

# Rapport de fin de projet

## DEMI-TON

### Description multimodale

### pour la structuration automatique

### des flux de télévision

## 1 Liste des équipes impliquées

**TEXMEX** : Techniques d'exploitation des données multimédias, UMR 6074 IRISA, Rennes

**METISS** : Modélisation et expérimentation pour le traitement des informations et des signaux sonores, UMR 6074 IRISA, Rennes

**DCA** : Description des contenus audiovisuels, Institut national de l'audiovisuel, Bry sur Marne

## 2 Liste des participants

### TEXMEX

- Gros, Patrick, directeur de recherche INRIA, 40 %
- Sébillot, Pascale, professeur INSA de Rennes, 35 %
- Kijak, Ewa, maître de conférences université de Rennes 1 depuis le 1/9/2006, 10 %.
- Naturel, Xavier, doctorant, contrat de recherche doctorale INRIA, 50 %, début de la thèse 1/2/2004, soutenue le 17/4/2007.
- Huet, Stéphane, doctorant, allocataire MENESR, 50 %, début de la thèse 1/10/2004, soutenue le 11/12/2007.
- Lecorvé, Gwénoélé, doctorant, allocataire MENRT, 50 %, début de la thèse en octobre 2007.
- Dupuis, Arnaud, post-doctorant payé par l'ACI du 1/5/2005 au 30/11/2005, 100 %, ingénieur expert INRIA payé sur un contrat du FNADT depuis le 1/12/2005, 100 %.
- Dufouil, Cédric, Ingénieur associé INRIA (payé sur subvention INRIA), 100 %, du 1/9/2005 au 31/8/2007.

### METISS

- Gravier, Guillaume, chargé de recherche CNRS, 75 %
- Moraru, Daniel, post-doctorant, du 1/2/05 au 30/4/05 sur financement propre, payé par l'ACI du 1/5/2005 au 31/1/2006, 100 %.
- Cauchy, Pierre, ingénieur associé INRIA (payé sur subvention INRIA), 100 %, depuis octobre 2007.

### DCA

- Brunie, Vincent, chercheur INA, 30 %, jusqu'à fin avril 2006.
- Carrive, Jean, chercheur INA, 30 %
- Vinet, Laurent, chercheur INA, 30 %
- Balin, Fabrice, ingénieur INA, 30 %
- Caillet, Marc, doctorant CIFRE INA, 20 %, du 25/10/2004 au 24/10/2007.
- Poli, Jean-Philippe, doctorant CIFRE INA, 100 %, début de la thèse 4/1/2004, jusqu'à janvier 2007. Soutenance le 29 mai 2007.

### 3 Changements significatifs intervenus dans le projet

- V. Brunie qui était responsable du projet pour l'INA a quitté l'INA. Il a été remplacé par J. Carrive qui était déjà impliqué dans le projet.

## 4 Avancées scientifiques dans le cadre du projet

### 4.1 Structuration de flux de télévision

Un des buts principaux du projet était de décrire des flux de télévision de longue durée (au minimum plusieurs semaines). Deux approches ont été étudiées.

**Approche prédictive.** La thèse de J.P. Poli à l'INA a adopté un point de vue top-down, en partant de l'observation des guides de programmes, ceux, de type prévisionnel, publiés dans la presse spécialisée et ceux établis par l'INA à partir de l'observation du flux réel. En modélisant les grilles de programmes d'une chaîne et en entraînant le modèle sur les grilles des années passées, il est possible de déduire tous les enchaînements possibles d'émissions, y compris celles qui ne sont pas annoncées dans les guides. Le modèle est constitué d'un modèle de Markov contextualisé, pour les transitions d'un programme à l'autre, et d'un arbre de régression, utilisé pour prédire la durée d'une émission. Ce modèle permet de prédire à quelle sorte de transition il faut s'attendre, afin de savoir quel objet doit être détecté (un jingle de publicité, un générique d'émission) et surtout à quel endroit du flux il doit être recherché. Ainsi les détections sont effectuées localement dans le flux et il a été montré expérimentalement que dans le pire des cas, moins de la moitié du flux était nécessaire pour le structurer. Ce travail a été réalisé en utilisant les guides de programme correspondant à un flux de plusieurs années de télévision.

**Approche descriptive.** La thèse de X. Naturel à l'IRISA a pris le point de vue inverse : à partir d'un guide programme prédictif (issu de la presse spécialisée) et du flux lui-même, on a cherché à identifier les différents programmes et inter-programmes dans le flux en détectant leurs bornes et en étiquetant les segments ainsi délimités. La méthode utilisée consiste à tout d'abord détecter les inter-programmes à l'aide de marqueurs spécifiques et d'une base de tous les inter-programmes déjà détectés, puis d'aligner les segments de programmes avec ceux du guide de programme par DTW (Dynamic Time Warping).

**Perspectives.** La deuxième approche développée nécessite un étiquetage manuel d'une journée de flux, ce qui représente une contrainte très forte dans la pratique. Dans le cadre du projet ANR Semim@ges, nous reprenons cette méthode pour supprimer cette contrainte. L'idée est de se baser sur une détection de toutes les répétitions dans le flux pour identifier les inter-programmes. On doit pouvoir alors identifier les programmes de manière identique à ce qui a été fait dans le présent projet. Par ailleurs, cette segmentation basée image sera couplée avec une segmentation basée sur la détection des répétitions de motifs sonores.

Une deuxième piste est de coupler ce processus de segmentation basée sur la parole et le traitement de la langue naturelle. Suite aux travaux effectués par Stéphane Huet et Gwénoél Lecorvé, nous devrions pouvoir disposer d'une transcription utilisable de la parole, et pouvoir segmenter celle-ci en plages de thème homogène. Une telle information est très complémentaire par rapport aux informations dont nous disposons avec le son et l'image. L'enjeu est de trouver un outil méthodologique d'intégration.

### 4.2 Traitement de la bande sonore

Plusieurs aspects du traitement de la bande sonore dans les flux télévisuels ont été étudiés au cours du projet, en particulier autour de la transcription automatique et de la segmentation en locuteur.

**Analyse du flux de parole télévisuelle.** Dans un premier temps, nous avons complété et amélioré la plate-forme IRENE pré-existante, adaptée à la transcription des informations diffusées à la radio. Nous avons d'une part ajouté un module permettant la structuration du flux selon le locuteur et étudié l'apport du suivi de locuteur pour améliorer la segmentation. D'autre part, nous avons exploité cette segmentation pour améliorer la transcription par l'ajout d'un module d'adaptation au locuteur. Cette nouvelle version de IRENE a ensuite été évaluée sur le flux télévisuel, mettant ainsi en évidence le manque de robustesse des modèles appris sur la radio devant la qualité sonore moindre de la télévision. Le travail s'est ensuite orienté sur l'utilisation de points d'ancrage macrophonétique pour améliorer cette robustesse et s'affranchir de la nécessité de disposer d'un grand corpus annoté de parole télévisuelle pour apprendre des modèles. Ces derniers travaux ont montré un fort potentiel pour améliorer la transcription automatique et suscitent des recherches à venir concernant la détection automatique de points d'ancrage macrophonétique (ISCA Intl. Workshop on Non-Linear Speech Processing, 2007).

Par ailleurs, de nombreuses améliorations ont été amené à la plate-forme IRENE au cours du projet permettant ainsi d'envisager son utilisation dans le cadre de la structuration de flux multimedias. Afin de favoriser l'intégration de la transcription automatique au processus de structuration audiovisuel, nous avons démarré l'intégration des différentes composantes du système de transcription au sein de la plate-forme d'indexation audiovisuelle, en y incluant les résultats des recherches effectués dans le cadre du projet Demi-Ton. Une première version de la plate-forme sera diffusée en août 2008.

**Traitement automatique des langues et reconnaissance de la parole.** Dans un premier travail, mené autour de la thèse de S. Huet, le lien entre reconnaissance automatique de la parole (RAP) et traitement automatique des langues (TAL) a été étudié. Le but était double : voir comment les techniques de TAL peuvent améliorer la transcription de la parole, et évaluer comment les techniques de TAL peuvent traiter la parole transcrite. Le premier aspect a donné lieu à un travail intéressant sur l'utilisation d'un tagger statistique PoS (Part of Speech = parties du discours) pour corriger les erreurs de la transcription : de nombreuses variantes morphologiques d'un même mot sont en effet homophones et ne peuvent être distinguées par un système de RAP classique. La prise en compte des caractéristiques PoS (genre, nombre, type du mot : nom, verbe...) permet une désambiguïsation dans de nombreux cas, ce qui se traduit par une amélioration sensible du taux d'erreur sur les mots (réduction d'environ 1 point). En outre, nous avons mis en évidence que l'utilisation d'informations morphosyntaxique permet d'améliorer les mesures de confiances fournies par le système de transcription pour chaque mot reconnu (Text, Speech and Dialogue, 2006; Intl. Conf. on Spoken Language Processing, 2007).

**Segmentation thématique des flux de parole.** Une première étape de structuration exploitant la transcription automatique de la parole concerne la segmentation thématique. En effet, de nombreuses émissions abordent successivement plusieurs sujets, ce qui leur confèrent une structure thématique inhérente qu'il est intéressant de détecter. Dans le cadre de la thèse de S. Huet, nous nous sommes donc intéressés à ce problème pour lequel un modèle statistique combinant de multiples sources de connaissance a été développé. Le modèle proposé se base en grande partie sur la notion de cohésion lexicale, *i.e.* sur le fait que la cohérence du vocabulaire utilisé au sein d'un même thème (répétition de même mots ou de mots sémantiquement reliés). En complément de cette information lexicale, nous avons également considéré des informations syntaxiques (séquences d'étiquette morphosyntaxiques) et acoustique (alternance de locuteurs, pauses, présence de musique), qui permettent une segmentation thématique exploitable (Traitement Automatique des Langues Naturels, 2008).

**Adaptation automatique de la transcription de parole dans les flux radiophoniques et télévisuels**  
Les systèmes de transcription automatique se basent sur un modèle de langage (ML) appris sur une vaste collection de textes aux sujets variés. Si ces ML synthétisent une bonne connaissance générale de la langue et permettent d'obtenir des performances globalement satisfaisantes, ils restent cependant peu adaptés pour transcrire des documents très marqués thématiquement. De plus, on est en droit d'attendre d'un modèle de langage adapté au sujet traité qu'ils permettent d'obtenir une meilleure transcription. En raison du nombre

élevé de thèmes pouvant être abordés dans un flux télévisé, nous avons étudié une technique d'adaptation non supervisée du ML à partir de techniques de recherche d'information permettant la récupération de documents thématiquement proche sur Internet (Intl. Conf. on Acoustics, Speech and Signal Processing, 2008). La technique proposée se base sur l'extraction de mots-clés dans une transcription initiale afin de caractériser un segment thématique. Cette caractérisation par mots-clés ouvrent de nombreuses perspectives concernant la génération automatique d'index, l'enrichissement des contenus par ajout de lien vers des documents pertinents ou encore la vérification du contenu par rapport aux informations de la grille de programme. Ces perspectives feront l'objet d'un travail de thèse dans le cadre du projet Quaero.

### 4.3 Langage de description pour l'audiovisuel

**FDL : Feria Description Language** La plate-forme de développement d'applications multimédias Feria, développée à l'Ina en concertation avec l'IRISA, repose en grande partie sur un langage pivot, appelé FDL (pour Feria Description Language). Ce langage, issu de projets précédents, a fait l'objet d'une refonte complète. Cette refonte a été réalisée par Marc Caillet dans le cadre de sa thèse effectuée à l'Ina en partenariat avec le projet WAM de l'INRIA Rhône-Alpes. L'objectif de ce travail était d'implémenter un langage de description permettant le stockage et la manipulation de données de description de documents audiovisuels, descriptions pouvant provenir soit de processus automatiques soit de processus manuel, en faisant reposer la définition de ce langage sur un paradigme de programmation par objets plutôt que sur une modélisation XML, comme c'est le cas par exemple pour MPEG-7. Ont ainsi été mis en place des mécanismes de définition des classes de descripteurs, d'instanciation, ainsi que des mécanismes de réflexion et d'introspection permettant de manipuler descripteurs et classes de descripteurs de manière générique. Enfin, une syntaxe XML a été définie pour le stockage en base de données. L'implémentation repose sur le framework .NET de Microsoft, et la sérialisation XML se fait à l'aide de la base de données MS SqlServer 2005. Une nouvelle implémentation du langage FDL est à présent à l'étude, reposant sur les mêmes principes de programmation par objets mais tirant un meilleur profit des fonctionnalités du framework .NET en termes de métalangage et de sérialisation XML. D'autre part, d'autres formats de sérialisation sont également à l'étude : sérialisation XML dans un système de fichiers, sérialisation en format texte tabulé et sérialisation binaire, afin de mieux correspondre aux besoins des outils d'analyse automatique.

## 5 Plates-formes d'indexation audiovisuelle

Au delà des échanges purement scientifiques, le projet DEMI-TON a favorisé de nombreux échanges technologiques entre les deux équipes. Initialement, chaque entité avait amorcé le développement d'une plate-forme informatique adaptée au traitement de documents multimédias. Dans le cadre de DEMI-TON, les deux équipes ont déterminé des points de convergence indispensables au partage et à la portabilité d'algorithmes de traitement entre les deux plates-formes.

**Projet de plate-forme multimédia** Les activités de recherche sur la structuration automatique de flux télévisés nécessitent de disposer de corpus audiovisuels significatifs représentant plusieurs semaines de programmes. Sachant que 24 heures de vidéo correspondent approximativement à 40 Go de données, les enregistrements de plusieurs semaines peuvent alors occuper des espaces disque de plusieurs To. À ces informations s'ajoutent des données de description (grille des programmes, résumé...) ainsi que les résultats des outils de traitement (mouvement de caméra, suivi de visages...).

Par ailleurs, les outils d'analyse de vidéos étant souvent gourmands en ressources processeur, le traitement de ces gros volumes de données requière de disposer de puissances de calcul dimensionnées en conséquence. A ces problèmes de stockage et de calcul, s'ajoute la protection des droits d'auteurs sur les données audiovisuelles réglementant leur diffusion ainsi que leur duplication. Enfin, les natures hétérogènes des informations à traiter (image, son, texte) pose de nombreux problèmes de synchronisation et de précision d'accès aux données. Typiquement, l'accès aléatoire à une image (rechercher une image précise sans lire les images précédentes) au sein d'un flux vidéo MPEG, pourtant primordiale dans un processus de développement d'algorithmes

d'analyse automatique de vidéos, peut rapidement s'avérer être complexe à mettre en oeuvre. Pour répondre à toutes ces contraintes techniques (stockage, limitation des droits d'accès, optimisation des calculs...), il apparaît donc nécessaire de développer une infrastructure informatique adaptée au traitement de données audiovisuelles.

**Avancement du projet.** Depuis mai 2005, une plate-forme de traitement de vidéos a été développée en collaboration avec l'INA. Cette plate-forme est d'une part destinée à favoriser les échanges d'outils de traitement entre l'INA et l'IRISA, et d'autre part, conçue pour répondre à de nombreux problèmes récurrents liés aux traitements de vidéos.

Une première phase de développement a été effectuée à l'IRISA de mai à novembre 2005 et a consisté à étudier et à développer un serveur centralisant des corpus audiovisuels pouvant représenter plusieurs semaines de programmes télévisés. Ce serveur devait permettre de programmer des acquisitions de séquences télévisées, de les diffuser en « streaming », de gérer les vidéos propres à chaque utilisateur et enfin de protéger les corpus en contrôlant les accès par le biais de login et mot de passe. L'accès à la plate-forme, via un outil baptisé « DIVA manager », s'effectue par le biais d'une interface web fonctionnant sous de nombreux systèmes d'exploitation (Windows, Linux et MacOSx).

Une seconde phase de développement a débuté en septembre 2005 et a consisté à développer une application client/serveur offrant un accès précis aux images et aux sons extraits de contenus audiovisuels. Cette application a été conçue dans l'objectif de simplifier le développement d'outils de traitement vidéo et propose aux utilisateurs de faire abstraction des nombreux problèmes liés au décodage de séquences audiovisuelles. L'étude et le développement de l'application « DIVA client/serveur » a contribué à lever de nombreux verrous techniques liés à la précision d'accès ou encore à la synchronisation des données.

L'ensemble de cette solution logicielle, intégrant « DIVA manager » et « DIVA client/serveur » a été baptisée « DIVA solution ».

La dernière phase du développement a porté sur l'exploitation des résultats obtenus à l'aide de « DIVA solution » et l'amélioration des captures.

Pour permettre la visualisation des résultats et la création des vérités terrains, le logiciel « Navitex » basé sur la solution « DIVA client » a été développé. Son interface graphique adaptée à la navigation dans de grand volume de données facilite la saisie de vérité terrain comme la visualisation des résultats. Comme pour « DIVA Solution », Navitex est disponible sur plusieurs systèmes d'exploitation (Windows, Linux et MacOSx).

Cette dernière phase a également permis d'améliorer l'acquisition des flux audiovisuels pour profiter des avantages de la diffusion numérique TNT. L'enregistrement bénéficie désormais d'une meilleure définition et qualité d'image, des flux de sous-titrage, du teletext, des métadonnées de programmation diffusées dans les flux et également des pistes sonores multilingues. Ce passage à l'acquisition numérique a également rendu possible la parallélisation des captures par utilisateur et par chaîne TV. L'interface web de l'outil « DIVA manager » a également été enrichi d'un guide des programmes en ligne pour faciliter la programmation des captures.

**Valorisation des activités de TEXMEX associées à la plate-forme.** L'application « DIVA client / serveur » (distant video access), développée à l'IRISA, a fait l'objet d'un dépôt de code au près de l'APP (agence pour la protection des programmes), elle est référencée sous le numéro : IDDN.FR.001.320006.000.S.P.2006.000.40000. Ce logiciel a été diffusé à plusieurs partenaires dans le cadre de collaborations : à l'équipe DCA de l'INA, à l'IRIT de Toulouse, à la société Kersonic. Il devrait être utilisé dans le cadre du programme de recherche Quaero soutenu par OSEO. Parallèlement, un article décrivant le système ainsi que les problèmes techniques résolus par le logiciel a été publié lors de la semaine du document numérique (SDN'06) dans le cadre de l'atelier SIAV (Systèmes d'Information AudioVisuels).

## 6 Bilan et perspectives

### 6.1 Avancées scientifiques

Le travail mené au sein du projet DEMI-TON a lancé un nouvel axe de recherche. En dehors d'un ou deux articles, ce sont les premiers travaux à étudier en profondeur l'analyse de longs flux de télévision. De nombreuses parties du flux avaient fait l'objet d'études : la détection de publicités, l'analyse des journaux télévisés, des retransmissions sportives. Mais aucun n'avait abordé la structuration complète des flux. Cette question n'a pas encore été incluse dans les campagnes d'évaluation telles TRECVID, mais son intérêt applicatif, tel que souligné dessous par l'INA est réel.

La travail entamé au sein de DEMI-TON a plusieurs suites en cours. Au sein du projet ANR Semimages d'une part, et au sein du programme Quaero. L'objectif est de rendre la chaîne de structuration la plus automatique possible vu les volumes de données concernées, et d'autre part de pousser l'analyse aussi loin que possible : reconnaissance des programmes, détermination des thèmes abordés dans ces programmes, caractérisation des inter-programmes.

### 6.2 Plate-forme

Un des enjeux importants pour pouvoir aborder le problème de la structuration des longs flux est de pouvoir disposer de tels flux et de pouvoir les exploiter. Cela exige de pouvoir les capter, les enregistrer et y accéder. Une partie importante de l'effort du projet a été dédié à ces aspects. Les briques logicielles développées ont déjà donné lieu à plusieurs transferts, de nombreuses équipes travaillant sur l'indexation de la vidéo ayant les mêmes difficultés sans avoir les moyens de développer une solution logicielle. Pour le moment, cette diffusion a été limitée aux collaborations que nous avons. Une diffusion plus large serait envisageable, mais demande une infrastructure plus lourde.

### 6.3 Publications, thèses

Le projet DEMI-TON a donné lieu à la soutenance de trois thèses. Xavier Naturel a obtenu une mention spéciale du prix de la recherche 2007 de l'Inathèque de France pour ses travaux.

Par ailleurs, les travaux menés dans le projet ont donné lieu à la publication de deux livres (relatant d'autres travaux aussi bien-sûr !), 5 chapitres de livre, deux articles dans des revues internationales et deux dans des revues nationales, 15 communications dans des conférences internationales et 9 dans des conférences nationales.

### 6.4 Suite de la collaboration

La suite la plus notable de cette collaboration sera le programme de recherche Quaero, qui regroupe les trois équipes de DEMI-TON avec de nombreux autres partenaires. La tâche 10.1 du Core Technology Cluster de Quaero sera consacrée à la suite de DEMI-TON, l'INA fournissant un corpus et récupérant les résultats obtenus via le projet applicatif "Digital Multimedia Asset Management" co-dirigé par l'INA et Thomson. Ce programme a débuté le 1<sup>er</sup> mai 2008 pour une durée de 5 ans.

Un autre projet prenant la suite a été proposé cette année à l'ANR dans le cadre des programmes internationaux de l'appel à projet du "programme blanc", avec deux équipes partenaires japonnaises. La première est celle d'Hirochi Murase à Nagoya, la seconde est celle de Shin'Ichi Satoh au NII de Tokyo. Le but de ce projet est de développer un système de recherche d'information audio-visuelle.

### 6.5 Intérêt applicatif

L'Inathèque de France, pour sa mission de dépôt légal de la télévision et de la radio, capte en permanence plusieurs dizaines de chaînes de radio et de télévision. Les émissions à documenter doivent tout d'abord être temporellement repérées dans le flux, ce qui constitue une tâche fastidieuse. Dans le but d'automatiser

cette tâche, l'intérêt applicatif du projet DEMI-TON est immédiat. On peut de plus noter que certaines des méthodes de structuration temporelle développées dans le cadre du projet ont une application à un niveau de granularité temporelle plus faible, c'est-à-dire au niveau d'éléments de grilles de programme dont la structure temporelle est régulière.

## 7 Réunions et Conférences organisées dans le cadre du projet

- Début du projet : 1er avril 2005
- Réunion plénière du projet : 11 octobre 2005, Rennes
- Séminaire technique : du 24 au 28 octobre 2005, Bry sur Marne
- Réunion plénière du projet : 30 janvier 2006, Paris
- Réunion plénière du projet : 6 juin 2006, Rennes
- Séminaire technique : 3 et 4 juillet 2006, Bry sur Marne
- Réunion plénière du projet : 28 novembre 2006, Rennes
- Soutenance de thèse de Xavier Naturel. Jean Carrive était examinateur au sein du jury.
- Soutenance de thèse de Jean-Philippe Poli. Patrick Gros était rapporteur de la thèse.
- Réunion plénière du projet : 18 décembre 2007, Rennes
- Soutenance de thèse de Stéphane Huet.
- Réunion de fin de projet : 3 avril 2008, Rennes.

## 8 Soutiens obtenus en liaison avec ce projet

### 8.1 Postes chercheurs

**Xavier Naturel** : doctorant, contrat de recherche doctorale INRIA, début de la thèse 1/2/2004.

**Jean-Philippe Poli** : doctorant, convention CIFRE, début de la thèse 4/1/2004.

**Stéphane Huet** : doctorant, allocation MENRT.

**Gwénolé Lecorvé** : doctorant. allocation MENRT.

### 8.2 Postes ingénieurs

**Arnaud Dupuis** : après son séjour post-doctoral payé sur les crédits obtenus dans le cadre du projet DEMI-TON, A. Dupuis a été embauché comme ingénieur de niveau post-doctoral par l'INRIA grâce à un financement octroyé par le FNADT (Fonds national d'aménagement et de développement du territoire). Cette embauche a été effective le 1er décembre 2005 pour une durée de 3 ans.

**Cédric Dufouil** : a été embauché par l'INRIA (sur sa dotation budgétaire) pour deux ans le 1er septembre 2005 comme ingénieur en CDD.

**Pierre Cauchy** : a été embauché par l'INRIA (sur sa dotation budgétaire) pour deux ans le 1er octobre 2007 comme ingénieur en CDD.

### 8.3 Contrats nationaux

**Semim@ges** : 2007 - 2009, contrat ANR RIAM, financement de 133000 EUR pour l'équipe TEXMEX.

**QUAERO** : 2008 - 2012, contrat A2I, financement total 100 000 000 EUR (subvention et avance remboursable), 4 300 000 EUR pour l'INRIA. Les points traités dans DEMI-TON seront repris par l'INRIA au sein de la tâche 10.1 du Core Technology Cluster et les aspect plus applicatifs au sein du projet Digital Media Asset Management codirigé par IINA et Thomson Grass Valley.

**Infom@gic** : projet du pôle de compétitivité Cap Digital financé par la DGE et plusieurs conseils généraux d’Ile-de-France, de décembre 2005 à mi 2009. Infom@gic vise à mettre en place, sur une période de trois ans, un laboratoire industriel de sélection, de tests, d’intégration et de validation d’applications opérationnelles des meilleures technologies franciliennes dans le domaine de l’ingénierie des connaissances et de l’analyse de données.

## 8.4 Contrats européens

**MUSCLE** : réseau d’excellence du 6e PCRDT Multimedia Understanding Through Semantics, Computation and Learning, 2004 - 2008. Dans ce cadre, nous menons des travaux sur la structuration multimodale des retransmissions sportives à la télévision et sur les modèles stochastiques qui permettent une telle structuration. Par rapport à nos activités dans DEMI-TON, il s’agit ici d’une structuration plus fine au niveau d’une seule émission. Il y a donc une forte complémentarité.

**K-SPACE** : réseau d’excellence du 6ème PCRDT Knowledge Space of Semantic inference for automatic annotation and retrieval of multimedia content, 2006-2008. L’objectif de ce réseau d’excellence est de bâtir un programme de recherche commun dans le domaine de l’analyse automatique et semi-automatique des contenus multimédias. Dans ce cadre ont été menés principalement des travaux sur des collections homogènes de documents. Ainsi, des démonstrateurs tirant profit de l’alignement temporel texte-parole au phonème près de six versions différentes du Misanthrope ont été développés à l’aide de la plate-forme Feria.

## 8.5 Contrats internationaux

**Équipe associée e\_TV** : dans le cadre du programme « équipes associées » de l’INRIA, nous avons créé une telle équipe commune entre TEXMEX, l’équipe d’Ichiro Ide et de Hiroshi Murase à l’université de Nagoya et celle de Shinichi Satoh au National Institute of Informatics de Tokyo. Cette équipe, d’une durée de 3 ans, a été créée au 1<sup>er</sup> janvier 2008. Le but du projet est de développer des techniques de structuration et description des flux de télévision. Financement : 20 000 EUR en 2008, dégressif ensuite.

## 8.6 Contacts internationaux dans le cadre de ce projet

**Stéphane Marchand-Maillet** : Centre universitaire d’informatique, université de Genève, Suisse.

**Ichiro Ide** : Université de Nagoya, Japon.

**Luis Villaseñor Pineda** : Laboratoire des Technologies de la Langue, Institut National d’Astrophysique, d’Optique et d’Électronique, Puebla, États Unis du Mexique.

## 8.7 Soutiens financiers

- Le FNADT nous a accordé 150 kEUR qui nous ont permis d’embaucher Arnaud Dupuis comme ingénieur en CDD pour une durée de 3 ans.
- L’université de Rennes 1 nous a accordé 28 kEUR de BQR pour participer à l’équipement de la plate-forme.
- L’INSA de Rennes nous a accordé 13,5 kEUR de BQR pour tester l’emploi d’une baie RAID au sein d’un cluster de la grille GRID5000, et 11 kEUR qui ont permis d’acheter 4 lames pour notre cluster de calcul.
- l’INRIA nous a accordé 330 493 EUR pour acheter une baie NAS et 4 serveurs de calculs et de données. Pour rappel, 85 kEUR du budget DEMI-TON ont été utilisés pour acheter du matériel pour notre cluster de calcul.

## 9 Publications obtenues dans le cadre du projet

### Livres et chapitres de livre

- *L'indexation multimédia - Description et recherche automatiques*. Sous la direction de Patrick Gros, Hermès - Lavoisier, Mai 2007.
- Fabienne Moreau, Vincent Claveau, Pascale Sébillot. Description des textes. In *L'indexation multimédia - description et recherche automatiques*, P. Gros (ed.), Chapitre 6, pages 163–190, Lavoisier - Hermès Science Publications, traité IC2, Juin 2007.
- Guillaume Gravier. Description multimodale de documents multimédias. In *L'indexation multimédia - description et recherche automatiques*, P. Gros (ed.), Chapitre 7, pages 191–214, Lavoisier - Hermès Science Publications, traité IC2, Juin 2007.
- *Multimodal Processing and Interaction : Audio, Video, Text*. Sous la direction de Petros Maragos, Alexandros Potamianos et Patrick Gros, Springer-Verlag, à paraître en 2008.
- Petros Maragos, Patrick Gros and Athanassios Katsmanis et George Papandreou. Cross-Modal Integration for Performance Improving in Multimedia : A Review. In *Multimodal Processing and Interaction : Audio, Video, Text*. Sous la direction de Petros Maragos, Alexandros Potamianos et Patrick Gros, Springer-Verlag, pp. 3–48, à paraître en 2008.
- Manolis Delakis, Guillaume Gravier et Patrick Gros. Stochastic Models for Multimodal Video Analysis. In *Multimodal Processing and Interaction : Audio, Video, Text*. Sous la direction de Petros Maragos, Alexandros Potamianos et Patrick Gros, Springer-Verlag, pp. 91–110, à paraître en 2008.
- Stéphane Huet, Gwénolé Lecorvé, Guillaume Gravier, Pascale Sébillot. Toward the Integration of Natural Language Processing and Automatic Speech Recognition : Using Morpho-syntax and Pragmatics for Transcription. In *Multimodal Processing and Interaction : Audio, Video, Text*. Sous la direction de Petros Maragos, Alexandros Potamianos et Patrick Gros, Springer-Verlag, pp. 201–220, à paraître en 2008.

### Journaux internationaux

- Xavier Naturel et Patrick Gros. Detecting Repeats for Video Structuring. In *Multimedia Tools and Application*, paru en ligne, <http://www.springerlink.com/content/f417v67462m89067/>, novembre 2007
- Manolis Delakis, Guillaume Gravier et Patrick Gros. Audiovisual Integration with Segment Models for Tennis Video Parsing. In *Computer Vision and Understanding*, 111(2) :142–154, August 2008.

### Conférences internationales

- Xavier Naturel, Patrick Gros. A Fast Shot Matching Strategy for detecting duplicate sequences in a television stream. In *CVDB'05 : Proceedings of the 2nd ACM SIGMOD International Workshop on Computer Vision meets DataBases*, Baltimore, USA, juin 2005.
- Daniel Moraru, Mathieu Ben, Guillaume Gravier. Experiments on speaker tracking and segmentation in radio broadcast news. In *Proceedings of the European Conference on Speech Communication and Technology – Interspeech*, Lisbonne, Portugal, septembre 2005.
- Jean-Philippe Poli Predicting program guides for video structuring. In *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence*, pp. 407-411, Hong-Kong, Chine, novembre 2005.
- Jean-Philippe Poli, Jean Carrive Vidéo Stream Structuring and Annotation Using Electronic Program Guides; in *Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation*, pp. 137-141, Galway, Irlande, novembre 2005.
- Xavier Naturel, Guillaume Gravier, P. Gros. Fast Structuring of Large Television Streams using Program Guides. In *Proceedings of the 4th International Workshop on Adaptive Multimedia Retrieval (AMR)*, Genève, Suisse, juillet 2006.
- Stéphane Huet, Guillaume Gravier, Pascale Sébillot. Are Morphosyntactic Taggers Suitable to Improve Automatic Transcription?. In *Proceedings of Text, Speech and Dialogue (TSD)*, Lecture Notes

- in Computer Science, Volume 4188/2006, pages 391-398, Brno, Tchéquie, septembre 2006.
- Jean-Philippe Poli et Jean Carrive. Improving Program Guides for Reducing TV Stream Structuring Problem to a Simple Alignment Problem. In *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation*, Sydney, Australie, novembre 2006.
  - Jean-Philippe Poli et Jean Carrive. Television Stream Structuring with Program Guides. In *Proceedings of the IEEE International Symposium on Multimedia (ISM2006)*, San Diego, Californie, USA, décembre 2006.
  - Jean-Philippe Poli et Jean Carrive. Modeling Television Schedules for Television Stream Structuring. In *Proceedings of the ACM International MultiMedia Modeling Conference*, Singapour, janvier 2007.
  - Marc Caillet, François Yvon, Cécile Roisin et Jean Carrive. Engineering Multimedia Applications on the Basis of Multi-Structured Descriptions of Audiovisual Documents. In *Proceedings of the International Workshop on Semantically Aware Document Processing and Indexing*, Montpellier, France, mai 2007.
  - Stéphane Huet, Guillaume Gravier et Pascale Sébillot. Morphosyntactic Processing of N-Best Lists for Improved Recognition and Confidence Measure Computation. In *Proceedings of the 8th European Conference on Speech Communication and Technology*, pages 1741–1744, Anvers, Belgique, août 2007.
  - Guillaume Gravier and Daniel Moraru. Towards phonetically-driven hidden Markov models : Can we incorporate phonetic landmarks in HMM-based ASR? In *Proc. ISCA Tutorial and Research Workshop on Non Linear Speech Processing*, pp. 161–168, Lecture Notes in Artificial Intelligence, vol. 4885, 2007.
  - Stéphane Huet, Guillaume Gravier et Pascale Sébillot. Morphosyntactic Resources for Automatic Speech Recognition. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, mai 2008.
  - Gwénoél Lecorvé, Guillaume Gravier et Pascale Sébillot. On the use of Web resources and natural language processing techniques to improve automatic speech recognition systems. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, mai 2008.
  - Gwénoél Lecorvé and Guillaume Gravier and Pascale Sébillot. An unsupervised Web-based topic language model adaptation method. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, 2008.

### **Journaux nationaux**

- Patrick Gros. Description et indexation automatiques des documents multimédias : du fantasme à la réalité. In *Documentaliste - Sciences de l'information*, 42(6) :383-391, décembre 2005.
- Bruno Bachimont et Patrick Gros. Recherche : des défis scientifiques. In *Les nouveaux dossiers de l'audiovisuel, numéro spécial : Internet : quelle place pour la vidéo ?*, (9) :28-30, mars 2006.

### **Conférences nationales ou francophones**

- Jean-Philippe Poli et Jean Carrive. Proposition d'une architecture pour un système de structuration automatique de flux audiovisuels. In *Actes de la conférence sur la compression et la représentation des signaux audiovisuels*, pp. 49-53, Rennes, France, novembre 2005.
- Xavier Naturel, Guillaume Gravier, Patrick Gros. Étiquetage Automatique de Programmes de Télévision. In *Actes de la conférence sur la compression et représentation des signaux audiovisuels*, Rennes, France, novembre 2005.
- Stéphane Huet, Guillaume Gravier, Pascale Sébillot. Peut-on utiliser les étiqueteurs morphosyntaxiques pour améliorer la transcription automatique?. In *Actes des 26èmes journées d'études sur la parole*, Dinard, France, juin 2006.
- Daniel Moraru, Guillaume Gravier. Ancres macrophonétiques pour la transcription automatique. In *Actes des 26èmes journées d'études sur la parole*, Dinard, France, juin 2006.

- A. Dupuis et C. Dufouil. Une solution client/serveur pour l'analyse de corpus audiovisuels. In *Actes de la conférence sur les systèmes d'information audiovisuels*, Fribourg Suisse, septembre 2006.
- Jean-Philippe Poli and Jean Carrive. Prédiction de séries temporelles : application à la structuration de flux audiovisuels. In *Actes de la conférence sur la compression et la représentation des signaux audiovisuels*, Caen, France, novembre 2006.
- Xavier Naturel. Applications de la détection de texte pour l'indexation de vidéos de télévision. In *Actes de la conférence sur la compression et la représentation des signaux audiovisuels*, Montpellier, France, novembre 2007.
- Stéphane Huet and Guillaume Gravier and Pascale Sébillot. Un modèle multi-sources pour la segmentation en sujets de journaux radiophoniques. In *Proc. Traitement Automatique des Langues Naturelles*, 2008.
- Gwénoùé Lecorvé and Guillaume Gravier and Pascale Sébillot. Vers une adaptation thématique non supervisée de modèles de langage : utilisation d'Internet comme un corpus ouvert. In *Proc. Journées d'Étude sur la Parole*, 2008.

### **Rapports de recherche**

- Stéphane Huet, Pascale Sébillot, Guillaume Gravier. Introduction de connaissances linguistiques en reconnaissance de la parole : un état de l'art. Rapport de Recherche IRISA, No1804, mai 2006.
- Xavier Naturel, Patrick Gros. Detecting Repeats for Video Structuring. Rapport de Recherche IRISA, No 1790, mars 2006.