

Vague C (2013-2017)

Unité de recherche : dossier unique

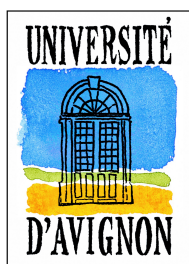
PROJET SCIENTIFIQUE

(Partie I)

LABORATOIRE INFORMATIQUE D'AVIGNON

Contrat Quinquennal 2013-2017

UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE



UNIVERSITÉ D'AVIGNON
ET DES PAYS DE VAUCLUSE



Sommaire

1. Introduction.....	6
2. Auto-évaluation : forces, faiblesses, risques et opportunités.....	6
3. Objectifs.....	8
4. Orientations scientifiques.....	8
4.1 Introduction.....	8
4.2 Culture et sociétés numériques : enjeux sociétaux	8
4.3 Culture et sociétés numériques : verrous scientifiques.....	9
4.4 Stratégie et modalité de mise en œuvre.....	9
4.5 Intégration au projet d'établissement.....	10
4.6 La Structure fédérative de recherche Agorantic.....	10
4.7 Projets transversaux.....	10
4.8 Structure thématique et politique de recrutement.....	10
5. Structure administrative et gouvernance	11
5.1 Le CERL.....	11
5.2 Gouvernance du LIA.....	11
6. Formation doctorale.....	12
7. Support et valorisation.....	12
8. Contrats de recherche.....	13
9. Politique financière.....	13
10. Liens avec les EPST.....	13
11. Appels à projets liés aux investissements d'avenir	13
12. Animation scientifique de l'unité.....	14
13. Projet scientifique par thématique.....	15
13.1 Projet de la thématique Langage.....	16
13.1.1 Orientations stratégiques.....	16
13.1.2 Orientations scientifiques.....	16
13.1.3 Analyse des opportunités et risques.....	21
13.2 Projet de la thématique Recherche Opérationnelle et Optimisation	23
13.2.1 Orientations stratégiques.....	23
13.2.2 Orientations scientifiques.....	23
13.2.3 Problèmes	23
13.2.4 Analyse des opportunités et risques.....	24
13.3 Projet de la thématique Réseaux.....	25
13.3.1 Orientations stratégiques.....	25
13.3.2 Orientations scientifiques.....	25
13.3.3 Analyse des opportunités et risques.....	28

Résumé

Le dynamisme du LIA s'est traduit par une forte progression des indicateurs objectifs de sa productivité scientifique (doublement des publications, du budget, organisations régulières de conférences de portées nationales et internationales). Ce dynamisme lui donne maintenant une visibilité consolidée par des collaborations internationales et confirmée par l'insertion professionnelle de ses docteurs et HDR (8 docteurs MCF ou CR sur les 22 ayant soutenus en 2006/2007, 6 des 8 HDR soutenues sont maintenant PR ou DR).

Cependant, cette croissance a été obtenue à effectif permanent quasi-constant (20 EC en 2005 pour 21 en 2011 plus 1 PR émérite, pas de progression du nombre de personnels IATOS). Il est clair que la charge qui pèse sur l'ensemble des personnels est maintenant très lourde ; ce point constitue une faiblesse de l'unité et pourrait devenir un handicap dans l'avenir. La stagnation de l'effectif permanent a conduit aussi l'unité à maintenir l'effectif des équipes récentes (RO et Réseaux) à la limite d'une taille critique qui leur permettrait de se structurer et de se stabiliser.

Par ailleurs, l'activité s'est essentiellement développée à l'intérieur de ses domaines de spécialité, sans que les collaborations inter-thématiques ne se renforcent significativement. Enfin, le projet de l'Université s'est construit autour de 2 axes identitaires, AgroSciences et Culture et patrimoine, qui l'ancrent dans le territoire mais dans lesquels la position du LIA n'était pas claire ces dernières années.

Après cette période de croissance rapide, l'objectif de l'unité est maintenant de consolider ses acquis et de se structurer à la fois aux niveaux organisationnel et scientifique :

- **organisationnel** : nous continuerons à œuvrer au développement du CERI, qui a été créé en avril 2010 et qui regroupe le département et le laboratoire d'informatique. Son objectif est de **développer les synergies entre enseignement et recherche** pour répondre aux problèmes d'administration et de support, ainsi que pour améliorer la qualité du recrutement étudiant et des formations, y compris au niveau doctoral. Le CERI s'est engagé dans une démarche qualité (Master en Alternance certifié ISO 9001 en 2011) que nous étendrons au suivi des doctorants.
- **consolidation des bases scientifiques** : nous **développerons notre expertise et notre visibilité internationale dans nos spécialités**, en soutenant les thématiques fragilisées par leur taille réduite, en organisant des manifestations nationales et internationales, en confirmant notre rôle de concepteurs et de diffuseurs d'environnements logiciels pour la communauté scientifique et en maintenant notre présence, forte, dans les sociétés savantes de nos domaines de recherche.
- **ouverture pluridisciplinaire** : le dynamisme culturel régional et la forte activité de l'UAPV dans les domaines de la culture et du patrimoine sont, pour nous, une occasion de jouer un rôle moteur dans le projet d'établissement. Au-delà de ce contexte local, le développement des réseaux, de l'Internet et des grandes bases multimédias pose des problèmes nouveaux liés à la diversité de fond et de forme des contenus, aux infrastructures, à la complexité et au passage à l'échelle. La configuration thématique du laboratoire nous place dans une situation privilégiée pour répondre à ces différentes questions, à la fois sur leurs aspects fondamentaux et applicatifs. Nous souhaitons orienter notre politique scientifique vers la **Culture et les sociétés numériques** ; ce projet devrait nous permettre de nous rassembler sous un bannière unique, visible, basée sur notre socle de compétences actuel et sur une ouverture vers les SHS.

La mise en œuvre de ce projet reposera sur une politique active de recherche du soutien des agences de financement nationaux et européens, par le financement de projets internes et par une intégration de cet objectif dans notre politique de recrutement.

Elle s'appuiera la **Structure Fédérative de Recherche** Agorantic, créée en 2011, et qui devrait nous permettre de susciter et de structurer des collaborations pluridisciplinaires et d'obtenir le soutien des EPST.

Executive summary

The dynamism of the LIA has led to a sharp improvement of its scientific quality indexes (doubling of the publication number, doubling of the budget, organization of national and international conferences). This dynamism has increased its visibility, which was consolidated through international collaborations.

This visibility is also confirmed by the professional integration of its doctors and HDR (8 of the 22 doctors that defended their PhD in 2006/2007 got permanent positions in public research laboratories, and 6 of the 8 HDR are now Professors).

However, these results were obtained with a poor growth of permanent staff : the lab included 20 researchers in 2005, and 21 in 2011 (plus 1 PR Emeritus) , with the number of technical and administrative employees (IATOS) remaining unchanged. It is clear that the workload of all the teams is now very high. This is a weakness of the lab that could become a handicap in the future.

The poor growth of the permanent researcher workforce also led the unit to maintain the recent teams (RO and Réseaux) under the critical size that could enable them to structure and to stabilize themselves. Moreover, the research was mainly conducted within the research domains, without any reinforcement of cross-domain collaborations. Finally, the project of the University was built around two axes : Culture and history on the one hand, AgroSciences on the other hand, which matches regional identity but in which the position of the LIA was not clear.

After this period of rapid growth, the objective of the lab is now to consolidate its achievements and to reinforce its organizational and scientific structures :

- *organizational: we will continue to work towards the development of CERI, which was established in April 2010 and which includes the teaching and research activities. Its objective is to develop synergies between teaching and research, in order to address the problems of administration and support understaffing, and to improve the quality of student recruitment and training, including at doctoral level. The Master was ISO 9001 certified in 2011 and we plan to extend the Quality management process to the supervision of PhD students.*
- *consolidation of scientific fundamentals: we will develop our expertise and our international visibility in our research fields, by supporting teams weakened by their small size, by organizing national and international conferences, by designing and distributing software environments to the scientific community, and by maintaining our presence in the societies of our research areas.*
- *opening to cross-domain collaborations : regional and cultural dynamism combined to the excellence of UAPV research in the field of humanities are, for us, a great opportunity to play a leading role in the University project. Beyond these local opportunities, the development of the Internet suggests new needs and new scientific problems, related to the diversity of contents, to the infrastructure supporting data exchange, and to complexity and scalability.*
- *The skills of the laboratory put it in a unique position to address these issues, both in their fundamental and technological aspects. Our project consists in grouping our research activities under a clear and unified banner of "Culture and digital societies", which is based on our current skills and which opens new perspectives of collaboration with UAPV digital humanities teams.*

The implementation of this project will be based on an aggressive policy of project proposals to funding agencies (at national and European levels) and by funding internal projects that fit our scientific goals. Moreover, our recruitment policy will take these scientific objectives into account.

Last, we will use federative research structure Agorantic, founded in 2011, to initiate and structure multidisciplinary collaborations and to obtain the support of EPST.

Projet du LIA pour la période 2013- 2017

1. Introduction

Le bilan que nous avons présenté montre, sur un certain nombre d'indicateurs objectifs, la progression de la productivité scientifique et du rayonnement du LIA. En parallèle, l'unité connaît des changements structurels liés aux efforts que nous avons réalisés pour améliorer notre cohésion interne et notre visibilité internationale. Ces efforts ont notamment abouti, en avril 2010, à la création d'un Centre d'Enseignement et de Recherche en Informatique (CERI) qui associe le laboratoire et le département d'informatique. Par ailleurs, le laboratoire s'est engagé dans la création d'une Structure Fédérative de Recherche sur les *sciences et technologies des cultures et sociétés numériques*, créée en juin 2011 par l'UAPV.

Ce document présente les aspects structurels et scientifiques de notre projet. Il s'appuie sur un bilan critique et identifie les axes que nous développerons dans les années à venir et les moyens que nous mettrons en œuvre pour atteindre nos objectifs.

2. Auto-évaluation : forces, faiblesses, risques et opportunités

Le LIA a développé ses **collaborations nationales et internationales** ces dernières années et nous poursuivons nos efforts pour conforter notre positionnement international. Ceci s'est traduit par une présence forte dans les sociétés savantes (IEEE SLTC, ISCA,...), par des participations régulières aux grandes campagnes d'évaluations internationales (NIST SRE, NIST LRE, TREC, ROADEF...) et par l'organisation de conférences (JEP-TALN 2008, WiOpt 2010, EACL 2012 sera organisée à Avignon, InterSpeech 2013 en co-organisation). De nombreuses collaborations internationales ont donné lieu à des publications communes et à l'accueil de 28 chercheurs étrangers sur des périodes de plus de 3 mois.

Cette ouverture à l'international doit cependant s'appuyer sur une base locale forte et cohérente. Les collaborations entre thématiques sont restées assez ponctuelles. Notre taille ne nous permet pas d'accentuer le **fractionnement thématique** de l'unité et la **cohésion du laboratoire** doit être renforcée. Ce renforcement doit notamment aider les thématiques les plus récentes à développer des liens qui les aideraient à se stabiliser.

Les moyens obtenus et la production scientifique du LIA ont sensiblement progressé au cours de la période précédente. Le budget moyen non consolidé est passé d'environ 350K€ à plus de 750K€. L'essentiel de cette augmentation est dû à la croissance des moyens contractuels issus de collaborations industrielles ou académiques, réalisées dans le cadre de projets de recherche financés. C'est, nous semble-t-il, le signe d'un dynamisme fort et d'une reconnaissance de notre savoir-faire. Ces éléments accompagnent la progression qualitative et quantitative de notre production scientifique : comme cela a été évoqué, le nombre de publications en revues référencées et conférences internationales par année a doublé par rapport au précédent contrat (95 en moyenne annuelle sur [2006,2010] contre 45 par an sur [2002,2005], en ACL+C-ACTi+DO+OS), alors que notre effectif permanent est resté globalement stable (22 enseignants chercheurs en 2011 contre en 2005). Cette forte progression du **niveau de publication et le dynamisme de nos activités contractuelles** nous semblent être des points positifs du bilan.

Cependant, l'augmentation des ressources contractuelles s'est accompagnée d'une croissance plus contenue des moyens récurrents, dont la part relative est réduite : elle représentera, en 2011, moins de 10% du budget total (non consolidé) du laboratoire. Cette modification de l'équilibre entre ressources affectées et récurrentes et la structuration en projets de nos activités ont une influence sur notre organisation et notre politique scientifique :

- l'implication dans les projets financés par l'ANR ou l'UE pourrait nous faire courir le risque d'un pilotage scientifique de l'unité par les agences de financement. Il est essentiel pour nous de garder la **maîtrise de notre politique scientifique** à moyen et long termes.
- le développement des projets financés demande à ce que la structure hôte assure un certain nombre de charges nécessaires à leur réalisation et à leur valorisation. Malgré le soutien que l'UAPV nous accorde, il s'agit d'une université notoirement sous-dotée en personnel IATOS – ce qui impacte naturellement le LIA. De façon plus générale, l'engagement des chercheurs dans des activités de support et d'administration nous fait courir le **risque d'une érosion du potentiel de**

recherche de l'unité et d'une sous-valorisation de nos activités. Le problème de la valorisation des plate-formes logicielles développées au LIA est emblématique de ce risque : certaines d'entre elles sont utilisées par les plus grands laboratoires mondiaux mais, en l'absence de support technique suffisant, l'investissement que nous avons réalisé sera, en partie, perdu. Les solutions que nous proposons, pour prévenir ce risque et **assurer le support et la valorisation de nos activités de recherche**, sont décrites dans la section 5.1 de ce document.

Forces	Faiblesses
Publications : niveau et progression	Charges administratives, manque de support
Dynamisme des activités contractuelles	Interactions thématiques insuffisantes
Reconnaissance nationale/internationale	Taille critique/structuration des thématiques récentes
Animation scientifique (conférences, sociétés savantes)	Difficulté de recrutement des doctorants
Insertion doctorants/MCF-HDR	
Opportunités	Risques
Axe identitaire de l'UAPV <i>Culture et Patrimoine</i> , ancré dans le territoire, évoluant vers les STIC.	Érosion du potentiel de recherche lié au manque de support et à la stagnation du nombre de permanents
Enjeux scientifiques/socio-économiques des Sciences du Web	Politique scientifique opportuniste (guidée par les contrats et les applications)
Association pluridisciplinaire STIC/SHS	Détournement des domaines d'expertise actuels
Dynamisme local (UAPV et Pôle de compétitivité SCS)	

Tableau 1: Auto-évaluation des forces, faiblesses, risque et opportunités.

Outre la charge supplémentaire liée à la croissance de nos activités, nous sommes impliqués dans l'administration de l'enseignement avec un dynamisme qui a porté ses fruits : le Master en alternance du CERI a été **certifié ISO 9001** en Juin 2011 et nous projetons d'étendre le périmètre de la certification à la licence et au suivi des doctorants du LIA ; une spécialité du Master, sur le thème *Informatique et Santé*, ouvrira en 2012 ; enfin, le **sous-encadrement** de nos formations devrait encore s'aggraver si la croissance des effectifs enregistrée en 2011 se confirme. L'**augmentation de l'effectif permanent du laboratoire** est une condition nécessaire au maintien, à long terme, du dynamisme de l'unité.

Depuis 2006, 8 HDR et 35 thèses ont été soutenues. L'**insertion de nos docteurs** est un des points que nous jugeons positif : 8 docteurs ont obtenu un poste de MCF ou de chercheur permanent en France et 6 autres un poste de chercheur statutaire à l'étranger sur les 22 ayant soutenu en 2006 et 2007. De plus, 6 MCF ont obtenu un poste de professeur depuis 2006. Cependant, les effectifs des filières recherche ont diminué ces dernières années alors que le **recrutement des doctorants** est un point critique pour le développement du laboratoire. Trouver des solutions à ce problème sera une de nos priorités de la prochaine période contractuelle.

L'Université affiche deux axes identitaires autour des Agrosociétés d'un côté, et de la Culture et du Patrimoine de l'autre, dans lesquels la position du LIA n'était pas claire. L'évolution vers de ce dernier axe vers les *cultures et sociétés numériques* converge avec le développement du Web et s'appuie sur le dynamisme culturel de la région et sur son patrimoine historique. Cette convergence nous offre l'opportunité de regrouper nos activités sous le cadre, très large, de l'étude des cultures et sociétés numériques. Ce thème fédère la plupart des sujets de recherche actuels des trois thématiques du laboratoire. Il ouvre des perspectives de collaboration avec d'autres laboratoires reconnus de l'Université, notamment en sciences sociales et humaines. Cette association originale d'équipes solides et complémentaires sur des sujets porteurs est une réelle opportunité pour nous d'afficher un thème identitaire fort, qui s'intègre à la politique globale de l'Université. La note d'orientation stratégique évoque cette possibilité.

L'affirmation d'un tel **axe identitaire** pourrait néanmoins porter le risque d'une marginalisation des sujets de recherche qui s'y inscrivent moins directement ou d'un éloignement de nos domaines d'expertise.

Notre projet précise nos objectifs, la stratégie que nous mettrons en œuvre pour les atteindre et pour prévenir les risques et saisir les opportunités que nous venons d'évoquer.

La tableau 1 présente une synthèse des points forts, des faiblesses, des risques et opportunités que nous avons présentés.

3. Objectifs

La politique du LIA pour la prochaine période quinquennale consistera à prévenir les risques et à tirer parti des opportunités évoquées dans notre bilan, ainsi qu'à développer les points qui nous semblent essentiels pour l'avenir de l'unité. Cette orientation que nous voulons donner à notre politique scientifique est détaillée dans les sections suivantes.

La politique de publication visant le meilleur niveau sera poursuivie et renforcée. Cette **productivité scientifique** devrait profiter des **collaborations internationales** dont nous voulons soutenir la progression.

Notre premier objectif est de **développer notre expertise et notre visibilité internationale dans nos spécialités** scientifiques, notamment en organisant des manifestations scientifiques internationales, en confirmant notre rôle de concepteurs et de diffuseurs d'environnements logiciels pour la communauté scientifique et en maintenant notre présence, forte, dans les sociétés savantes de nos domaines de recherche.

Par ailleurs, nous consoliderons les bases qui permettent ce développement, à la fois au niveau du **support à la recherche**, qui doit être renforcé, de la **cohésion thématique** de l'unité et de son assise locale qui devraient tirer parti d'un contexte favorable.

4. Orientations scientifiques

4.1 Introduction

Nous avons montré, dans l'inventaire des risques et opportunités liés au contexte et à la configuration actuelle du laboratoire, qu'il nous semblait important de renforcer sa cohésion interne et son ancrage dans l'université. La richesse culturelle de la région et les compétences particulières de notre Université dans le domaine de la culture et du patrimoine nous en donnent l'opportunité. Cette section présente un projet scientifique construit autour des *cultures et sociétés numériques*, qui inscrit notre projet dans celui, plus global, de l'Université.

Ces orientations scientifiques doivent mettre en cohérence les thèmes et les perspectives de recherche de l'unité autour d'un espace commun, non exclusif, dans lequel chacun trouve sa place.

4.2 Culture et sociétés numériques : enjeux sociétaux

La numérisation de l'information et sa mise en réseau mondial changent nos rapports à l'information, à la connaissance, à la culture, au travail,... Certains de ces changements sont clairement perceptibles (voir par

exemple et à des niveaux différents, le rôle des réseaux sociaux dans le *printemps arabe*, la croissance du commerce électronique, le bouleversement du modèle économique des industries culturelles, etc...). D'autres sont latents : l'évolution de la connectivité sociale (So), de la géo-localisation (Lo) et de la mobilité (Mo) conduit à une évolution des usages que les acteurs du domaine appellent SoLoMo et qui pourrait nous amener à faire un pas de plus dans la banalisation de l'Internet et son intégration dans le quotidien.

Notre projet scientifique repose sur l'idée qu'il s'agit d'un changement profond des organisations sociales et économiques qu'il est essentiel de comprendre et de maîtriser. Anticiper et tirer le meilleur parti de ces mutations dépend de la capacité des STIC à produire l'outillage méthodologique et technologique qui permettrait de rendre l'Internet utilisable et intelligible.

Les thématiques de recherche du LIA ne couvrent pas l'ensemble du spectre scientifique concerné par ce défi ; néanmoins, nous sommes dans une situation privilégiée pour répondre à certaines des questions scientifiques majeures qui sont posées.

Les verrous scientifiques et la stratégie de mise en œuvre du projet sont développés dans les sections suivantes.

4.3 Culture et sociétés numériques : verrous scientifiques

Les verrous scientifiques sont liés à la dimension du Web, à sa dynamique, mais aussi à la dispersion des sources et à la fragmentation de l'information. Les principaux axes sur lesquels nous sommes en situation de faire avancer l'état de l'art concernent l'extraction et l'analyse des contenus, leur mise en relation, la recherche des structures sous-jacentes, les questions d'infrastructure et de passage à l'échelle. Nous dressons une liste non exhaustive de quelques points clefs sur lesquels nous nous concentrerons :

- mise en relation des niveaux contenus/structures/usages : cette imbrication des différents niveaux est une particularité essentielle du Web ; la traiter implique d'extraire et d'analyser les contenus, de modéliser les flux d'information, d'intégrer les usages dans les modèles numériques,... Dans cette optique, différents projets ont été proposés à l'ANR sur les appels à projets 2011 ; ces projets regroupent plusieurs thématiques de recherche du LIA et des laboratoires de sciences humaines et sociales de l'Université, notamment le laboratoire Culture et Communication.
- structuration des données hétérogènes : une des principales limites à l'utilisation du Web est la diversité de fond et de forme des contenus : Internet est multilingue, de plus en plus multimédia et interactif ; les sources et les contextes de production et d'échange des données sont extrêmement hétérogènes. Pour extraire de l'information ou de la connaissance, accéder ou naviguer efficacement, il faut produire des représentations à plusieurs niveaux des contenus. Ce type de problèmes est traité au LIA, souvent depuis plusieurs années dans des contextes différents : recherche d'information, indexation multimédia, résumé automatique, traduction automatique, détection d'opinions, etc.
- mobilité et connectivité : le développement des terminaux légers et de la connectivité suscite des nouveaux besoins à la fois du côté des infrastructures (sûreté et sécurité, protocoles dynamiques, réseaux hétérogènes, énergie) et des stratégies de gestion et d'accès aux contenus. Les liens entre l'infrastructure, les usages et les contraintes qu'ils génèrent sur la recherche et la représentation de l'information nous semblent particulièrement intéressants. Là aussi, nous avons déposé des projets ANR qui associent plusieurs thématiques du LIA et des laboratoires de SHS.

4.4 Stratégie et modalité de mise en œuvre

Notre priorité est de maintenir le niveau d'expertise dans nos spécialités et de limiter la fragmentation thématique du LIA en développant des axes de recherche, liés aux sociétés numériques, autour desquels toutes nos thématiques de recherche peuvent se retrouver.

La mise en œuvre du projet respecte ce double objectif de consolidation de nos spécialités et d'harmonisation de nos perspectives scientifiques : elle reposera sur une politique volontariste de recherche de financements sur des thèmes fédérateurs, sur une intégration de ce double objectif dans la politique de recrutement et sur le financement de projets internes au laboratoire.

Elle s'appuiera sur la structure fédérative de recherche *Agorantic*, qui a été créée en Juin 2011. Ces différents points sont détaillés dans les sections suivantes.

4.5 Intégration au projet d'établissement

Le dynamisme culturel régional et la forte activité de l'UAPV dans les domaines de la culture et du patrimoine sont, pour nous, une occasion de jouer un rôle moteur sur ces thèmes au niveau de l'établissement. La note d'orientation stratégique de l'UAPV présente ses 2 axes identitaires : *Sciences et AgroSciences* d'une part, et *Culture et Patrimoine* d'autre part. La note évoque une évolution de ce deuxième axe vers les *sociétés numériques*. Notre objectif est que cette perspective se réalise à l'horizon du mi-contrat, ce qui affirmerait le rôle transversal et fédérateur du LIA dans le projet scientifique de l'Université.

4.6 La Structure fédérative de recherche Agorantic

L'UAPV a créé une structure fédérative de recherche sur les Sciences et Technologies des Cultures et Sociétés Numériques. Cette SFR, appelée *Agorantic*, fédère des laboratoires de sciences et technologies de l'information et des laboratoires de sciences humaines et sociales, autour du Web vu à la fois comme ressource et comme objet d'étude. Son ambition est de créer une dynamique pluridisciplinaire sur ces sujets. Elle est co-dirigée par Marc El-Bèze (LIA) et Marie-Sylvie Poli, du centre Norbert Elias. Ses objectifs s'inscrivent parfaitement dans les perspectives scientifiques du LIA et vont dans le sens des collaborations internes et externes que nous voulons renforcer, en particulier avec les SHS. Son développement sera un de nos objectifs prioritaires des prochaines années.

4.7 Projets transversaux

La mise en œuvre de notre projet scientifique passe impérativement par le soutien des agences de financement nationales et européennes. Nous voulons nous donner les moyens d'amorcer des travaux dans les axes que nous jugeons prioritaires et nous mettre en situation de proposer les meilleurs projets possibles aux agences.

Dans cette perspective, nous avons mis en place des projets internes au LIA, financés sur ressources propres. Les projets soumis sont évalués par deux experts extérieurs au laboratoire, l'arbitrage final étant effectué par le conseil scientifique du CERI.

Ce dispositif a été expérimenté en 2011 ; il a permis de financer, à hauteur de 20 000 euros, un projet associant les thématiques Langage et Recherche Opérationnelle. Cette première expérience s'est révélée très positive ; nous la renouvellerons chaque année, le budget alloué à cette opération sera réévalué chaque année.

4.8 Structure thématique et politique de recrutement

Le laboratoire était, pour la précédente période contractuelle, organisé en thématiques et sous thématiques parfois très proches. Cette proximité a conduit les sous-thématiques liées au Traitement automatique du langage à développer leurs interactions et à se regrouper. Par ailleurs, les activités conjointes des groupes RO et Réseaux se sont développées, avec des soumissions d'articles et le dépôt de projets de recherche communs (le projet ANR RESPET, associant les 2 thématiques, a été accepté en 2011). Les collaborations ente RO et réseaux devront être soutenues et augmentées. Il s'agit d'un élément stabilisant pour deux thématiques qui sont très dynamiques.

Notre politique de recrutement sera conforme à l'objectif de consolidation de la structure thématique de l'unité : nous préserverons le potentiel de recherche du pôle Langage en veillant à ce que les sujets charnières de cette thématique aient les moyens de leur développement, en particulier ceux qui sont à l'articulation des domaines du traitement de l'écrit et de la parole.

Par ailleurs, notre objectif est d'amener les thématiques plus récentes, en Recherche Opérationnelle et Réseaux, à atteindre une masse critique qui les stabilise. Ils devraient bénéficier de l'augmentation, que nous demanderons, du potentiel de recherche de l'unité.

Enfin, nous privilégierons les recrutements qui iront dans le sens d'un renforcement de la cohésion thématique du laboratoire et qui pourront participer à son projet scientifique. Ils devront, bien entendu, être conformes aux besoins spécifiques des formations qui sont dispensées au CERI.

5. Structure administrative et gouvernance

Les perspectives d'évolution structurelle de l'unité sont liées à celles du CERI, avec un double objectif d'amélioration des missions de recherche et de formation en informatique à l'UAPV.

5.1 Le CERI

Le CERI est un département de Recherche et d'Enseignement de l'UFRip Sciences et Technologies de l'Université d'Avignon. Créé en avril 2011, il regroupe le département et le laboratoire d'informatique. Il définit un ensemble d'objectifs communs qui devraient conduire au développement des synergies entre enseignement et recherche, à optimiser l'administration des deux entités et à améliorer leur visibilité. Du point de vue du LIA, le CERI devrait nous permettre de nous adosser au département pour répondre à un certain nombre de problèmes concrets qui se posent à nous :

- **administration** : la mutualisation des compétences et d'un certain nombre de ressources des deux entités devrait nous permettre d'optimiser leur gestion administrative. Ce regroupement doit aussi amener l'équipe administrative à atteindre une masse critique qui lui donnera les moyens d'assurer les missions dont elle a la charge, y compris en cas d'absence ou de difficultés personnelles d'un de ses membres.
- **communication et recrutement** : un des points clefs de la réussite du LIA tient à la qualité des doctorants recrutés. Le Master d'informatique constitue un vivier de doctorants potentiels de très bon niveau : nombre d'entre eux ont entamé une brillante carrière universitaire après être passés par le Master d'informatique suivi d'un doctorat au LIA. L'existence et la qualité du Master est une chance pour le laboratoire qui doit contribuer à la qualité de la formation et des recrutements et doit avoir une politique volontaire d'intéressement des étudiants aux activités de recherche. Le contexte global d'une baisse d'intérêt pour les sciences et pour la recherche augmente encore l'importance de cette démarche. La communication autour d'une entité regroupant enseignement et recherche est un atout de ce point de vue.
- **support et valorisation** : nous avons évoqué, dans le bilan, des difficultés potentielles liées à la faiblesse de notre dotation en personnels IATOS. Une des solutions envisagée repose sur la mutualisation de certaines ressources affectées à l'administration, la communication, la gestion du parc matériel. Nous pourrions aussi nous appuyer sur la formation en alternance pour soutenir notre activité de support et de valorisation de la recherche. Par ailleurs, nous encouragerons la proposition de sujets de fin d'études et de stages proposés aux étudiants qui pourront les impliquer dans le support et la valorisation de nos activités.
- **formation doctorale** : le CERI s'est engagé dans une **démarche Qualité** qui a aboutie, à la certification ISO 9001 du Master en Alternance. Un de nos objectifs est, à terme, d'associer la formation doctorale à ce processus et d'étudier la possibilité de généraliser la démarche Qualité à d'autres activités du laboratoire.
- **politique partenariale** : les formations en informatique de l'UAPV ont, depuis plusieurs années, développées des relations fortes et durables avec le monde industriel. Le CERI doit nous permettre de tirer le meilleur parti du réseau relationnel que le département d'un côté et le LIA de l'autre ont développé ces dernières années.

5.2 Gouvernance du LIA

Le LIA est piloté par un comité scientifique constitué de représentants démocratiquement élus des personnels et de chercheurs extérieurs à la structure qui joueront un rôle de conseil sur les questions de politique scientifique à long terme. Plus précisément, le comité scientifique réunit le directeur du CERI, un représentant par thématique de recherche, des enseignants chercheurs élus, des enseignants chercheurs non permanents et des doctorants, du personnel administratif. Tous ces représentants sont élus par les

membres de leurs collèges respectifs. Le comité est dirigé par le directeur du laboratoire, qui est élu par le comité sur proposition du directeur du CERI. Les membres extérieurs participant au CS seront sollicités sur les questions majeures de politique scientifique ; ils contribueront au pilotage scientifique de la structure. Il s'agit du VPCS de l'UAPV et de 3 chercheurs seniors (1 par thématique) extérieurs à l'UAPV. Le comité scientifique est un lieu de décision et un espace de discussion dans lequel la politique scientifique du laboratoire est discutée et définie. Ses réunions seront ouvertes à tous les membres permanents du LIA. Par ailleurs, nous organiserons des séminaires de laboratoire qui, tous les 2 ans, réuniront les membres permanents et non-permanents, avec des présentations de l'ensemble des travaux réalisés et des tables rondes autour desquelles des projets et des perspectives à plus long terme devront être présentés et discutés. Les modalités de pilotage et, de façon générale, les règles de fonctionnement de l'unité sont définies dans un règlement intérieur qui a été démocratiquement adopté par l'ensemble des membres du CERI.

6. Formation doctorale

Le LIA était lié à l'école doctorale (ED) I2S de Montpellier. Récemment, l'UAPV a obtenu la création de 2 écoles de site, dont l'ED *Sciences et AgroSciences* (ED 536). Le LIA jouera un rôle majeur dans cette structure.

Nous avons mis en place, avec notre école doctorale, des *comités de thèse* qui se réunissent chaque année pour évaluer le niveau d'avancement des travaux de thèse et, éventuellement, offrir une assistance au doctorant. Nous souhaitons aller plus loin dans l'accompagnement des doctorants en les incitant à effectuer des séjours de recherche dans des laboratoires internationaux. Ces séjours devraient avoir une durée minimale de 2 mois ; leur financement pourra être assuré par des bourses de l'UAPV, par les laboratoires d'accueil ou par le LIA sur fonds propres. Le laboratoire accompagnera, avec le service des relations internationales de l'université, les étudiants dans leur projet de visite, à la fois pour identifier les laboratoires qui présentent un intérêt particulier pour le doctorant, pour établir un premier contact et définir le contenu scientifique de la collaboration.

Ce dispositif a été expérimenté en 2010/2011 avec le soutien financier du laboratoire (de 1500 à 2500 euros d'aide par étudiant) ; il a permis à 3 étudiants de passer de 3 à 4 mois dans des laboratoires de renommée internationale.

Notre objectif est de systématiser ces séjours. Outre l'intérêt pédagogique évident d'un tel dispositif, nous pensons qu'il devrait nous permettre de tisser de nouveaux liens avec des acteurs majeurs de la communauté scientifique et de conforter notre visibilité internationale.

Le CERI a mis en place un système de management par la qualité (dont une formation est d'ores et déjà certifiée Iso 9001). L'administration y participe activement et nous envisageons une certification sur le suivi des doctorants en 2013.

7. Support et valorisation

Une part importante de nos recherches et de nos activités contractuelles s'appuie sur un ensemble de plateformes technologiques. Ces environnements sont souvent d'une dimension très conséquente et utilisées par de nombreux partenaires académiques ou industriels. C'est par exemple le cas d'ALIZÉ, dédiée à l'identification biométrique, qui est utilisée par des laboratoires et des industriels de renommée internationale (CMU, MIT, Thalès, etc.). Il s'agit non seulement d'un outillage nécessaire à la validation expérimentale de nos études mais aussi d'un élément contribuant notablement à l'animation de la communauté scientifique et au rayonnement de l'unité.

Pour pérenniser et valoriser ce capital logiciel que les projets nous permettent de constituer, nous devons disposer des moyens humains qui nous font défaut actuellement. Nous chercherons à obtenir ces moyens en suivant les 3 pistes suivantes :

- demandes renouvelées de postes IATOS affectés au LIA,
- développement des interactions enseignement/recherche, de façon à faire converger l'intérêt pédagogique de nos formations et les ambitions du laboratoire en terme de diffusion des produits dérivés de notre recherche. C'est un des attendus du CERI qui est présenté en section 4.1.
- financement, sur fonds propres, d'actions ponctuelles visant à mettre en place les environnements

de support et de valorisation des résultats de notre recherche.

- Le WEB est une vitrine de nos activités scientifiques. Nous développerons notre politique de communication sur ce média, par la diffusion des outils et des ressources produites par le laboratoire et par une présentation complète des résultats de nos recherches.

8. Contrats de recherche

L'activité contractuelle de l'unité a été fortement augmentée ces dernières années. Il s'agit d'un atout essentiel pour le laboratoire, qui lui permet de développer ses collaborations et de développer ses perspectives de recherche. Il est critique de maintenir le niveau de ces activités, en continuant à être sélectif sur la nature et les sujets des projets. Être une force de proposition de projets et en être les porteurs nous met en situation de jouer un rôle moteur dans la définition de leurs objectifs scientifiques. C'est un point important pour préserver la maîtrise de notre politique scientifique tout en obtenant les moyens de la mettre en œuvre.

Par ailleurs, nous jouerons pleinement notre rôle auprès des agences de financement et d'orientation de la recherche, aux niveaux national et européen, pour faire valoir notre vision des axes qui doivent être financés en priorité.

9. Politique financière

Nous avons évoqué, dans le chapitre *Autoévaluation*, les risques liés au déséquilibre entre la dotation récurrente et le budget généré par nos activités contractuelles. Pour limiter ce déséquilibre et financer notre politique de développement, nous mettrons en place un certain nombre de mesures qui doivent nous aider à financer une politique scientifique à long terme, qui ne soit pas directement dépendante des projets de recherches financés par les agences.

La première d'entre elles consiste à effectuer un prélèvement sur les contrats ANR, EU ou industriels. Pour les projets ANR il s'agit d'un reversement des frais de structures pris en charge par l'ANR. Cette part du financement était, jusqu'à présent, gérée par les porteurs de projet. Depuis 2011, le laboratoire réinvestit ces sommes dans les projets transversaux (cf. Section 3.7). Nous envisageons d'étendre ce dispositif aux autres types de contrats dans les années à venir.

Par ailleurs, l'activité contractuelle soutenue de la plupart des membres du LIA permet désormais un financement des missions et du matériel par les contrats de recherche. Les moyens récurrents du laboratoire sont utilisés notamment pour donner à chacun les moyens de développer son activité en dépit des aléas liés à l'obtention de contrats industriels ou collaboratifs. Ce soutien du laboratoire porte essentiellement sur le financement de missions ou, plus marginalement de postes de travail. Les demandes de financement de missions sont arbitrées en conseil de laboratoire, en fonction de règles prédéfinies.

10. Liens avec les EPST

Nous avons évoqué, dans le bilan, l'historique de nos demandes d'association au CNRS. Malgré de très bonnes évaluations, cette démarche n'a pas abouti. Nous renouvelerons cette demande qui nous semble légitime et qui pourrait nous donner les moyens directs d'une politique scientifique à long terme. Par ailleurs nous développerons nos liens avec l'INRIA, avec la perspective de signature d'un accord cadre prenant appui sur la convention de collaboration déjà signée dans le cadre du projet MAESTRO.

11. Appels à projets liés aux investissements d'avenir

Le laboratoire s'est associé aux réponses de l'UAPV aux appels à projets *Investissements d'Avenir*. Une première proposition d'Equipex a été soumise en 2010; elle concernait l'équipement accompagnant le projet scientifique de la SFR Agorantic. Ce projet n'a pas été retenu, et nous avons déposé en 2011 une nouvelle demande (Patrimonia) en réponse au second appel à projet Equipex. Enfin nous projetons de soumettre un projet de laboratoire d'excellence sur le langage et la cognition, avec différents laboratoires des Universités Marseillaises. La réussite d'un tel projet nous permettrait de structurer nos relations, déjà fortes, avec le LPL et le LIF.

12. Animation scientifique de l'unité

Le LIA organise des séminaires hebdomadaires auxquels peuvent s'ajouter des présentations exceptionnelles, liées aux visites de chercheurs accueillis par l'une des thématiques, et chaque thématique organise des réunions de travail régulières. Ces activités seront poursuivies et nous mettrons en place, dès 2012, un séminaire de 2 jours réunissant l'ensemble des membres du laboratoire pour faire un point de l'état d'avancement des différents projets et échanger sur les difficultés éventuelles et les pistes de développement de l'unité à moyen et long termes.

13. Projet scientifique par thématique

13.1 Projet de la thématique Langage

13.1.1 Orientations stratégiques

Le dynamisme scientifique de la thématique s'est traduit, ces dernières années, par une augmentation du nombre de publications, par un engagement dans les campagnes d'évaluation nationales internationales, par la diffusion, vers la communauté scientifique, de ressources produites dans le cadre nos activités de recherche. Par ailleurs, nous avons significativement développé nos collaborations nationales et internationales, en particulier dans le cadre de collaborations contractualisées. Nos priorités, dans les années à venir, sont de 3 ordres :

- **optimiser notre potentiel de recherche** : il est essentiel pour nous de maintenir notre potentiel de recherche et d'encadrement. Notre dynamisme repose sur un investissement très important de chaque membre de l'équipe dans la recherche ainsi que sur une activité contractuelle efficace mais chronophage. Par ailleurs, la thématique contribue très significativement au pilotage et à l'administration de la structure, à la fois dans ses aspects scientifiques et pédagogiques (direction des études, du laboratoire, vice présidence du conseil d'administration de l'Université, responsabilité du Master, responsabilité de la SFR, ...). Une augmentation des moyens dédiés au support et à la valorisation de la recherche pourrait nous permettre d'éviter un effritement de notre potentiel de recherche dû à la prise en charge, par des personnels enseignant chercheurs, de diverses tâches qui pourraient être assurées par des personnels IATOS.
- **structurer nos collaborations** : les collaborations nationales et internationales de la thématique sont nombreuses, mais souvent orientées vers des laboratoires dont la culture scientifique est proche de la nôtre. Notre objectif, pour la prochaine période quinquennale, est de structurer ces collaborations en allant dans le sens de la complémentarité. La structure fédérative de recherche Agorantic et son orientation sur la culture et les sciences du Web devrait nous offrir cette occasion d'ouvrir et de structurer des collaborations multi-disciplinaires. Sur un autre plan, nous projetons de structurer nos collaborations avec des équipes proches scientifiquement et géographiquement.
- **valoriser les résultats de notre recherche** : le domaine du traitement automatique du langage naturel implique des infrastructures logicielles et matérielles lourdes. De plus, les recherches réalisées reposent souvent sur l'analyse de corpus de grandes tailles. Ces dernières années, la thématique a fourni un effort très conséquent pour le développement de plate-formes logicielles et de bases de données qui sont des composants stratégiques du domaine. Nous chercherons, dans les années à venir, à tirer le meilleur parti de cet investissement en mettant en place les moyens d'une valorisation durable de ressources qui peuvent contribuer au dynamisme de l'ensemble de la communauté scientifique du domaine.

13.1.2 Orientations scientifiques

La thématique langage a pour objectif à moyen terme de travailler sur les technologies qui sont au cœur du domaine et de développer quelques axes émergents, dont nous pensons qu'ils deviendront des sujets majeurs dans les années à venir et qui s'inscrivent dans le projet scientifique global du LIA.

D'un point de vue méthodologique, la communauté scientifique TAL s'est concentrée sur les méthodes statistiques avec une orientation système très marquée. Nous avons contribué à ce mouvement qui a permis, sur de nombreux sujets, de faire des progrès considérables. Nous envisageons de poursuivre cet effort. Néanmoins, dans les années à venir, nous pensons qu'une ouverture des technologies vers les utilisateurs, l'environnement, les sources d'informations de natures et de niveaux différents devrait permettre de lever un certain nombre de verrous scientifiques importants et, en aval, d'améliorer l'utilisabilité des technologies.

Par ailleurs, le développement du WEB et des grandes bases audiovisuelles offre, pour la thématique, l'opportunité de développer des aspects complémentaires de ses activités – notamment entre l'oral et l'écrit – avec des problèmes nouveaux liés à l'exploitation conjointe des différentes modalités pour l'extraction,

l'analyse et l'exploitation des contenus. Dans cette optique, nous envisageons de renforcer nos collaborations scientifiques internes, de les étendre au périmètre de l'université en particulier vers les disciplines liées à la production des contenus, à l'analyse des usages, et, au delà, de structurer nos collaborations avec les laboratoires proches qui sont des acteurs majeurs du domaine.

Nos perspectives portent sur 3 axes de recherche complémentaires :

a) Technologies au cœur des systèmes de traitement automatique du langage. Il s'agit de poursuivre les travaux que nous avons mené sur les modèles et les algorithmes dédiés au TAL, qu'il soit écrit ou oral, dans différents contextes.

Les systèmes développés reposent le plus souvent sur de l'apprentissage automatique et des modèles statistiques. Dans ce type d'approches, les deux points critiques concernent les informations fournies aux systèmes, les modèles et les stratégies d'identification adoptées.

Outre le développement de modèles et de stratégies de reconnaissance originales, nous nous concentrerons sur l'intégration de différents niveaux de représentation et de sources de connaissances hétérogènes. Cette approche a été initiée sur différents sujets au LIA, par exemple en recherche d'information, en reconnaissance et en compréhension de la parole ou en caractérisation du locuteur.

b) Interprétabilité, extraction de connaissances

Les méthodes d'apprentissage automatique sont très largement utilisées en TAL. Elles ont permis de construire des systèmes efficaces sans s'appuyer sur une connaissance explicite des objets d'études, connaissance souvent très partielle ou incertaine. Cet axe considère les systèmes d'apprentissage automatique comme des outils d'analyse, qui peuvent permettre d'extraire de l'information de haut niveau des contenus. Cette question a été abordée dans différents contextes au LIA (voir par exemple la thèse de J. Khan, en reconnaissance du locuteur, ou les travaux de I. Da Cuna mariant linguistique et TALN, les travaux sur l'analyse des voix pathologiques). C'est aussi un point essentiel pour l'analyse des contenus parlés (*Speech Analytics*), qui est un sujet de recherche fédérateur de la thématique et qui suscite un intérêt croissant dans la communauté scientifique.

Nous voulons développer cet axe en nous associant à d'autres laboratoires, souvent du domaine des SHS, qui ont une approche analytique de l'étude du langage ou *par* le langage (linguistique, culture, communication, médecine, ...). Cette ouverture pourra se faire dans le cadre du projet de Structure Fédérative de Recherche autour des sciences du Web, porté par l'UAPV.

c) Approches centrées sur utilisateur

A un certain niveau de maturité, les performances des systèmes de TAL plafonnent et leur amélioration devient coûteuse et hypothétique. Une de voies permettant de dépasser ces limites consiste à replacer l'utilisateur au centre du système, en identifiant clairement les contraintes et les opportunités liées au contexte et en utilisant l'information fournie de façon directe ou indirecte par les opérateurs ou les utilisateurs des systèmes.

Cette approche peut être adoptée dans les phases d'apprentissage, d'utilisation et d'analyse des systèmes et de leurs usages. Elle a connu des développements importants dans les domaines de l'apprentissage automatique avec des applications au traitement de la vidéo, de la parole ou de l'écrit. Nous avons ouvert des chantiers sur ce sujet (thèses de G. Senay, de P. Gotab, projet ANR SUMACC). De façon plus générale, la question de la place de l'utilisateur dans les systèmes d'interaction homme-machine est fondamentale, quels que soient la modalité, le contexte ou les attendus de l'interaction.

13.1.2.1 Projet en Traitement Automatique du Langage Naturel Écrit

Dans l'ensemble des projets en Traitement Automatique du Langage Naturel Écrit (TALNE), le LIA privilégie le paradigme d'intégration des différents composants développés. Quel que soit le thème, en TALNE on s'attache à combiner des méthodes symboliques et numériques dans le but d'atteindre le double objectif de traiter des problèmes complexes et de taille réelle. Les résultats attendus sont des publications de haut niveau, la cession de licences, la mise à disposition d'outils d'utilisation libre (sous licence GPL) et la participation aux conférences et campagnes nationales et internationales d'évaluation (DEFT, TREC, EQUER, Technolangue, INEX, ARCADE EVALDA, SENSEVAL, TDT TREC, TAC TREC, MEDIAEVAL). Nous voulons aussi développer et élargir les collaborations internationales existantes. Nos perspectives s'articulent autour de 4 axes :

a) Extraction de contenus et Recherche d'Information

Le cœur de cette activité concerne la prise en compte de méta-information dans la phase d'Indexation. Il est aussi question de l'utilisation de Bases de Connaissances disponibles ou extraites automatiquement et la proposition de modèles nouveaux pour mesurer l'appariement entre questions et réponses potentielles présentes dans des documents textuels. Les méthodologies et outils utilisés sont principalement les modèles vectoriels ou probabilistes, les algorithmes d'apprentissage automatique supervisé ou non, les réseaux de neurones (RN), les arbres de décision et les Support Vector Machines (SVM). Les moyens spécifiques sont les ressources lexicales incluant les niveaux linguistiques les plus élevés, les bases de données textuelles et les corpus de référence sans exclure l'acquisition et la collecte de documents appropriés au traitement d'une tâche spécifique. Nous visons des applications dans le cadre des moteurs de recherche documentaire et de questions réponses et à l'activité Dialogue avec entrée vocale.

b) Catégorisation, Classification, Segmentation de documents

L'objectif ici est de réaliser un travail fondamental sur les méthodes de classification des données textuelles portant sur les algorithmes eux-mêmes. Nous réalisons des études sur la possibilité de ne pas dissocier la classification des autres problématiques traitées par les différents composants d'une des applications de TALNE. Pour y parvenir nous utilisons des algorithmes d'apprentissage automatique supervisé ou non, les RN et SVM, arbres de décision et méthodes discriminantes. Les applications sont les mêmes que celles citées précédemment sans pour autant exclure la possibilité d'appliquer les méthodes proposées à d'autres domaines.

c) Résumé automatique de documents

Les objectifs scientifiques en résumé automatique concernent le développement de méthodes pour la génération automatique de résumés de haute qualité (extraits, abstraits, résumé multilingue, multi-document, multi-média, résumé sémantique,...). Nous étudions aussi la génération de texte par transformation de phrases (compression, paraphrase, génération) et des recherches sur l'évaluation automatique de résumés. Les méthodologies et outils utilisés sont les modèles de Markov cachés, les RN et SVM, les méthodes symboliques, les stratégies à base de vote ainsi que les méthodes (symboliques/ numériques) de représentation et de transformation de phrases. Nous utilisons aussi les ressources lexicales (niveaux linguistiques les plus élevés) les corpus alignés de résumés et sources, l'utilisation de corpus de référence sans exclure l'acquisition et la collecte de documents appropriés au traitement d'une tâche spécifique. Les applications concernent les systèmes de filtrage et de résumé automatique de documents dans un sens large.

d) Annotation de corpus et enrichissement de ressources

Les objectifs scientifiques de cet axe consistent à développer des méthodes pour l'enrichissement des corpus permettant l'annotation linguistique (syntaxique, morphologique, sémantique, pragmatique, etc.). Nous portons un intérêt particulier à l'étude sur les corpus annotés et non annotés, ainsi que les graphes terminologiques. Les méthodologies et outils employés sont les modèles de Markov cachés, les RN et SVM, les arbres de décision ; les algorithmes de graphes et les approches symboliques. Les applications se situent dans le cadre des systèmes mettant en œuvre l'extraction de contenus.

13.1.2.2 Projet en Traitement Automatique de la Parole

D'une façon générale, la technologie du traitement automatique de la parole a atteint une certaine maturité mais de très nombreux problèmes restent ouverts. Notre objectif est de consolider notre positionnement international, de valoriser les plate-formes logicielles développées notamment pour la reconnaissance du locuteur et de travailler sur des approches alternatives et complémentaires susceptibles de répondre à des problèmes de fond du traitement automatique de la parole.

Par ailleurs, le développement des grandes bases audiovisuelles et du WEB suscite des problèmes nouveaux, liés à la diversité des conditions, des usages, et au besoin d'outils permettant la structuration et la navigation par les contenus.

a) Voix pathologique

Jusqu'en 2009, les travaux du LIA dans le domaine de la caractérisation de pathologies se sont focalisés sur les troubles de la voix (dysphonie). Ce domaine d'activité s'est diversifié en 2009 dans le cadre du projet ANR DesPho-APady. L'expertise du LIA dans le cadre du traitement automatique s'applique à présent également sur la caractérisation des troubles de la parole (dysarthrie) dont peuvent souffrir des patients atteints de maladies d'origine neurologiques (ex : maladie de Parkinson, sclérose amyotrophique latérale, maladies génétiques rares...). Ces travaux ont diversifié les collaborations du LIA avec de nouveaux partenaires cliniques telles que les hôpitaux la Pitié Salpêtrière et Pitié-Salpêtrière et Pitié-Salpêtrière (Paris), d'Aix-en-Provence ou encore le CHU de Toulouse. Ces collaborations seront intensifiées dans les années à venir. Le LIA continuera à adapter les techniques statistiques de reconnaissance/classification automatique du locuteur dans le cadre du traitement des dysphonies. Il engagera également son expertise dans le cadre des dysarthries pour proposer des techniques de classification non supervisée de signaux de patients présentant des troubles de la voix et de la parole. Par ailleurs, nous mettrons au point des mesures d'intelligibilité basées sur la reconnaissance de la parole pour l'évaluation objective de l'intelligibilité des patients atteints de dysarthrie.

b) Variabilité acoustique, robustesse, espaces de représentation

Il s'agit d'un thème central en TAP que nous avons abordé par la recherche et la combinaison de paramètres complémentaires et, au niveau modèle, en proposant des architectures originales basées sur l'analyse factorielle ou la généralisation de HMM semi-continus. Ces travaux ont ouvert des pistes que nous explorerons, notamment pour le développement de modèles de décomposition à base d'analyse factorielle en reconnaissance de la parole et la recherche d'espaces de représentation compacts et/ou discriminants, par exemple en poursuivant les études entreprises sur les clefs d'identification binaires en reconnaissance du locuteur (thèse de P.M. Bousquet, brevet en cours),

c) Catégorisation audio et approches hybrides

Il s'agit d'un thème très large, qui a connu des développements importants avec la croissance des données multimédia sur le Web et le besoin d'annotation fine des caractéristiques extra-linguistiques des signaux de parole. Nous renforcerons cet axe, en particulier en tirant parti du savoir faire et des environnements techniques développés pour l'authentification biométrique. Des approches hybrides combinant différentes modalités (audio, texte, vidéo) ou différents niveaux de représentation seront proposées.

d) Stratégies d'identification, combinaisons et intégration de systèmes

Les systèmes de TAP *état de l'art* reposent le plus souvent sur des formulations statistiques qui présentent l'avantage décisif d'offrir un cadre théorique clair pour l'intégration de sources d'information diverses, dès lors qu'elles peuvent être projetées dans un espace probabiliste. La fusion de sources très hétérogènes reste cependant problématique et oblige à un important travail de formalisation et, souvent, de mise au point des systèmes. Ces questions seront abordées sur la combinaison de systèmes de reconnaissance de la parole et en reconnaissance du locuteur, domaine dont le LIA est devenu un acteur majeur au niveau international.

e) Traitement de la parole en situation de ressources limitées

Un des points critiques des technologies de traitement automatique de la parole est lié à la quantité de données nécessaire à l'estimation des modèles sous-jacents, à la fois pour les tâches de transcription et pour celle de caractérisation extra-linguistique des contenus (locuteur, langue, thèmes, etc.). Nous avons abordé ce sujet dans différents contextes, liés à la transcription de langues peu dotées (projet ANR PI), ou pour l'adaptation des systèmes de reconnaissance de la parole ou du locuteur. Nous développerons ces travaux qui sont connexes à la fois à la question de la préservation des patrimoines culturels et à celle du passage à l'échelle pour le traitement des grandes collections multimédias. Ceci pourra notamment porter sur l'utilisation des techniques de caractérisation audio par clefs binaires, développées au LIA pour l'authentification biométrique, pour la structuration de très grandes bases multimédias.

13.1.2.3 Projets en Interactions vocales et Traitements linguistiques de la Parole

Les activités du LIA pour l'interaction vocale et les Traitements linguistiques de la Parole continuent de

s'inscrire dans un double objectif :

- collaborer avec les meilleurs laboratoires universitaires et industriels travaillant dans le domaine du dialogue homme-machine, à travers la participation à des projets, afin de continuer à produire de la recherche académique de haut niveau,
- garder des partenariats avec les grands opérateurs de télécommunication et les entreprises développant des services vocaux interactifs (Orange, AT&T, Loquendo) afin d'avoir à disposition des cadres applicatifs réalistes et travailler sur des données "réelles" issues de services mis en fonction auprès du public.

Dans ce cadre, les activités vont être organisées autour de 2 axes comportant plusieurs activités principales :

- **Axe 1 - Interactions vocales : dialogue homme-machine**
 - Compréhension de la parole multilingue et multi-domaine
 - Gestion du dialogue stochastique
 - Interactions Homme-Robot et Homme-Avatar
- **Axe 2 - Traitements linguistiques de la Parole**
 - Traduction automatique de la Parole
 - Extractions de méta-informations de données audio

Nous ne détaillons plus avant que quelques unes de ces activités.

a) Serveurs Vocaux interactifs

Les serveurs vocaux interactifs, implémentant un dialogue en langage naturel entre un utilisateur et un système d'information, sont un domaine en pleine expansion. Ce type de serveurs s'apparente à une interface entre un utilisateur et une base de données. Le but du dialogue est alors de formaliser les requêtes qui seront adressées à la base de données. On peut trouver dans cette famille de systèmes l'ensemble des systèmes de réservations (d'hôtels, de restaurants, de trains, d'avions, etc.) ainsi que les systèmes de relation clientèle (accès, questions sur des factures, suivi de commande, questions sur un service ou un produit, etc.).

b) Compréhension de la parole multilingue et multi-domaine

Le cœur de cette activité concerne le problème de la compréhension de la parole. Après avoir proposé des modèles de représentation des connaissances de base du processus de compréhension et des algorithmes de décodage conceptuel intégrés au processus de reconnaissance automatique de parole, nous allons désormais développer des modèles de compréhension plus riches permettant d'avoir une représentation sémantique complète d'un message, dans un cadre de dialogue homme-machine.

Ces travaux sont menés dans le cadre du projet ANR Port-MEDIA qui a démarré en mars 2009. Port-MEDIA est un projet CONTINT qui regroupe les partenaires suivants : LORIA-Nancy, LIG, LIUM, ELDA et le LIA. Le thème du projet est la compréhension robuste de la parole spontanée pour les applications de serveurs vocaux. L'originalité du projet réside dans la multilingualité (développer une application dans une langue et la traduire dans une autre) ainsi que sur la diversité des systèmes de dialogue homme-machine envisagés.

c) Interactions Homme-Robot et Homme-Avatar

Dans la cadre d'une collaboration avec Supélec-Metz et le LASS, nous entendons étudier l'interaction homme-machine sous l'angle de la mise en situation. Afin de mettre en situation les systèmes interactifs, nous proposons d'utiliser des robots humanoïdes particulièrement mais aussi d'étudier l'usage de têtes parlantes (avatar virtuel). Par ailleurs, la mise en situation implique la nécessité pour l'interface de s'adapter à son environnement et à son interlocuteur. Des méthodes d'apprentissage automatique sont donc nécessaires afin d'obtenir cette adaptation. Supélec-Metz sera le coordinateur du projet, le LAAS apportera son expertise en robotique et Aldeberan Robotics son soutien par le biais de la fourniture de robots humanoïdes.

Cette approche située du dialogue homme-machine implique un certain nombre de problèmes scientifiques et techniques : par exemple des problèmes d'acquisition du signal de parole qui dégraderont les taux de

reconnaissance et de compréhension du langage. Les ambiguïtés générées par ces taux de reconnaissance dégradés devront être traitées au sein des systèmes de gestion de l'interaction. Cette problématique est étudiée dans la communauté du dialogue homme machine par le biais de modèles d'optimisation statistiques de prise de décision (e.g. les processus décisionnels de Markov partiellement observables) mais reste néanmoins un verrou important. La faculté d'adaptation à des situations contextuelles et dialogique différentes ainsi qu'à l'utilisateur par apprentissage constituera aussi un axe de recherche. L'application robotique supposant par ailleurs un usage plus régulier par un utilisateur identique, l'adaptation sur le long terme est une caractéristique désirée. Un autre verrou à attaquer sera l'étude de l'attitude à faire adopter aux robots humanoïdes et avatars afin de rendre l'interaction la plus naturelle et efficace possible.

L'originalité principale de ce projet est de se baser sur des méthodes d'apprentissage numérique visant initialement à réaliser du contrôle optimal. Typiquement, il s'agit de méthodes d'apprentissage par renforcement permettant d'explorer différentes possibilités d'interaction et d'apprendre par essais-erreur. Ce type de méthode a déjà prouvé son efficacité dans le domaine du dialogue purement vocal (applications téléphoniques essentiellement). Il s'agit ici d'étudier la possibilité de transposer les résultats de ces méthodes dans le cadre d'une approche située du dialogue homme-robot avec une méthodologie progressive. La mise en situation des systèmes de dialogue grâce à la robotique et aux avatars va en effet augmenter la complexité et le nombre des actions réalisables par le système et des situations qu'il pourra rencontrer. Par conséquent, il faudra faire passer à l'échelle les méthodes d'apprentissage utilisées dans le domaine du dialogue ce qui nécessitera de nouveaux développements théoriques.

d) Traduction automatique de la parole

Une activité dans le domaine de la traduction automatique a débuté en 2011 dans le cadre d'une collaboration avec l'équipe de traitement automatique de la langue naturelle du LIG. Ce travail a donné lieu au développement au LIA d'un système de traduction statistique basé sur le toolkit Moses, pour la paire de langue anglais-français. Ce système, combiné avec celui du LIG pour une soumission commune pour la campagne internationale en traduction automatique WMT de 2011, a obtenu de bonnes performances puisqu'il s'est classé 2ème sur 16 participants pour la direction de traduction français vers anglais, d'après les évaluations humaines. Nous souhaitons dans les années à venir continuer à développer une activité de traduction automatique, en nous appuyant sur les spécificités du LIA qui possède une bonne expertise dans le domaine de la parole. En continuant à faire évoluer notre système de traduction en nous appuyant sur Moses de manière à conserver un système à l'état de l'art, nous voulons principalement nous concentrer sur la traduction de documents audio, qui sont encore peu étudiés comparativement aux actes parlementaires ou aux articles d'information. Nous souhaitons en particulier nous attaquer à l'utilisation de connaissances sémantiques, comme celles extraites dans le cadre des interactions vocales, pour guider la traduction des documents audio. Cette approche nous paraît intéressante dans ce cadre car les systèmes doivent baser leurs hypothèses sur les sorties imparfaites des systèmes de reconnaissance de parole et ont ainsi intérêt à prendre leur décision sur un maximum d'informations.

13.1.3 Analyse des opportunités et risques

Cette section décrit les risques et opportunités liés à l'environnement de la thématique : laboratoire, Université, contexte scientifique global et aux évolutions sociétales et économiques.

Opportunités

- l'affirmation par l'UAPV d'un axe identitaire autour de la culture et du patrimoine et le mouvement vers le numérique présente pour nous une bonne opportunité de développer des interactions avec d'autres thématiques du laboratoire, d'autres laboratoires de l'UAPV et d'afficher nos compétences actuelles sous une **bannière Cultures et sociétés numériques** qui nous place au centre de la stratégie de l'Université et qui nous inscrit dans une perspective d'avenir, dont les enjeux sociétaux vont croissants.
- la thématique réunit des compétences en extraction, analyse des contenus parlés et traitement de l'écrit : cette **mixité écrit/parole** présente un réel potentiel qui peut être encore développé.

Risques

- l'activité contractuelle pourrait guider les orientations scientifiques de la thématique ou la pousser à investir des ressources qu'elle n'a pas dans la valorisation. Nos priorités doivent être clairement

définies et l'engagement dans un projet scientifique fédérateur et ambitieux sur le long terme devrait nous permettre d'éviter cet écueil.

- les orientations scientifiques du laboratoire vers les *cultures et sociétés numériques* pourraient détourner la thématique des sujets qui sont au cœur de son expertise. Ses sujets de recherche actuels s'inscrivent très naturellement dans ce projet et nous affirmerons cette ambition double de consolidation dans nos spécialités et d'ouverture autour de la *culture et des sociétés numériques* vues comme un objet d'étude commun.

13.2 *Projet de la thématique Recherche Opérationnelle et Optimisation*

13.2.1 Orientations stratégiques

La thématique entend d'abord **corriger ses faiblesses**. Si un virage a été amorcé, il convient de confirmer les orientations encore naissantes : le recrutement d'un Maître de Conférence vient en partie combler le déficit en taille de l'équipe et permettra sans doute de mieux répondre aux exigences, notamment en terme de **projets subventionnés par l'ANR** et par la **Communauté Européenne**, des activités de recherche modernes. Nous continuerons également à développer nos collaborations internes d'une part avec les autres thématiques du LIA mais également avec d'autres chercheurs de l'UAPV.

En pratique, cela se traduira par une **participation active** aux activités de la **Structure Fédérative Agorantic (SFR)** réunissant, outre le LIA, les laboratoires de Culture et Communication, de Géographie et de Droit de l'Université et des partenaires extérieurs. Ainsi, certains des projets de la thématique devraient avoir un caractère pluridisciplinaire plus marqué. Il s'agira de participer et, dans certains cas de porter, des projets liés à l'usage du numérique dans des domaines comme la culture et le patrimoine.

Par ailleurs, mais avec tout autant de détermination, nous maintiendrons une **politique volontariste d'échanges internationaux**. Nos collaborations avec l'**Amérique Latine** (Brésil, Chili, Pérou notamment) et **le Canada** seront dès lors, pour le moins préservées et éventuellement encore renforcées. Là aussi, l'arrivée d'un membre supplémentaire dans la thématique sera une occasion d'augmenter et intensifier les échanges avec ces pays mais aussi d'étendre nos réseaux de collaboration, en particulier vers **l'Afrique (Sénégal notamment)**. La thématique joue un rôle moteur dans le cadre de la participation de l'UAPV au consortium fondateur de l'**Université des Sciences et Techniques de Hanoï** (Vietnam) où elle est partie prenante dans le master du département STIC. Un étudiant de doctorat issu de l'USTH est d'ailleurs déjà membre de l'équipe. De plus, nous entendons nous ouvrir à l'**Europe** en établissant des échanges avec le MIT-Zaragoza Logistic center (MIT-ZLC), centre international de recherche en Transport et Logistique, situé à Saragosse (Espagne), avec lequel nous sommes en contact. Il s'agira ainsi de se positionner pour être en mesure de **répondre aux appels du PCRD** et de s'inscrire dans une mouvance, inspirée par la LRU, qui assigne comme mission aux universités la construction de l'espace européen de la recherche.

Nous maintiendrons également avec conviction, une **politique active de recrutement et d'encadrement des doctorants**. Ceux-ci sont le plus souvent porteurs d'un souffle qui ne peut-être que bénéficiaire à la thématique, au laboratoire et à ses membres permanents qui, en échange, contribuent à leur formation et à leur insertion professionnelle. Nous insisterons en particulier sur la mobilité de nos doctorants, sur leur participation à des conférences nationales et internationales et sur les formations de préparation à l'insertion professionnelle.

Nous nous efforcerons aussi de garder les trois axes à partir desquels nous avons défini le bilan de la thématique : **méthodes, problèmes académiques et applications**, tout en développant une activité contractuelle que nous souhaitons forte.

13.2.2 Orientations scientifiques

13.2.2.1 Méthodes

Nos travaux sur le développement de méthodes originales pour l'optimisation discrète resteront, à moyen terme, dans la lignée de ce qui a été entrepris au cours du quinquennal précédent.

Une part importante de ceux-ci concerneront la résolution de **problèmes quadratiques en variables binaires** pour lesquels nous tenterons de généraliser (et d'exploiter !) la notion de persistance. On parle de persistance d'une variable lorsqu'elle prend la même valeur dans toutes les solutions optimales. Nous espérons pouvoir généraliser ceci à des expressions logiques impliquant plusieurs variables et dont la valeur serait donc la même pour toute les solutions optimales.

Nous continuerons également à travailler sur **Resolution Search**, méthode alternative au branch-and-Bound et sur les méthodes exploitant à la fois les caractéristiques de la **programmation mathématique** et celles de la **programmation par contraintes**.

13.2.3 Problèmes

Pour ce qui a trait aux problèmes, nous développerons nos activités en **transport**, dans le cadre de problèmes académiques mais aussi dans une optique d'applications innovantes que nous évoquerons plus loin. Nous nous intéresserons à des problèmes de **localisation et design de réseau**. Nous nous intéresserons aussi à des problèmes de **planification, affectation et ordonnancement** de tâches sur des ressources, bi-objectif (quadratique : minimisation des coûts de déplacement des flots inter-tâches, linéaire : minimisation des sommes des temps d'attente).

Nous poursuivrons également nos travaux sur l'usage de méthodes de décomposition pour les problèmes d'ordonnancement en cherchant à généraliser nos résultats à des classes plus larges de problèmes.

Enfin, nous continuerons à nous intéresser aux problèmes de la famille du sac-à-dos, en particulier en considérant le **problème d'affectation généralisée**.

13.2.3.1 Applications

Dans la Structure Fédérative de Recherche Agorantic, nous contribuerons aux problématiques issues des **Transports Intelligents** et de **l'Intelligence ambiante** en collaboration avec d'autres équipes du LIA mais également avec les géographes des UMR ESPACE (Université d'Avignon) et UMR ESPACE-DEV (Université de Montpellier). Nous chercherons à optimiser le transport dans une grande ville (du sud) en agissant à la fois sur la localisation de ses activités et sur le design de son réseau de transport. Nous chercherons également à développer des techniques permettant un usage optimisé d'appareils embarqués sur des véhicules pour un meilleur contrôle des flux de véhicules (via les feux de signalisation). Ces systèmes capables de recevoir et de transmettre de l'information (sur l'état du trafic, sur l'état des stocks pour les livraisons etc.) ne seront pleinement utiles que si cette information est traitée et exploitée de façon à déterminer des politiques optimales de déplacement, de stockage etc. Cela inclut notamment la **planification d'un réseau local Internet**, en collaboration avec Eitan Altman, DR INRIA associé à la thématique Réseaux du LIA. Dans ce type de réseaux, les télécommunications transitent entre les usagers en étant stockées sur des véhicules utilitaires (bus). Il faut alors optimiser le trajet du véhicule qui fait transiter l'information de manière à ce que, globalement, le réseau soit le plus efficace possible.

En collaboration avec l'Universidad Tecnica Federico Santa Maria, Santiago, Chili, nous interviendrons, en amenant nos compétences en optimisation, dans le projet **ALMA** (<http://www.eso.org/sci/facilities/alma/index.html>) qui est l'un des plus gros projets (terrestres) d'astronomie. Il nous faudra trouver des techniques pour optimiser, en incluant des notions de robustesse, l'usage d'un réseau de télescopes situés dans le désert de l'Atacama (Chili). Un projet ECOS-CONICYT a été déposé en ce sens.

Par ailleurs, nous continuerons nos travaux sur la **recherche de cibles intelligentes**.

13.2.4 Analyse des opportunités et risques

La réflexion sur le bilan de la thématique nous a permis de proposer une analyse de nos forces et faiblesses mais nous amène également à évaluer les risques et opportunités qui lui sont attachés.

Les risques sont en réalité mineurs et résident essentiellement dans le fait que beaucoup des axes d'étude envisagés sont en collaboration avec des chercheurs d'autres disciplines et sur des problématiques nouvelles (pour certains projets). Cela peut induire des lenteurs, dues à des difficultés de compréhension entre les différents partenaires, des difficultés également de «gestion» du projet, et, enfin, dues à des contextes nouveaux qu'il nous faudra, appréhender (projet ALMA et optimisation de l'utilisation des systèmes embarqués). De plus, ces projets collaboratifs ont une consonance appliquée bien réelle et vont orienter un peu plus qu'elles ne l'étaient nos activités vers des applications. Il se peut ainsi que notre rythme de publication soit amené à baisser légèrement. Ce risque sera amenuisé par le maintien d'activités plus traditionnelles et plus théoriques de Recherche Opérationnelle.

Néanmoins, ces projets sont également de formidables opportunités pour développer des concepts innovants et pour s'inscrire dans un processus de développement technologique. Par ailleurs, la fréquentation assidue de collègues de disciplines voisines ou éloignées (géographes, spécialistes de la Culture, de la Communication, juristes) devrait générer, outre les projets rapidement évoqués ici, d'autres synergies et une dynamique que nous envisageons avec enthousiasme. Nul doute non plus que de tels projets offriront davantage de possibilités en terme de financement par l'ANR ou dans le cadre du PCRD.

C'est donc globalement une stratégie orientée vers les échanges que nous proposons ici : échanges avec l'étranger (partenaires traditionnels ou nouveaux), échanges avec des disciplines connexes ou éloignées, et échanges avec les doctorants.

13.3 Projet de la thématique Réseaux

13.3.1 Orientations stratégiques

Notre thématique souhaite renforcer des liens privilégiés avec plusieurs laboratoires nationaux ou internationaux. Ces partenariats reposent sur la mise en place de thèses en cotutelles, projets de recherches communs, d'accueil de visiteurs ou de séjours dans des laboratoires étrangers. Plusieurs projets en cours nous permettent de développer nos activités de recherche avec des laboratoires très connus dans notre domaine comme l'INRIA, Telecom SudParis, Eurecom, EPFL Suisse, University of Illinois (USA), IISC India, Université de Montréal, Ecole Polytechnique de Montréal, University of Concordia, Université Mohammed V Rabat Maroc, Université Abdelmalek Essassidi Tanger Maroc, ENSI/Université de Menouba Tunis Tunisie, University of Wisconsin, Saint-Petersburgh State University, CNRS/LASS, LIRMM, etc..

Nous continuerons à garder des partenariats avec les grands opérateurs de télécommunication et les entreprises développant des services dans les réseaux. Les collaborations avec Alcatel-Lucent, Orange Labs, CREATE-NET et Microsoft, nous permettent d'avoir des cadres applicatifs réels et de travailler sur des applications issues de services mis en fonction auprès du consommateur des réseaux.

Les activités du LIA pour la thématique *Réseaux* portent de façon générale sur les architectures de réseaux sans fil. Dans ce cadre, nos activités vont être organisées sur plusieurs axes principaux :

- Développer des outils mathématiques et des outils logiciels pour évaluer, optimiser, et contrôler les systèmes complexes et en particulier les applications dans les réseaux ;
- Développer des algorithmes distribués pour les réseaux autonomes ;
- Développer des plate-formes d'expérimentation et une méthodologie d'évaluation pour les nouvelles architectures de réseaux.

Dans les cinq prochaines années, nous poursuivrons a priori nos recherches sur les thèmes mentionnés dans le bilan 2006-2011. Ces thèmes de recherche sont importants pour la société et engendrent de nombreux défis scientifiques. De plus, ces projets pluridisciplinaires montrent la grande variété de domaines auxquels s'intéresse notre groupe et les interactions que nous pouvons avoir avec d'autres équipes en économie, en biologie, en communications numériques et en mathématiques. Les membres de la thématique « Réseaux » participent activement sur plusieurs projets transversaux au sein du LIA. Nous continuerons dans cette voie en proposant des projets pluridisciplinaires susceptibles de fédérer les thématiques du LIA autour des problèmes liés à l'analyse des contenus, des infrastructures, de la sécurité et des usages.

13.3.2 Orientations scientifiques

Le développement des nouvelles architectures de réseau, ou de nouveaux protocoles et mécanismes sur ces réseaux, nécessite un cycle de vie incluant plusieurs étapes dont la modélisation, la vérification, la conception, le test, l'expérimentation et l'évaluation afin de montrer leur faisabilité, leur bon fonctionnement et/ou leur supériorité par rapport à l'existant. Dans le domaine des réseaux, la simulation et l'expérimentation sur des plate-formes sont souvent utilisées, mais de façon individuelle et différente par chaque équipe de recherche. Nous nous intéressons à définir une méthodologie qui pourrait servir à toute la communauté, garantissant la reproductibilité des expériences (en plate-forme ou en simulation). Pour cette nouvelle activité, nous allons utiliser des outils qui sont développés et utilisés depuis de nombreuses années dans d'autres domaines comme le «experimental design» et les statistiques, mais qui restent encore méconnus du monde de la recherche en réseaux. L'objectif est assez ambitieux et, même si le résultat n'était finalement que peu utilisé par la communauté, les étapes vers la définition d'une telle méthodologie vont nous permettre de renforcer notre phase d'évaluation et d'expérimentation pour toutes nos activités dans les autres axes de recherche.

Nos perspectives portent sur plusieurs thèmes de recherche :

a) Sécurité dans les réseaux sans fil

Nous souhaitons contribuer au niveau de sécurité dans les réseaux autonomes et réseaux de véhicules en développant de nouvelles architectures distribuées et hiérarchiques permettant l'authentification sécurisée,

la gestion des clés et des certificats, de façon hiérarchique où les nœuds seront organisés en clusters. Chaque cluster dispose d'un clusterhead élu avec un certain niveau de confiance et autour duquel une certaine protection doit prendre place. Ensuite, nous nous intéresserons à la conception inter couches MAC et Routage pour la détection des nœuds malveillants avec différentes attaques, pour la réaction contre ces nœuds et pour la coopération entre les nœuds : systèmes de notation et punition.

...dans les réseaux de véhicules

Nous travaillerons au développement d'un nouveau modèle de confiance dans les réseaux de véhicules où les entités sont les UB (véhicules), les RSU (Road Side units), les fabricants et les autorités. Un mécanisme de surveillance et de notation permettra aux véhicules de bien se comporter en annonçant une information vraie et non falsifiée. Un conducteur malhonnête aura tendance à diffuser des informations fausses concernant la situation des routes, par exemple, présence de verglas, de congestion, etc., pour se faire libérer la route. Ainsi, toute annonce sur l'état de la route sera authentifiée. Le système, ainsi constitué, nécessite l'authentification du conducteur ce qui peut être néfaste pour sa vie privée. Nous souhaitons proposer un système d'authentification anonyme entre véhicules.

...dans les réseaux de capteurs

Nous nous intéresserons à la conception inter-couches pour la sécurité dans les réseaux 6LoWPAN (Low Power Personal Area Networks). Un réseau 6LoWPAN est constitué d'un ensemble d'équipements (capteurs), avec de faibles capacités en ressources, connectés à Internet avec du bas débit via le protocole IPv6. Plusieurs couches OSI seront impliquées dans la sécurisation des transferts des données afin d'assurer la disponibilité et l'intégrité et la confidentialité des données transmises sur un réseau 6LoWPAN : IP, MAC et application. AES à lui tout seul au niveau 2 (IEEE 802.15.4) n'est pas suffisant, IPSec doit être allégé et plusieurs mécanismes au niveau application seront étudiés (fragmentation, cryptage, surveillance). La sécurisation des protocoles de gestion de communication inter-réseaux et inter-opérateurs est un problème important.

b) Réseaux autonomes

Les nouvelles architectures de réseaux sans fil seront basées sur des systèmes en autogestion et auto-adaptation afin de venir à bout de la complexité grandissante de réseaux et permettre leur expansion future au-delà de leur taille actuelle. Le but est de limiter l'intervention de l'homme et rendre le système plus autonome. Pour cela, nous souhaitons développer des fonctions d'autonomie aux réseaux et aux services et également aux protocoles. Ceci dans le but qu'ils puissent interagir entre eux de manière indépendante, sans le « soutien de l'homme ». Pour apporter des solutions à ces problèmes, la méthodologie utilisée consiste à proposer des algorithmes distribués qui prennent en considération les contraintes d'implémentation dans les réseaux existants ou ceux du futur. Nous continuerons à développer des algorithmes d'apprentissage automatique et la théorie de la psychologie animale utilisée par les biologistes. Par ailleurs, nous continuerons à contribuer à la localisation dans les réseaux de capteurs en proposant des solutions efficaces permettant une meilleure précision en milieu indoor. Nos propositions, dans ce cadre, ont permis d'obtenir de meilleures solutions en outdoor en considérant des capteurs équipés de différentes techniques RSSI, TDOA et AOA en utilisant de la RF. Cependant, ces techniques restent inappropriées dans le cas indoor. Nous pensons combiner de l'ultrason avec la RF et utiliser des algorithmes de correction basés-filtre tout en respectant les contraintes d'énergie pour le calcul et le routage.

c) Réseaux Ad-hoc

Dans cet axe de recherche, nous nous intéresserons essentiellement aux protocoles pour le *multicast*, le routage épidémique, la sécurité, la coopération et la garantie de la QoS « Qualité de service » dans un environnement de réseaux hétérogènes. Il s'agit, plus particulièrement des réseaux Ad-hoc, de capteurs mobiles, des réseaux Ad-hoc à connexion intermittente et des réseaux de véhicules.

L'explosion du trafic mobile sur les réseaux d'opérateurs est telle que la capacité n'est plus suffisante pour proposer des offres « data » illimitées. Nous nous sommes intéressés à la mise en œuvre de techniques issues des réseaux tolérants aux délais DTN (Delay Tolerant Networks) et les réseaux véhiculaires (VANET) afin de fournir des services de distribution de contenus, déchargeant les réseaux d'infrastructure. Ces techniques permettent également à des nœuds mobiles de s'échanger de l'information au gré de leurs rencontres avec des technologies sans licence telles que le Wi-Fi. Ces nouvelles capacités sont utiles lorsque l'infrastructure est saturée (e.g., lors d'une manifestation culturelle) ou inexistantes (e.g. Dans le métro) dans le cadre d'échanges avec des nœuds présents dans le voisinage. Les réseaux DTN adoptent une logique «*store and forward*», qui permettent d'augmenter la tolérance aux problèmes liés à une connectivité intermittente et aux liens à latence et taux d'erreur binaire élevés. Ce type de réseau ouvre des

perspectives intéressantes pour plusieurs applications dans la culture, le trafic routier, les zones où les infrastructures de télécommunications sont peu développées, etc. Le but de cet axe est de développer des techniques DTN permettant d'utiliser la mobilité des utilisateurs comme support de dissémination de données afin de décharger les réseaux d'infrastructure et d'assurer la capacité d'échanger efficacement des données de manière opportuniste avec les autres nœuds. Nous envisageons d'utiliser cette technologie pour décharger les réseaux cellulaires 3G et de réseau Hot-Spots, en utilisant des techniques DTN. Les technologies Ad-hoc comme le Wi-Fi permettant d'utiliser la mobilité comme un vecteur de transport de l'information. Puisque l'université d'Avignon est très impliquée dans le festival d'Avignon et d'autres festivals dans la région, notre équipe vise à développer une application, appelée MobileFaun, qui permettra une communication de masse basée sur la technologie DTN pendant des événements médiatiques comme les manifestations culturelles. Nous envisageons de mettre en place cet outil notamment durant le Festival d'Avignon. Cette application offrira donc un fort potentiel pour faire, par exemple, de la publicité pour des spectacles ou autres dans ce type d'événements culturels où de nombreuses personnes utilisant ou possédant un téléphone portable se croisent dans une zone bien définie. Cette plate-forme va nous permettre d'évaluer les mécanismes de routage dans DTN et la capacité du réseau pour acheminer le trafic. La sécurité sera au cœur de nos recherches dans cet axe. En particulier, nous nous intéresserons à la conception inter couches MAC et routage pour la détection des nœuds malveillants avec différentes attaques, pour la réaction contre ces nœuds et pour la coopération entre les nœuds : systèmes de notation et punition.

d) La gestion des ressources et de l'énergie dans les réseaux

La gestion des ressources et de l'énergie dans les réseaux, notamment dans les réseaux de communication sans fil, est devenue un des thèmes prioritaires de notre domaine de recherche. Les architectures de réseau basées sur les technologies de radio cognitive en sont de bons exemples. Avec l'ouverture à l'utilisation de nouvelles plages de transmission (comme les fréquences de la télévision hertziennes) ou la possibilité de transmettre sur des fréquences peu utilisées (radar), un système radio intelligent pourra choisir sa propre fréquence de transmission. De récentes mesures effectuées sur l'utilisation du spectre montrent un usage très limité de certaines fréquences sous licence, pour certaines zones géographiques. Il existe de nombreuses définitions d'un système radio intelligent qui diffèrent suivant l'emploi de tel système. On peut trouver une définition commune qui est : « Un système radio intelligent est un système radio qui possède des mécanismes de contrôle lui permettant une adaptation autonome et en temps réel à son environnement. » On peut donc résumer les trois propriétés d'un système radio intelligent. *Observation* : directement ou indirectement, le système radio est capable d'acquérir des informations à propos de son environnement *Adaptabilité* : le système radio est capable de changer ses paramètres de transmission. *Intelligence* : le système radio adapte ses paramètres de transmission suivant un objectif précis et de façon intelligente. Nous nous intéressons particulièrement à définir des mécanismes d'observation optimaux pour les mobiles équipés de radios intelligentes avec pour objectif des critères de performance avec prise en compte de l'énergie. En effet, la capacité des radios de pouvoir mesurer l'utilisation d'un canal donne des informations très importantes et qui permettent de prendre des décisions de façon optimale. Ces décisions peuvent porter tout d'abord sur l'utilisation ou non d'un canal particulier, la puissance de transmission, etc. Pour toutes ces études, nous utilisons des méthodologies issues de la théorie du contrôle (MDP) et de la théorie des jeux (car les radios intelligentes sont en compétition pour l'accès aux ressources libres). Ces méthodologies seront aussi utiles pour étudier des problèmes de gestion des ressources dans les réseaux de transport. Nous avons répondu conjointement avec des membres de la thématique RO, à un appel à projet Transport Terrestre Durable de l'ANR qui traite de problèmes d'optimisation de l'énergie dans les réseaux de transport avec des modèles de jeu hiérarchique et des modèles de choix discrets.

e) Réseaux de capteurs Véhiculaires

Les réseaux de capteurs véhiculaires sont des réseaux qui combinent entre capteurs sans fil (stationnaire ou mobiles), des véhicules et éventuellement une infrastructure (Internet et/ou cellulaire). Les enjeux sont importants : optimisation de la sûreté, de la sécurité, du confort des usagers de la route en environnement autoroutier ou urbain.

D'une part, nous souhaitons contribuer aux architectures des réseaux, à la collecte des données, à la gestion de la communication et à l'agrégation des données. Les informations de sécurité utiles doivent être triées des informations de confort et traitées en priorité. Les protocoles MAC IEEE 802.11p et DSRC dans l'environnement WAVE doivent être étudiés conjointement avec 802.15.4 pour les capteurs.

D'autre part, avec des membres de l'équipe RO, nous avons proposé un projet ANR pour l'optimisation du trafic et des cycles de feux. Habituellement, la programmation des cycles de feux se faisait de manière statique. Dans ce projet, nous proposerons une optimisation dynamique. Pour cela, nous utiliserons de nouvelles technologies qui sont les capteurs sans fil communiquant avec les véhicules et l'infrastructure. Le travail commence d'abord par la collecte des données et ensuite l'optimisation du trafic.

13.3.3 Analyse des opportunités et risques

La thématique réseaux est très complémentaire des autres thématiques du laboratoire ; cette **complémentarité** la met en situation de collaborer à la fois avec la thématique Recherche Opérationnelle, avec laquelle un projet ANR commun démarrera fin 2011, mais aussi avec la thématique langage, avec laquelle plusieurs projets sont en phase d'évaluation à l'ANR. De plus, la thématique pourrait étendre ses **collaborations industrielles** sur la conception de plate-formes expérimentales, la définition des politiques de brevets et standards. Ce potentiel de collaboration est une des opportunités que nous devons saisir dans les années à venir.

Par ailleurs, le contexte scientifique et technologique est au développement du Web est des usages mobiles, voir géolocalisés. En particulier, la mobilité et la connectivité requièrent la conception de nouveaux protocoles pour les réseaux autonomes hétérogènes. Nous avons la possibilité scientifique de répondre à de nombreux appels à projets sur ces thèmes qui sont au cœur du projet "société numérique". Ainsi, notre priorité sera de consolider cet axe pour nous mettre en situation de répondre à de tels besoins.

Les principaux risques sont liés à la **dispersion thématique** que ces collaborations pourraient entraîner et à la **taille réduite de la thématique**, qui la fragilise et qui pourrait limiter sa capacité de diversifier ses thèmes de recherche.