à la planification du plan d'informatisation.
- Elaboration de supports pédagogiques.
- Tâches administratives.

### **Participations honorifiques relatives aux activités pédagogiques**

- Membre du jury de l'examen national du BTS "Comptabilité et gestion", Académie de Nice (Alpes Maritimes, Var et Corse), 1993-1994.
- Participation à un groupe national de travail sur le télé-enseignement, 1998.

*Fin de la 1<sup>ère</sup> partie du curriculum vitæ correspondant à la période  
antérieure à l'intégration de France Télécom Orange Labs  
(avant août 1999)*

---

## **II. Partie 2 : Période au sein de France Télécom Orange Labs**

*Cette partie du curriculum vitæ correspondant  
à la période postérieure à l'intégration de  
France Télécom Orange Labs  
(après août 1999)*

## 1. Activités de chef de projets de R&D

### **Projet de Recherche Pluridisciplinaire TRANSAT**

Nom du Projet : TRANSAT (Translation of speech and text).

Objet du projet : Traduction Automatique de l'écrit et de l'oral.

Date de début / fin : Mai 2004 / Mai 2007.

Mon rôle dans le projet : chef de projet.

Partenaires du projet : Plusieurs équipes des laboratoires de la division R&D, partenaires universitaires, partenaires externes techniques.

Le projet TRANSAT avait pour objet l'étude et la mise en œuvre d'une approche, validée par des démonstrateurs, pour la traduction automatique de l'écrit et de l'oral, à travers le couplage des technologies du langage naturel écrit et oral, développées à la division R&D de France Télécom (traduction automatique linguistique et statistique, reconnaissance et synthèse de la parole). Le projet visait également la combinaison des approches symboliques et statistiques en traduction automatique. J'ai assuré la responsabilité de ce PRP et la coordination des travaux menés par les membres de six équipes de R&D appartenant à trois laboratoires de la division R&D de France Télécom. J'ai également assuré la coordination des partenariats avec des laboratoires universitaires (CRE avec l'université Joseph Fourier et CRE avec l'université de Provence) et avec des partenaires industriels (sociétés de consulting et prestataires de services).

### **OMTS (Orange Mobile Translation Service) : piste de transfert A-TTM des résultats de TRANSAT**

Le projet TRANSAT a donné lieu à une piste de transfert opérationnel baptisée OMTS et qui est en cours d'instruction, sous ma coordination, dans le cadre du processus A-TTM et en partenariat avec le Marketing Stratégique. OMTS est proposé sous la forme d'un service d'ORANGE à destination de ses clients de téléphonie mobile en situation de déplacement à l'étranger. Il leur permet de soumettre, à partir d'un téléphone mobile, des requêtes de traduction par SMS ou par voie vocale. Une Unité d'Affaires a été identifiée et le dossier est en cours d'instruction pour un lancement éventuel du service en fonction des priorités TTM.

Il est à noter que OMTS a été candidat aux Orange Labs Awards 2007 et 2008. Un représentant de l'Unité d'Affaires d'Orange et un représentant du Marketing Stratégique s'étaient associés à moi pour cette candidature, qui n'avait malheureusement pas été retenue.

### **Projet collaboratif EURESCOM P923**

Nom du projet : Babelweb.

Objet du projet : "Multi-lingual Web Sites: Best Practice Guidelines and Architectures".

Date de début / fin : Juin 1999 / Décembre 2000.

Mon rôle dans le projet : PMC member ou chef de projet pour France Télécom.

Partenaires du projet : Eurescom (superviseur), KPN (chef de projet), FT, BT, PT et TI.

Ce projet avait pour objet de définir une architecture globale, validée par un démonstrateur, pour l'élaboration des sites web multilingues. Les aspects liés à l'internationalisation et à la localisation des applications informatiques étaient fortement abordés dans ce projet. Outre ma contribution au projet, j'avais également encadré la contribution de plusieurs collègues à FTR&D, notamment dans la phase de réalisation et d'évaluation du démonstrateur. Ce qui est important à souligner est que nous (FTR&D) étions responsables de la tâche centrale du projet (Work-Package 4), qui consistait à concevoir l'architecture globale de la plateforme. Ce projet a produit des résultats très satisfaisants (brochures Eurescom, démonstrateur, présentation dans plusieurs congrès, etc.) et il a été évalué de façon très favorable. Une continuité sur le thème du

multilinguisme a été fortement encouragée. Justement, un deuxième projet (Eurescom P1104) a été proposé par les mêmes partenaires pour 2001. Il a été accepté et a démarré en février 2001 pour une période de deux ans (voir ci-dessous).

### **Projet collaboratif EURESCOM P1104**

Nom du projet : MUST.

Objet du projet : "Multimodal, multilingual information services for small mobile terminals".

Date de début / fin : Février 2001 / Décembre 2002.

Mon rôle dans le projet : PMC member ou chef de projet pour France Télécom.

Partenaires du projet : Eurescom (superviseur), KPN (chef de projet), FT, BT, PT et Telenor.

Ce projet avait pour objet d'étudier et de mettre en œuvre une plateforme multimodale et multilingue, permettant à l'utilisateur d'accéder à des informations, dans sa propre langue, à travers des terminaux mobiles de type PDA et téléphones mobiles (UMTS était au premier plan au niveau applicatif). J'ai été le PMC member de France Télécom dans ce projet, dans lequel Orange Labs avait une participation diversifiée. En effet, j'avais sollicité pour ce projet un certain nombre d'équipes appartenant à différentes laboratoires et divisions d'Orange Labs. Le rôle de France Télécom dans le projet était relativement important en tant que fournisseur de technologies et d'expertise de haut niveau.

### **Programme collaboratif national TECHNOLANGUE**

J'ai contribué à la coordination de la participation de France Télécom aux deux projets :

1. Nom du projet : EVALDA-EASY.

Titre du projet : Evaluation des analyseurs syntaxiques.

Date de début / fin : 2005/2006.

Partenaires du projet : Nombreux acteurs universitaires et industriels du domaine du TALN.

2. Nom du projet : EVALDA-EQUER.

Titre du projet : Evaluation des systèmes de questions-réponses.

Date de début / fin : 2005

Partenaires du projet : Nombreux acteurs universitaires et industriels du domaine du TALN.

### **Pilotage du projet de développement Suite logicielle TiLT 2010**

Nom du projet : Suite logicielle TiLT 2010.

Objet du projet : Industrialisation de la technologie TiLT et études économiques.

Date de début / fin du lot : Avril 2010 / Décembre 2010.

Mon rôle dans le projet : Responsable du projet.

### **Pilotage du projet de développement TiLT industrialisation des composants**

Nom du projet : TiLT industrialisation des composants.

Objet du projet : Industrialisation des composants de la technologie TiLT.

Date de début / fin du lot : Janvier 2011 / Décembre 2011.

Mon rôle dans le projet : Responsable du projet.

### **Pilotage du projet de recherche NLP & Semantics 2011**

Nom du projet : Natural Language Processing & Semantics 2011.

Objet du projet : Extraction d'informations (entités nommées et relations).

Date de début / fin du lot : Janvier 2011 / Mars 2012.

Mon rôle dans le projet : Responsable du projet.

### **Pilotage du projet de recherche S&LP 2012**

Nom du projet : Semantic and Language Processing 2012.

Objet du projet : Acquisition des connaissances et extraction d'informations sémantiques.  
Date de début / fin du lot : Avril 2012 / Décembre 2012.  
Mon rôle dans le projet : Responsable du projet.

### **Pilotage du projet de recherche S&LP 2013**

Nom du projet : Semantic and Language Processing 2013.  
Objet du projet : Extraction d'informations sémantiques, désambiguïsation syntaxique et sémantique, acquisition de données sémantiques.  
Date de début / fin du lot : Février 2013 / Décembre 2013.  
Mon rôle dans le projet : Responsable du projet.

### **Pilotage du projet d'anticipation TiLT\_Products**

Nom du projet : TiLT\_Products.  
Objet du projet : Industrialisation et valorisation des composants de la technologie TiLT.  
Date de début / fin du lot : Janvier 2014 / Décembre 2014.  
Mon rôle dans le projet : Responsable du projet.

### **Pilotage du projet d'anticipation Composants techniques TiLT et Reperio**

Nom du projet : Composants techniques pour l'accès aux contenus.  
Objet du projet : Industrialisation et valorisation des composants des technologies TiLT pour le traitement automatique des langues naturelles et Reperio pour la recommandation.  
Date de début / fin du lot : Janvier 2015 / Décembre 2015.  
Mon rôle dans le projet : Responsable du projet.

## **2. Activités de responsable de lots dans des projets ou des pôles de recherche**

### **Pilotage de lot (Work Package) du projet de recherche QUAERO/MSSE**

Nom du projet : QUAERO/MSSE.  
Objet de la tâche : Intégration d'une fonctionnalité de CLIR ou recherche d'information cross-lingue à un moteur de recherche vidéo et multimédia.  
Date de début / fin du lot : Janvier 2008 / Décembre 2013.  
Mon rôle dans le projet : Responsable de lot (WP).

Cette fonctionnalité de CLIR (Cross Language Information Retrieval) a permis, via une requête en français, d'accéder à des contenus dans plusieurs autres langues (anglais, allemand, espagnol, portugais, arabe, russe, chinois, ...). Cette fonctionnalité de CLIR utilise des mécanismes de traduction automatique des métadonnées associées aux vidéos (presse, news, sport, etc.).

### **Pilotage de lots et de tâches des Pôles et Macro-pôles de Recherche**

Depuis plusieurs années, je contribue de manière significative à la coordination des activités du Programme Vision, du Pôle ou du Macro-pôle de Recherche auxquels sont rattachées les activités de recherche en traitement automatique des langues naturelles (HyperArtLedge, DATALEDGE et Macro-Pôle Knowledge Sciences) :

- Responsabilité des lots et des tâches consacrés au thème de la traduction automatique.
- Contribution aux évaluations WORLD-CLASS.
- Coordination des activités de valorisation des résultats de la recherche.
- Coordination d'actions de partenariat avec le monde de la recherche académique (CREs).

- Contribution aux orientations stratégiques du pôle de recherche.
- Contribution à la définition des objets de la recherche.
- Etc.

### **Pilotage de tâche du projet PAVIS**

Nom du projet : PAVIS 2008.

Objet de la tâche : Valorisation, transfert et partenariats autour de la technologie TiLT

Date de début / fin du lot : Janvier 2008 / Décembre 2008.

Mon rôle dans le projet : Responsable de tâche.

La tâche dont j'étais responsable vise la coordination des activités suivantes :

- Documentation bilingue des manuels d'installation et d'intégration des logiciels d'application de TiLT.
- Mise en place de partenariats internes et externes autour des logiciels de TiLT.
- Coordination du déploiement de certains logiciels de TiLT sur les intranets de France Télécom.

### **Pilotage du lot OMTS du projet NEWCOM**

Nom du projet : NEWCOM 2007 et NEWCOM 2008.

Objet du projet : Instruire, maquetter et expérimenter des idées de services.

Titre du lot : OMTS (Orange Mobile Translation Service)

Date de début / fin du lot : Septembre 2007 / Juin 2008.

Mon rôle dans le projet : Responsable de lot.

Le lot OMTS du projet NewCom héberge l'instruction du dossier A-TTM OMTS en vue de la mise en place, par ORANGE, d'un service de traduction accessible sur téléphones mobiles et à destination des clients en situation de mobilité.

### **Pilotage de tâche du projet REPERIO/QUAERO sur la traduction**

Nom du projet : REPERIO 2007.

Objet du projet : Recherche d'information multimédia.

Titre de la tâche : Traduction automatique pour la recherche de vidéos

Date de début / fin du lot : Janvier 2007 / Décembre 2007.

Mon rôle dans le projet : Responsable de tâche.

L'activité consiste à coordonner la mise en place, l'évaluation et l'intégration d'un module de traduction statistique (technologie externe) au sein du moteur de recherche de vidéos. Cette tâche a été confrontée à la difficulté d'acquérir les corpus nécessaires pour la phase d'apprentissage du système de traduction.

### **Pilotage d'un lot du projet SERACH-05 sur un prototype de moteur arabe**

Nom du projet : SEARCH 2005.

Objet du projet : Opération de développement sur la recherche d'information.

Titre du lot : Prototype de moteur de recherche en langue arabe

Date de début / fin du lot : Septembre 2005 / Janvier 2006.

Mon rôle dans le projet : Responsable de WP.

L'idée de mettre en place un moteur de recherche en langue arabe à France Télécom n'est pas récente. Cette réflexion a été lancée en 2000 et l'étude de faisabilité de ce projet m'a été confiée. Depuis, j'ai mené un certain nombre d'études sur l'existant, sur la faisabilité technique (en mettant l'accent sur l'apport du TALN en recherche d'information) et sur le potentiel de

valorisation d'un tel moteur, en partenariat avec Wanadoo qui avait mis en place le portail trilingue HAHOOA ( <http://www.hahooa.com/> ).

Le lot "moteur arabe" du projet Search 2005, mené avec le soutien d'un partenaire externe, a produit les résultats et les livrables suivants :

- Étude du marché de la recherche d'information dans le monde arabe.
- Analyse et évaluation de plusieurs moteurs de recherche en langue arabe.
- Analyse d'usage des moteurs arabes.
- Réalisation d'un démonstrateur de moteur de recherche en langue arabe.

### **Pilotage de tâche du projet STILTON**

Nom du projet : STILTON.

Objet du projet : Valorisation et industrialisation des logiciels de la technologie TiLT.

Titre du lot : Valorisation externe des logiciels de TALN

Date de début / fin du lot : Janvier 2005 / Décembre 2005.

Mon rôle dans le projet : Responsable de la tâche de valorisation externe.

Principaux résultat :

- Coordination de la commercialisation d'une licence de l'abrégeur de textes à THALES.
- Présentation de la technologie TILT au SMSI 2005 à Tunis (Sommet Mondial sur la Société de l'Information).
- Nombreuses actions de communication et de valorisation interne et externe au groupe France Télécom.

### **Pilotage de lot du projet TILT**

Nom du projet : TILT.

Objet du projet : Valorisation des logiciels langues naturelles TiLT et du moteur Q2.

Titre du lot : Valorisation des logiciels de traitement des langues naturelles TiLT

Date de début / fin du lot : Janvier 2004 / Décembre 2004.

Mon rôle dans le projet : Responsable de lot.

Principaux résultat :

- Construction d'une offre de produits issus de la technologie TILT de traitement automatique des langues naturelles.
- Nombreuses actions de communication et de valorisation interne et externe au groupe France Télécom.
- Coordination d'un partenariat avec THALES.

## **3. Reconnaissance scientifique nationale et internationale**

### **Membre du Conseil d'Administration de l'ATALA**

L'ATALA (Association pour le traitement automatique des langues) se consacre depuis 1959 au développement de la linguistique informatique en France. Les activités de l'ATALA se concrétisent à travers sa revue TAL (Traitement Automatique des Langues), l'organisation ou le parrainage de conférences (notamment la conférence annuelle TALN), etc. La plupart des membres de l'association sont des chercheurs et des doctorants dans le domaine du TALN. Le Conseil d'Administration de l'ATALA réunit des chercheurs de renommée nationale et internationale. Je fais partie de ce Conseil d'Administration depuis 2003 (mandat renouvelé en 2006 pour une durée de 3 ans).

<http://www.atala.org/>

### **Reconnaissance en qualité d'expert sur les NTIC en langue arabe**

J'ai été invité, en qualité d'expert sur les Nouvelles Technologies de l'Information et de la Communication, à intervenir aux "Premières Assises de l'enseignement de la langue et de la culture arabes en France" qui ont eu lieu jeudi 9 octobre 2008 à l'Assemblée Nationale, dans le cadre du projet de l'Union pour la Méditerranée. Ces assises ont été organisés sous le haut patronage du Président **Nicolas Sarkozy** et le parrainage de l'**Emir de l'Etat du Qatar**. Elles ont été ouvertes par **Marc LAFFINEUR**, Vice-Président de l'Assemblée Nationale et clôturées par les membres du Gouvernement **Xavier Darcos** et **Fadela Amara**, juste après une allocution de **Dominique Baudis**, Président de l'Institut du Monde Arabe.

### **Reconnaissance en qualité d'expert en recherche à France Télécom**

- Reconnaissance en qualité d'expert confirmé en R à France Télécom R&D, juillet 2001.
- Reconnaissance en qualité d'expert senior en R à France Télécom R&D, juillet 2003.
- Reconnaissance en qualité d'expert senior en R à France Télécom R&D, juillet 2006.

### **Membre du réseau d'excellence FLAReNet**

FLAReNet est un réseau d'excellence qui a été proposé dans le cadre du programme eContentplus en fin 2007. Il vise la mise en place de recommandations, à destination de la Commission Européenne et des organismes nationaux (ANR en France), pour l'orientation des activités de recherche collaborative dans les domaines des ressources et des technologies des langues écrites et orales.

### **Review de livres**

J'ai été sollicité par l'éditeur de la revue internationale Speech Communication Journal pour évaluer un livre sur la traduction automatique (Translation Engines: techniques for Machine Translation, Arturo Trujillo, Springer-Verlag, Applied Computing, Heidelberg, ISBN 1-85233-057-0, 1999). La review que j'ai réalisée a été publiée dans la revue Speech Communication, Issue 32(3), en octobre 2000.

Par ailleurs, je suis régulièrement sollicité, de manière nominative, pour des reviews de livres. Cependant, par manque de disponibilité, je suis souvent amené à décliner ces demandes.

### **Expert évaluateur du programme Technolangue**

J'ai été sollicité en 2002 par le Ministère de la Recherche et j'ai fait partie des évaluateurs des projets proposés dans le cadre du programme Technolangue, juin 2002.

### **Expertise d'un projet de traduction pour l'INRIA**

J'ai été sollicité en 2000 par l'INRIA pour expertiser un projet dans le domaine de la traduction automatique en vue de son financement.

### **Expertise de projets de recherche pour le compte du CRSNG au Canada**

Le CRSNG, organisme de financement de la recherche au Canada, m'a sollicité deux fois (en 2001 et en 2008) pour évaluer des projets de recherche, en lien avec le TALN et la traduction automatique, proposés par des laboratoires de l'université de Montréal.

### **Participation aux jurys de thèse et présidence de sessions de conférences**

- Membre d'un jury de thèse en informatique, université de Provence, 08/12/2006.
- Membre d'un jury de thèse en informatique, université Joseph Fourier, Grenoble, 31/10/2000.

- Président d'une "session plénière" de la conférence TALN 2004 consacrée au traitement automatique de la langue arabe écrite et parlée, Fès, Maroc, 2004.
- Président d'une session de la conférence internationale CII'99, Annaba, novembre 1999.

### **Participation à des comités scientifiques de revues et de conférences**

- Membre du comité scientifique des conférences TALN 1998, TALN 2002, TALN 2003, TALN 2004, TALN 2005, TALN 2006, TALN 2007, TALN 2008, TALN 2009, TALN 2010, TALN 2011, TALN 2012.
- Membre du comité scientifique de la conférence CITALA 2014.
- Membre du comité scientifique de l'un des workshops de la conférence LREC 2008.
- Membre du comité scientifique de l'un des workshops de la conférence ACL 2007.
- Membre du comité scientifique de la conférence SIIE2008.
- Membre du comité scientifique de la revue Computer Speech and Language, mars 2006.
- Membre du comité scientifique de la revue Information Retrieval Journal (Kluwer), mars 2004 et mars 2005.
- Membre du comité scientifique de la revue CHUM, Computers and the Humanities, septembre 2003 et août 2004.
- Membre du comité scientifique de la revue RIST (revue de l'information scientifique et technique) édité par le centre de recherche CERIST à Alger.
- Membre du comité scientifique de la conférence UAHCI 2005 Special Session on "Multilingual, Multimodal User Interfaces and Avatars", Las Vegas, USA, 22-27 July 2005.
- Membre du comité scientifique du Workshop, Arabic Language Resources and Evaluation, Post-Conference Workshop of LREC 2002, Las Palmas (Spain), 1st June 2002.
- Membre du comité scientifique du Workshop on Arabic Language Processing, ACL 2001, Toulouse juillet 2001.
- Membre du comité scientifique de COLING 2000 (Computational Linguistics, Saarbrücken).
- Membre du comité scientifique de la conférence I3S'2000, Constantine, Algérie.
- Membre du comité scientifique de la conférence CII'1999, Annaba, Algérie.

### **Conférences invitées ... quelques exemples**

- Conférence invitée à l'université de Paris 7 sur la traduction automatique, 2001.
- Conférence invitée à l'université de Grenoble2 sur le TALN, 2004.
- Conférence invitée à l'IMA (Institut du Monde Arabe) sur le thème de la recherche d'information en langue arabe, juin 2001.
- Conférencier invité aux congrès TELMI 2002 et TELMI 2006, dédiés aux technologies linguistiques pour le management de l'information, 2002 et 2006.
- Conférencier invité à LREC2002, Workshop on Arabic Language Resources and Evaluation, Las Palmas, Espagne, 29 mai – 2 juin 2002 (invitation déclinée par manque de disponibilité).
- Conférencier invité sur le thème de la traduction automatique au congrès international sur le TALN, Alger, juin 2005 (invitation déclinée par manque de disponibilité).
- Autres conférences invitées, ...

### **Organisation de manifestations scientifiques**

- **TALN 2004, Session spéciale sur le traitement de la langue arabe**  
Co-organisation d'une session plénière et d'une session de posters sur le traitement automatique de la langue arabe écrite et orale, TALN 2004, Fès, Maroc, Avril 2004.
- **TALN 2003, Workshop sur le TALN et le multilinguisme**  
Co-organisation d'un workshop sur le thème "TALN et du multilinguisme", Conférence TALN 2003, Batz-sur-Mer, France, Juin 2003.
- **TALN 2005, Workshop sur les langues peu dotés**

Co-organisation d'un workshop sur le thème "Traitement des langues peu dotés", Conférence TALN 2005, Dourdan, France, Juin 2005.

- **TALN 2006, Workshop sur Les Ressources dans le traitement de la langue**  
Co-organisation d'un atelier sur Les Ressources dans le traitement de la langue écrite : L'apport de Technolangue et les enjeux industriels, Conférence TALN 2006, Louvain, Belgique, Avril, 2006.
- **HCII 2003, Special session on Ontologies and Multilinguality in User Interfaces**  
Co-organisation d'une session spéciale sur le thème "Ontologies and Multilinguality in User Interfaces", 2nd International Conference on Universal Access in Human-Computer Interaction, Crete, Greece, Juin 2003.
- **UAHCI 2005, Special session on Multilingual, Multimodal User Interfaces and Avatars**  
Co-organisation d'une session spéciale sur le thème "Multilingual, Multimodal User Interfaces and Avatars", conférence UAHCI 2005, Las Vegas, Nevada, USA, 22-27 July 2005.

## 4. Activités de recherche

### Traduction automatique<sup>1</sup>

#### Réalisation d'un état de l'art en traduction

J'ai réalisé un état de l'art riche et assez exhaustif des systèmes de traduction automatique existants sur le marché (caractéristiques, fournisseurs, tarifs, etc.). Cet état de l'art inclut également les systèmes prototypes ou les travaux de recherche en cours dans le domaine de la traduction automatique. Cette étude est disponible sur le site Intranet d'Orange.

#### Encadrement d'une thèse en traduction automatique

J'ai assuré l'encadrement d'une thèse en informatique dans le domaine de la traduction automatique, en collaboration avec le Professeur J.Véronis de l'université de Provence. La soutenance de cette thèse a eu lieu le 8 décembre 2006.

#### Génération automatique en langage naturel

Notre équipe a démarré une activité en génération automatique de texte en langage naturel sur la base des outils d'analyse linguistique déjà en place. Pour cela, un ingénieur apprenti a travaillé dans notre équipe en 2001 et en 2002. Il a travaillé sur la mise en place d'un module de génération morphologique (flexionnelle) et d'un module de génération syntaxique. J'ai assuré l'encadrement de ce stage en collaboration avec mes collègues de l'équipe.

### Traitement automatique de la langue arabe

J'ai participé activement au lancement de l'activité de traitement automatique de la langue arabe à France Télécom Orange Labs. Parmi les actions menées :

- Prospection pour l'acquisition d'un lexique arabe (étude approfondie et exhaustive de l'offre au niveau mondial, identification d'un fournisseur, discussions, rencontre et acquisition du lexique).
- Recrutement et encadrement de deux Post-Doc spécialisés en traitement automatique de la langue arabe : la rareté de ce type de profil a été à l'origine d'une grande difficulté pour trouver de bons candidats.
- Construction d'une table de translittération des caractères arabes. Nous avons même étudié la possibilité de proposer cette table aux organismes de normalisation.

---

<sup>1</sup> Le terme "traduction automatique" dans ce document désigne toutes les variantes de la traduction associant l'homme et la machine (traduction automatique, traduction assistée par ordinateur, mémoires de traduction, etc.).

## **Internationalisation et localisation des applications et des services**

Cette activité concerne les applications et les services de France Télécom qui sont destinés à un marché international. La dimension internationale, qui est de plus en plus prononcée au sein du groupe France Télécom Orange, se traduit nécessairement par des besoins en localisation des applications qui sont destinées au marché international. L'objectif de cette activité est de mettre à la disposition des unités du groupe, qui sont des développeurs d'applications ou des concepteurs de services, une plateforme experte pour la localisation de leurs produits dans les langues du marché ciblé.

### **Encadrement de la recherche**

- Co-encadrement avec le Professeur Jean Véronis d'une thèse en informatique sur le thème de la traduction automatique, soutenue en 2006.
- Encadrement d'un postdoc sur le traitement automatique de la langue arabe, 2000/2001.
- Encadrement d'un postdoc sur le traitement automatique de la langue arabe, 2005/2006.
- Encadrement d'un postdoc sur la recherche d'information cross-lingue et multimédia, 2010/2011.
- Encadrement d'un ingénieur apprenti en informatique sur le thème de la génération en langage naturel, 2001 et 2002 (14 mois).

## **5. Activités de Conseil au sein du groupe ORANGE**

### **Conseil pour le projet JEI (Jordan Education Initiative)**

Ce projet est piloté par la Direction des Programmes de Transformation et Animation des Réseaux Formation de France Télécom, en partenariat avec Jordan Telecom et d'autres partenaires, pour le compte du Ministère de l'Education Jordanien. Il a pour but de mettre en place une plateforme pour le e-teachning de la langue arabe (aide à l'enseignement de la langue). J'ai été sollicité pour apporter mon expertise sur l'arabisation de l'application informatique.

### **Conseil dans le cadre de la mise en place du laboratoire ILAB du Caire**

Dans le cadre du projet de mise en place du laboratoire ILAB-Caire (démarré en janvier 2008), j'ai été régulièrement sollicité pour participer aux réunions de brainstorming ou pour aider dans la définition des thèmes de recherche du laboratoire en lien avec le traitement automatique de la langue arabe.

### **Conseils pour les unités d'Affaires de France Télécom**

J'ai mené plusieurs actions de conseil pour les unités d'affaire de France Télécom, notamment dans le domaine de la traduction automatique. A titre d'exemples :

- Conseils auprès de la Division Réseaux, Opérateurs et Système d'Information (ROSI) pour l'acquisition et le déploiement d'un logiciel de traduction automatique au sein du groupe France Télécom, février-juin 2006.

- Conseils auprès de la Direction Intégration et Services (DIS) pour l'acquisition et le déploiement d'un logiciel d'aide à la traduction basé sur la technologie des mémoires de traduction, février-juin 2006.
- Contribution au projet STRATINO, membre d'un groupe de travail sur le MTT16, 2003.
- Etc.

### **Projet de localisation des services du portail Voila en différentes langues**

Ce projet qui a été piloté par DMM/PAM en 2001 à la demande de FTMMS, a sollicité mon expertise dans le domaine de la localisation multilingue des applications. J'ai participé à différentes étapes de ce projet et j'ai mis à la disposition des partenaires un certain nombre de ressources liées à la localisation et à l'édition des textes multilingues, ainsi qu'à l'état des techniques les plus récentes dans ce domaine.

### **Etudes des opportunités des projets collaboratifs nationaux ou européens**

De façon générale, compte tenu de mon rôle d'expert et de ma connaissance des réseaux collaboratifs nationaux et internationaux, je suis souvent amené à étudier les opportunités de participation à des consortiums ou à des propositions de projets collaboratifs, aussi bien à l'échelle nationale qu'à l'échelle européenne.

## **6. Diffusion des résultats des projets au sein de France Télécom**

De manière générale, inspiré par mon rôle d'expert, j'ai systématiquement le réflexe de diffuser les résultats des projets auxquels je participe auprès des collègues, des projets ou des entités que ces résultats peuvent intéresser.

## **7. Activités de formation interne et externe à FTR&D**

### **Réunion d'Animation Technique à France Télécom**

Participation aux formations internes à France Télécom par l'intermédiaire des sessions de Réunions d'Animations Techniques (RAT). J'ai contribué aux RAT de 2001 sur le thème du traitement automatique du langage naturel. J'ai été sollicité pour intervenir sur le thème de la traduction automatique.

### **Enseignement universitaire à l'ENSTB**

Participation aux enseignements de fin de cycle d'ingénieurs de l'ENSTB (ENST Bretagne à Brest). Le thème du cours porte sur la traduction automatique, février 2006.

### **Enseignement à l'université de Marne La Vallée**

Participation aux enseignements du DEA des industries des langues, organisé par les universités Marne La Vallée, Paris VII, Genève, INSTN, CNAM et Aupelf-Uref. Le thème des enseignements porte sur la traduction automatique, théorie et réalisation de mini-systèmes de traduction, 1996 à 2000.

## **8. Activités de valorisation des travaux de R&D au sein et en dehors de France Télécom**

### **Déploiement de l'abrégeur sur le portail intranet France**

Coordination du déploiement du logiciel de résumé automatique et d'extraction des mots-clés sur le portail intranet France du groupe (décembre 2007). Le déploiement de l'abrégeur sur le portail international du groupe (en cours de restructuration) est également envisagé.

### **Déploiement de logiciels sur la plateforme du bureau-du-futur**

Coordination du déploiement du logiciel de résumé automatique et du service sms2fixe sur la plateforme du bureau-du-futur destinée aux Jardins de l'innovation (décembre 2007).

### **Protection des technologies**

Ce volet consiste à piloter les procédures de protection et de licenciement des technologies développées dans notre équipe. Ce volet comprend plusieurs facettes :

- Dépôt de brevets.
- Dépôt de logiciels.
- Cession de licences.

### **Construction d'une offre globale en TALN**

Ce volet consiste à contribuer à la construction d'une offre globale des technologies développées. Cette offre se décline en une offre du type projet ou conseil spécifique et une offre du type produit ou licence.

### **Conseil et aide à la mise en place d'une plateforme de démonstration**

Un autre rôle dans cette activité consiste à aider à la mise en place d'une plateforme de démonstration des outils qui soit accessible sur Intranet et Internet.

### **Activité de veille et étude de l'état du marché et de la concurrence**

Cette activité de veille et d'étude actualisée de l'état du marché et de la concurrence revêt un caractère important dans la valorisation des travaux de R&D. Bien entendu, cette activité requiert une connaissance approfondie des domaines concernés et une connaissance avertie de l'état de la technique.

### **Valorisation de la R&D auprès des industriels**

La valorisation des activités de R&D auprès des organismes industriels requiert une bonne connaissance des milieux industriels et une bonne connaissance de l'état des technologies existantes. Cette action est menée dans le cadre de conférences invitées à des salons ou à des manifestations de communication sur les technologies de TALN.

A titre d'exemple, j'ai présenté des conférences ou participé à des salons tels que :

- SMSI, Sommet Mondial sur la Société de l'Information, Tunis, novembre 2005.
- LANGTECH, salon des technologies du langage écrit et parlé, Paris, 2003.
- TELMI 2002 et 2006, congrès sur les outils de veille, organisé par la CCI du Nord à Lille.
- Etc.

## **Valorisation de l'abrégeur auprès de THALES**

Parmi les actions les plus concrètes de valorisation externe, j'ai coordonné et piloté, pendant près de deux ans, une action lobbying auprès de THALES qui s'est concrétisée par l'acquisition, par THLES Communications, d'une licence de l'abrégeur développé à France Télécom R&D (versions française et anglaise).

THALES a intégré l'abrégeur dans une plateforme de veille qu'elle envisageait de commercialiser. Chaque commercialisation de la plateforme de veille de THALES devra bénéficier à France Télécom par une commercialisation de licences de l'abrégeur.

Des contacts sont maintenus avec THALES qui s'intéresse actuellement aux outils développés par France Télécom autour de la langue arabe (abrégeur, lemmatiseur, etc.).

## **Valorisation des technologies linguistiques auprès de Wanadoo Jordanie**

J'ai piloté des contacts étroits avec WANADOO JORDANIE (démonstrations, réunions, etc.). Ces échanges portaient sur un certain nombre d'outils développés à France Télécom R&D :

- Abrégeur arabe et anglais,
- Interprétation des requêtes sur les Pages Jaunes.

## **9. Représentation de France Télécom dans des manifestations internationales**

### **Membre observateur dans des groupes de normalisation**

Je participe, en qualité de membre observateur, aux rencontres des groupes de travail et de réflexion sur la normalisation des ressources linguistiques. Ces groupes sont réunis autour de l'instance ISO TC37/SC4 et des actions de normalisation menées notamment par l'INRIA à travers les projets SYNTAX et RNIL.

### **Atelier Semantic Web Technologies à Luxembourg**

Dans l'action clé III du programme de travail IST 2001, l'unité 5 de la DG INFSO, en charge de ce thème, a organisé un atelier de deux jours pour identifier et discuter des actions de R&D prioritaires, pour des applications du Web sémantique. Cet atelier a eu lieu à Luxembourg au mois de novembre 2000. Ma participation a permis de se rendre compte des domaines d'investigation correspondant à des axes d'études possibles.

### **Séminaire du Centre Français du Commerce Extérieur**

Le Centre Français du Commerce Extérieur a organisé à Paris, le 9 mars 2001, un séminaire intitulé "Internet et informatique : quelles opportunités de marché aujourd'hui sur le Maroc, la Tunisie et le Liban ?". Ma participation a permis de se rendre compte de l'état du marché de l'internet dans ces pays. Cela nous permet de mieux appréhender nos travaux de R&D en traitement automatique de la langue arabe.

### **Autres manifestations**

De façon générale, mon rôle en qualité d'expert, m'amène de plus en plus à des actions exploratoires similaires dans le cadre de mon domaine d'expertise.

## 10. Publications scientifiques

### Publications internationales

#### Reuves internationales avec comité de lecture

- 1997 Boualem A.M., Harié S., "MtScript: a multilingual text editor", Computers and the Humanities, Volume 31, No. 2, Kluwer Academic Publishers, 135-151, 1997.
- 1998 Boualem A.M., "Edition des textes multilingues", Revue de l'Information Scientifique et Technique, Centre d'Etudes et de Recherche sur l'Information Scientifique et Technique, Alger, ISSN 1111-0015, Vol. 8, No 1, 1998.
- 2000 Boualem A.M., Book Reviews, Translation Engines: techniques for Machine Translation, Arturo Trujillo, Springer-Verlag, Applied Computing, Heidelberg, ISBN 1-85233-057-0, 1999, Speech Communication Journal, Issue 32(3), October 2000.
- 2001 BabelWeb Project Partners, BabelWeb Multilingual Guidelines, best practices architecture and procedures for international web sites, Multilingual Computing & Technology, No 45 Volume 13 Issue 1, 2001.
- 2008 J.Heinecke, G.Smits, C.Chardenon, E.G. De Neef, E.Maillebua, M.Boualem, "TiLT : une plateforme pour le traitement automatique du langage naturel", revue TAL, Volume 49, no 2/2008.

#### Reuves spéciales à audience internationale

- 2001 BabelWeb Project Partners, "Multilingual web sites, Best practice guidelines and architectures", Revue Eurescom, January 2001.
- 2002 MUST Project Partners, "MUST – multimodal and multilingual services for small mobile terminals", Revue Eurescom, May 2002.

#### Congrès internationaux avec comité de lecture et publication d'actes

- 1989 Boualem A.M., Lin C., "Ideogram-based processing system", ISCIS Conference, Cesme, Turquie, Octobre 1989.
- 1989 Boualem A.M., Lin C., Mlouka M., "Une approche d'aide à la saisie des caractères chinois", IASTED, Applied informatics, Grindelwald, Suisse, Février 1989.
- 1989 Boualem A.M., Amara F., Lin C., Franjaud J.C., Mlouka M., "Chinese and Japanese characters input", WECC'89, Western Educational Computing Conference, Palo Alto, USA, Novembre 1989.
- 1993 Boualem A.M, Mlouka M., "Aspects de la traduction automatique", Universal Knowledge Tools Conference, Toronto, Canada, Juin 1993.
- 1993 Boualem A.M., Mlouka M. "ML-TASC: système de traduction automatique multilingue", SS'93 Symposium, High Performance Computing, Calgary, Canada, Juin 1993.

- 1995 Boualem A.M., "Multilingual text editing", SNLP'95 Symposium on Natural Language Processing, Bangkok, Thaïlande, Août 1995.
- 1995 Boualem A.M., "Arabic language processing", SNLP'95 Symposium on Natural Language Processing, Bangkok, Thaïlande, Août 1995.
- 1996 Boualem A.M., Harié S., Véronis J., "MtScript: un éditeur de textes multilingues", CARI'96, 3ème Colloque Africain sur la Recherche en Informatique, Organisé par l'INRIA, Libreville (Gabon), Octobre 1996, 262-271.
- 1999 Boualem A.M., Leisher M., "Encoding script-specific writing rules based on the Unicode character set", 14th International Unicode Conference, Boston, USA, March 22-25, 1999.
- 1999 Boualem A.M., Zajac R., "Unicode-based Arabic text", ATLAS'99, Arabic Translation and Localisation Symposium, Tunis, May 26-28, 1999.
- 1999 Boualem A.M., Leisher M., Ponsford D., "Concordancing on underlying forms: a case study in Arabic", ATLAS'99, Tunis, May 26-28, 1999.
- 1999 Boualem A.M., "Vers un analyseur de contexte multilingue universel", CII'99, Conférence Internationale en Informatique, Annaba, Algérie, Novembre 21-23, 1999.
- 1999 Zajac R., Boualem A.M., Amtrup J., "Specification and implementation of input methods using finite-state transducers", 14th International Unicode Conference, Boston, Massachusetts, March 22-25, 1999.
- 2001 Boualem A.M., & al. "Recherche d'information en langue arabe", ACL/EACL 2001 Workshop, Arabic Language Processing: Status and Prospects, Toulouse, July 6, 2001.
- 2001 Malek Boualem (France Telecom R&D), Régina Sneifer (Wanadoo), "Hahooa Arabic Web Directory & Natural Language Processing for Arabic Information Retrieval", ACL 2001, Workshop on Arabic Language Processing, Toulouse, France, Juillet 2001.
- 2002 Boualem A.M. & partenaires MUST, "*Eurescom MUST project, Multimodal, multilingual information services for small mobile terminals*", Conférence TALN 2002, workshop NLP techniques for speech analysis, Nancy, France, 24-27 juin 2002.
- 2002 Luis Almeida, Ingunn Amdal, Nuno Beires, Malek Boualem, Lou Boves, Els den Os, Pascal Filoche, Rui Gomes, Jan Eikeset Knudsen, Knut Kvale, John Rugelbak, Claude Tallec, Narada Warakagoda, "Implementing and evaluating a multimodal and multilingual tourist guide", International CLASS Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Copenhagen, Denmark, 28-29 June 2002.
- 2002 Emilie Guimier De Neef, Malek Boualem, Christine Chardenon, Pascal Filoche, Jérôme Vinesse, "Natural Language processing software tools and linguistic data developed by France telecom R&D", IEMCT - Indo European Conference on Multilingual Technologies, Pune, Inde, 2002.
- 2002 Luis Almeida, Ingunn Amdal, Nuno Beires, Malek Boualem, Lou Boves, Els den Os, Pascal Filoche, Rui Gomes, Jan Eikeset Knudsen, Knut Kvale, John Rugelbak, Claude Tallec, Narada Warakagoda, "User-friendly Multimodal Services - Going the Multimodal

- route: making and evaluating a multimodal tourist guide service", Eurescom Summit 2002, Powerful Networks for Profitable Services, Heidelberg, 21-24 October 2002.
- 2003 Alain Léger, Malek Boualem, "Technology Survey on Knowledge Based Multilinguality", 2nd International Conference on Universal Access in Human-Computer Interaction 5, Crete, Greece, Juin 2003.
- 2008 Abbes R, Boualem A.M, Hassoun M., "Où en est la recherche d'information en langue arabe ?", Conférence SIIE 2008, Tunisie, 14-16 février 2008.
- 2008 Abbes R, Boualem A.M., "Dissymétrie entre l'indexation et la recherche d'information en langue arabe", Conférence TALN 2008, Avignon, France, 9-13 juin 2008.
- 2008 Boualem A.M., Abbes R, "Arabic language information retrieval : some thoughts", LREC 2008 workshop on HLT & NLP within the Arabic world, Arabic Language and local languages processing, Status Updates and Prospects, Marrakech, Morocco, 26 May - 1 June 2008.
- 2010 Benoît Gaillard, Jean-Léon Bouraoui, Emilie Guimier de Neef, Malek Boualem, "Expansion de requêtes pour l'optimisation de la recherche d'information multilingue basée sur la traduction des contenus", CORIA 2010, mars 2010, Sousse, Tunisie.
- 2010 Benoît Gaillard, Malek Boualem, Olivier Collin, "Query translation using Wikipedia-based resources for analysis and disambiguation", 14th Annual Conference of the European Association for Machine Translation, May 2010, Saint-Raphaël, France.
- 2010 Benoît Gaillard, Jean-Léon Bouraoui, Emilie Guimier de Neef, Malek Boualem, "Query Expansion for Cross Language Information Retrieval Improvement", RCIS 2011 Research Challenges in Information Science, May 2010, Nice, France
- 2010 Jean-Léon Bouraoui, Benoît Gaillard, Emilie Guimier de Neef, Malek Boualem, "Annotation of linguistic phenomena and topics in query logs", May 2010, Coruña, Spain.
- 2010 Benoît Gaillard, Olivier Collin, Malek Boualem, "Traduction de requêtes à l'aide de Wikipédia pour la recherche d'information multilingue", TALN 2010, Montréal, juillet 2010.
- 2011 Malek Boualem, Jean-Philippe Cabanal, Nabil Bouzerna, "Cross-Language Information Retrieval in QUAERO-MSSE Project", EAMT 2011, Leuven, Belgium, Mai 2011.
- 2012 Malek Boualem, "Integration of speech translation to enrich the cross-language information retrieval of the Quaero/MSSE prototype", QUAERO CTC/CORPUS workshop, Grenoble, France, Avril 2012.
- 2014 Malek Boualem, Johannes Heinecke, Ramzi Abbes, "Aspects de la recherche d'information en langue arabe", CITALA 2014, Oujda, Maroc, 26-27 Novembre 2014.

### **Congrès nationaux avec comité de lecture et publication d'actes**

- 1995 Boualem A.M., "Multilingual text editing and Arabic language processing", AI'95, Conférence en Génie linguistique, Montpellier, France, Juin 1995.

2001 Boualem A.M., "Traitement automatique de la langue arabe pour la recherche d'information", Séminaire de l'institut du monde arabe, Paris, France, juin 2001.

### **Publications dans le cadre de projets internationaux**

2000 Boualem A.M., "Overview of machine translation techniques", EURESCOM, P923, 2000.

1999 Boualem A.M., "Multilingual text processing difficulties", EURESCOM, P923, 1999.

2000 Boualem A.M., Publication d'une brochure du projet Babelweb, EURESCOM, P923, 2000.

### **Publications internes de rapports techniques**

1987 Boualem A.M., "Un analyseur de contexte multilingue", Rapport technique, INRIA.

1990 Boualem A.M., "Le terminal multilingue", Rapport de recherche, INRIA Sophia Antipolis.

2001 Boualem A.M., "Traduction automatique et traduction assistée par ordinateur", Publications des Réunions d'Animation Technique, RAT, France Télécom, 2001.

### **Publications COMET à France Télécom**

2007 Boualem A.M., "De TRANSAT à OMTS : service de traduction pour les clients d'Orange en situation de mobilité", WebR&D en direct des labos, 17 décembre 2007.

2008 Fluck F, Boualem A.M, "TiLT-abrégeur, une innovation Orange Labs déployée sur le portail Intranet France du Groupe !", WebR&D A LA UNE, 15 janvier 2008.

2002 Boualem A.M, Barbara Huc, "Offre logiciels linguistiques du Top 15", COMET, la UNE du 28 mars 2002.

2004 Boualem A.M, "Projets de recherche en perspectives en collaboration avec le laboratoire de France Télécom R&D à Pékin", COMET du 13 octobre 2004.

2005 G.Le Calvez, M. Boualem, Article COMET au sujet du déploiement de l'abrégeur sur NetR&D, COMET, 2005.

2005 G.Le Calvez, M. Boualem, Article COMET au sujet de la présentation de l'abrégeur arabe au SMSI 2005, COMET, novembre 2005.

2002 Anne Boutroux, Olivier Daridan, Karel Mittig, Jean-François Ravier, Jérôme Vinesse, Malek Boualem, "L'architecture et les services de personnalisation Web en cœur de réseaux", septembre 2002.

### **Publications techniques sur le Web**

1996 Boualem A.M., Véronis J., Harié S., Documentation technique détaillée du logiciel MtScript (en anglais) comprenant le manuel d'utilisation et les packages de téléchargement (équivalent 30 pages) à l'URL: <http://www.lpl.univ-aix.fr/projects/multext/MtScript/>