

# Equipe GEOD

(Groupe d'Etude sur l'Oral et le Dialogue)

Responsable : SERIGNAT, MCF-HC, INPG

## 1. PERSONNEL

NOM Prénom	Section	Etab.	Grade	%Recherche	Quotité
<b>Chercheurs</b>					
BIGI Brigitte	07	CNRS	CR2	100%	1
CAELEN Jean	07	CNRS	DR2	50%	0,5
CASTELLI Eric	07	INPG	CR2 (dét. CNRS)	50%	0
<b>Enseignants-chercheurs</b>					
BESACIER Laurent	27	UJF	MCF	67%	1
FOUQUET Yannick	27	UPMF	ATER	67%	1
LECOMTE Alain	72	UPMF	PR	67%	0,5
SERIGNAT Jean-Franç.	61	INPG	MCFHC	67%	0,8
<b>ITA/IATOS<sup>1</sup></b>					
HOLLARD Solange	07	CNRS	IE	50%	1
VACHER Michel	07	CNRS	IR	50%	1
<b>Personnels sous contrat</b>					
EVENO Nicolas	07	CNRS	Post-doc	50%	1

LISTE DES DOCTORANTS AU 01/09/2005			
NOM Prénom	Etablissement	Financement	Partenaire
CHAILLOL Stéphane	CNAM	Assedic	
DOMINGUEZ Patricia	UJF	Bourse CONACYT	Mexique
LE Viet-Bac	UJF	Bourse Excellence Eiffel	Ministère Affaires Etrangères
QUANG Vu-Minh	UJF	Bourse MAE, Ambassade de France au Vietnam	MICA-INP Hanoi Co-tutelle
TRAN Do-Dat	INPG	Bourse AUF	MICA-INP Hanoi Co-tutelle
TRAN Vu-Truc	UPMF	Bourse MAE, Ambassade de France au Vietnam	INP Hanoi Co-tutelle
TAN Tien-Ping	UJF	Bourse Gouvernement Etranger	Malaisie
POPESCU Vladimir	UJF	Bourse MIRA	Bucarest Fac.Electroni. Co-tutelle
OUAYOUCH Anas	UJF	Salarié	

LISTE DES ENSEIGNANTS-CERCHEURS ASSOCIES <sup>2</sup> ET COLLABORATEURS BENEVOLES <sup>3</sup> AU 01/09/2005			
NOM Prénom	Etablissement	Grade	Statut
TUFFELLI Denis	UJF	PR	Collaborateur bénév.
ISTRATE	ESIGETEL	MCF	Associé

Nombre d'équivalents chercheurs (NE) : 9,87 (avec doctorants)

<sup>1</sup> Personnels explicitement rattachés à l'équipe

<sup>2</sup> Un enseignant-chercheur associé est un collègue d'un autre site que Grenoble et d'un établissement avec lequel le laboratoire n'a pas de relation formalisée

<sup>3</sup> toute personne qui, sans être usager ou personnel d'un des établissements partenaires du laboratoire, concourt bénévolement à la réalisation de ses missions et n'a pas d'employeur

## 2. BILAN DES ACTIVITES DE RECHERCHE 02-05

### 2.1 *Thématique scientifique et objectifs généraux*

Depuis plus d'une quinzaine d'années, les moyens de communication (téléphone mobile, Internet) et les média de diffusion électronique de l'information (émissions de radio et de télévision numériques) ont connu un essor sans cesse croissant. Dans le même temps, le progrès des techniques de traitement numérique de l'information et de la technologie des calculateurs a été gigantesque. Cette évolution a ouvert des perspectives prometteuses à de nombreuses applications dans le domaine de la communication orale homme-machine ou homme-homme médiatisée, mais aussi à des applications spécifiques dans le domaine médical comme la télésurveillance des malades à domicile (habitat intelligent). En parallèle, grâce à la facilité de stockage due en partie à des algorithmes de compression très efficaces, les corpus de documents audio et vidéo ne cessent de croître. Pratiquement toute information multimédia se trouve aujourd'hui disponible sous format numérique et son exploitation ouvre le champ à de nouvelles applications d'indexation et de recherche de documents par le contenu.

Dans ce contexte, la thématique de recherche de GEOD est centrée sur l'**Interaction Orale**, articulée autour de *deux axes scientifiques principaux* : *Reconnaissance* (parole, audio et locuteur) et *Dialogue* (modélisation et compréhension). Pour ces deux axes, subsistent encore un certain nombre de verrous liés à la *généricité* des modèles : cette caractéristique reste un objectif essentiel, situé au centre de nos préoccupations de recherche pour le long terme. Pour l'axe *Reconnaissance*, les efforts de recherche de GEOD pendant la période considérée ont porté sur deux sous thèmes : (§2.1.1) la réalisation de systèmes de reconnaissance multilingue pour la parole continue et l'amélioration de leur robustesse, (§2.1.2) l'exploitation de la parole et des sons comme composante de l'interaction multimodale dans les espaces perceptifs. Pour l'axe *Dialogue*, l'objectif a été (§2.1.3) le développement de systèmes de dialogue homme-machine multimodaux.

#### 2.1.1. Reconnaissance robuste multilingue de la parole continue

Participants à ce thème de recherche : **L. Besacier** (Resp.), B. Bigi, E. Castelli, N. Eveno, R. Lamy, V.B. Le, P. Mayorga-Ortiz, D. Moraru, Q.C. Nguyen, J.F. Serignat, T.P. Tan, D.D. Tran, D. Tuffelli, D. Vaufraydaz, Q. Vu-Minh.

Mots-clés : Reconnaissance de parole continue, robustesse, grand vocabulaire, modèle acoustique, modèle de langage, multilingue, transcription, transcription enrichie.

##### a) Résumé

Les activités de recherche de GEOD dans ce thème se sont concentrées sur le développement de modèles acoustiques multilocuteurs et de modèles de langage pour le système de reconnaissance automatique de parole continue en français du laboratoire. L'originalité réside dans l'approche qui consiste à "aspirer" un grand nombre de sites Web dans une langue donnée et à filtrer les données textuelles récupérées afin de les rendre exploitables pour calculer des modèles statistiques de langage. Une adaptation de cette méthodologie à des langues peu dotées marque une tendance vers le multilinguisme qui prend une importance de plus en plus grande dans ces recherches. Des applications à la langue vietnamienne, à la langue khmère et à l'espagnol-mexicain (castillan) ont été envisagées et ont permis d'obtenir des résultats très encourageants. Des extensions de ce thème de recherche, dans le sens d'une "transcription enrichie" (segmentation en locuteurs, détection de zones d'intérêt, détection de "jingles" audio,...) pour des applications de recherche d'information par le contenu dans les bases de données, ont aussi été menées dans le cadre de diverses participations à des campagnes d'évaluation internationales. Enfin, des travaux pour des applications en biométrie ont été conduits en tenant compte du caractère souvent multimodal du domaine.

##### b) Description détaillée

Le développement de systèmes de reconnaissance multilingue de parole continue, grand vocabulaire et multi-locuteur, vise des applications de reconnaissance en langage naturel, dans un contexte de dialogue finalisé homme-homme ou homme-machine (systèmes de dialogue ou de traduction automatique pour le renseignement touristique ou pour le commerce électronique). Les problèmes scientifiques couverts par cet axe de recherche sont le *développement des modèles acoustiques multilocuteurs en environnement réaliste de travail* (tenant compte aussi du canal de transmission, par exemple pour la voix téléphonique ou la voix sur IP), la *modélisation du langage*, éventuellement adaptée à l'application et la *portabilité rapide vers une nouvelle langue*. Ils confèrent à cet axe de recherche un caractère véritablement pluridisciplinaire, sur lequel GEOD a concentré une part importante de ses efforts en développant un système de reconnaissance de parole continue efficace pour le français (RAPHAEL : Reconnaissance Automatique de Phrases, d'Acronymes et d'Expressions Langagières,

Thèse D. Vaufreydaz - 2002). Ce système, dont le schéma d'ensemble est présenté figure 1, a été exploité avec succès dans divers projets internationaux de dialogue homme-homme multilingue médiatisé (CSTAR III, 2002-2006 et NESPOLE, 2000-2002). Il a également été à la base du développement d'un ensemble de travaux de recherche menés à GEOD dans le sens du multilinguisme pour la portabilité rapide vers d'autres langues (vietnamien, khmer, castillan) et pour l'indexation de documents audio pour la recherche par le contenu.

En ce qui concerne le vietnamien, GEOD a d'abord développé un système de reconnaissance en mots isolés (Thèse Q.C. Nguyen, 2002) dans le cadre de la coopération scientifique du CLIPS avec l'Institut Polytechnique de Hanoi (IPH). En effet, le vietnamien est une langue tonale comportant 6 tons lexicaux, dans laquelle chaque syllabe est prononcée avec un ton lexical distinct. Dans les langues tonales, pour reconnaître les mots, il faut réaliser la reconnaissance des syllabes prononcées avec le ton, c'est-à-dire réaliser deux processus parallèles de reconnaissance du ton et de reconnaissance de la syllabe sans le ton. Cette étude a aussi été à l'origine du démarrage des activités de recherche du Centre de Recherche International MICA, créé à Hanoi, fin 2001, sous l'impulsion d'Eric Castelli (en détachement CNRS depuis septembre 2001) et inauguré officiellement en novembre 2002.

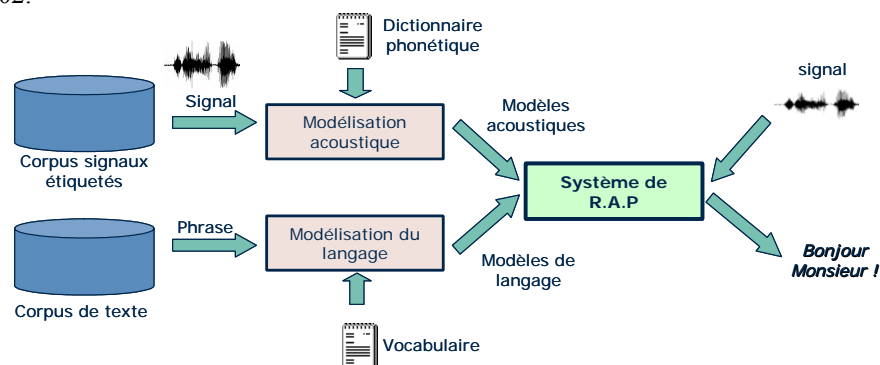
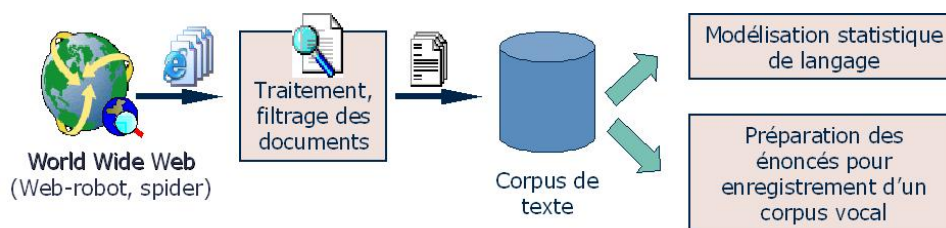


Fig. 1 : Schéma d'ensemble de la reconnaissance automatique de la parole.

Pour développer un système de reconnaissance automatique de la parole continue dans une nouvelle langue, il est souvent nécessaire de rassembler une grande quantité de corpus, contenant à la fois des signaux de parole (pour l'apprentissage des modèles acoustiques du système) mais également des données textuelles (pour l'apprentissage des modèles de langage du système). Pour une langue peu ou mal dotée il faut alors collecter soi-même les ressources nécessaires : signal de parole, lexique, corpus textuels.

L'originalité de la méthodologie mise au point à GEOD (V.B. Le, thèse en cours et L. Besacier, HDR en cours) consiste à adapter, aux langues peu dotées (vietnamien et khmer dans un premier temps), l'approche qui a déjà été validée pour une langue bien dotée, telle que le français (Thèse de D. Vaufreydaz, 2002). Concernant le recueil de données textuelles en grande quantité (figure 2), une approche intéressante consiste à "aspérer" un grand nombre de sites Web dans la langue donnée et à filtrer les données récupérées pour les rendre exploitables. Ces données textuelles peuvent servir d'une part à calculer des modèles de langages statistiques, et d'autre part à obtenir un corpus pouvant ensuite être prononcé par des locuteurs en vue de la constitution d'une base de signaux conséquente.



(Cf. Thèse Vaufreydaz 2002)

Fig. 2 : Récupération de données textuelles en utilisant le Web.

Les problèmes spécifiques pour les langues peu dotées concernent le nombre de sites Web qui est souvent assez réduit, la vitesse de transmission et la qualité des documents qui nécessitera alors plus d'outils de traitement. On pourra préférer des sites de nouvelles, au fort contenu rédactionnel, tels que VNexpress<sup>4</sup> par exemple pour le vietnamien. Afin de rendre les données exploitables, un certain nombre de traitements sont

<sup>4</sup> <http://www.vnexpress.net>

nécessaires tels que : 1) transformation html vers texte, 2) normalisation des tags, 3) conversion des encodages (nous avons choisi de tout convertir vers une représentation interne unique utilisant l'encodage UTF-8 d'Unicode), 4) séparation en phrases et 5) en mots, 6) groupement de mots composés, 7) transcription des symboles et 8) filtrage en fonction d'un vocabulaire donné. Alors que certains traitements peuvent être considérés comme relativement indépendants de la langue cible (1-2-6-8), d'autres doivent être repensés (3-4-5-7) pour chaque nouvelle langue cible : par exemple la séparation en mots est triviale pour les écritures latines mais problématique pour d'autres systèmes d'écriture comme le khmer, surtout si l'on ne dispose pas d'un vocabulaire (i.e. une liste de mots) au départ. Une boîte à outils open source rassemblant quelques uns de ces outils de traitement a été développée à GEOD<sup>5</sup>.

Cette méthodologie appliquée au vietnamien (décrite dans ACT-31, V.B. Le et al., 2003) a permis de collecter environ 2,5 Go de pages Web, ce qui, après filtrage, représente une quantité de données textuelles pouvant servir à l'apprentissage d'un modèle de langage statistique d'environ 400 Mo (5 millions de phrases). A titre de comparaison, une année complète du journal Le Monde en français correspond à 120 Mo en moyenne. Pour le khmer, qui est une langue encore moins dotée que le vietnamien, la collecte de ressources linguistique est plus difficile. Les travaux sont en cours dans le cadre de la collaboration avec l'ITC (Institut de Technologie du Cambodge) et la quantité de pages Web collectées est actuellement de 80 Mo (environ 6000 pages), ce qui reste encore faible et nécessite de poursuivre la collecte.

Enfin, pour le castillan (espagnol mexicain), des travaux similaires (décrits dans ACT-45 et ACT-46, L. Villaseñor et al., 2003) ont été menés en coopération avec le laboratoire INAOE<sup>6</sup> à Puebla (Mexique) dans le cadre du projet IPHM (Interaction Parlée Homme-Machine, 2002-2004), soutenu par le CNRS dans son action pour la création du laboratoire Franco-Mexicain d'Informatique (LAFMI). La méthodologie développée à GEOD et les outils de récupération de corpus de texte ont été implantés à l'INAOE et mis en œuvre sur le Web mexicain. L'application au castillan a permis, dans un premier temps, d'élaborer un lexique d'environ 200 000 mots, puis, à partir d'un corpus (23 Go) recueilli sur le Web, de constituer un corpus de texte (839 phrases, moyenne 45 mots) destiné à l'élaboration d'un premier modèle de langage pour un système de reconnaissance automatique de parole continue spontanée en castillan.

Pour le recueil de signaux de parole, GEOD a développé (depuis le début de ses travaux sur les ressources linguistiques pour la reconnaissance de parole) un outil logiciel ne mettant en œuvre que du matériel standard sur PC : EMACOP (Environnement Multimédia pour l'Acquisition et la gestion de Corpus Parole) (DEA de D. Vaufreydaz, 98). La plupart du temps, les campagnes d'enregistrement mobilisent d'importantes ressources humaines pour guider ou assister les locuteurs dans leur tâche de diction, pour organiser l'enregistrement, pour préparer les scénarios et les données, etc. Il faut pouvoir contrôler les différents scénarios pour varier les conditions de capture : la lecture d'un texte ou d'une suite de mots ou de mots isolés, la répétition après écoute d'une phrase, le dialogue en réponse à des questions, etc. Les méthodes d'acquisition rigoureusement contrôlées sont donc lourdes et les difficultés sont amplifiées dans le cas des langues mal dotées où les locuteurs ne sont pas forcément rôdés à l'utilisation de moyens informatiques. C'est pourquoi, le développement d'un utilitaire portable de gestion et d'acquisition de grands corpus sur un matériel standard est absolument indispensable. Les interfaces ont été adaptées pour manipuler respectivement les caractères vietnamiens et khmers. La figure 3 montre un exemple de l'interface d'EMACOP adaptée pour la langue khmère.

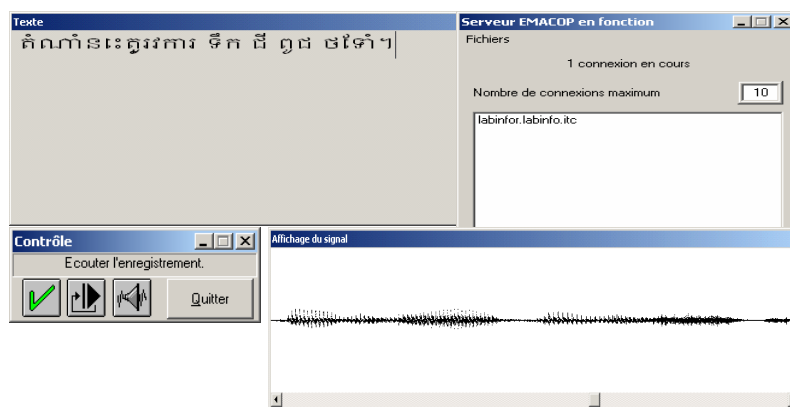


Fig. 3 : Interface EMACOP adaptée pour le Khmer

<sup>5</sup> <http://www-clips.imag.fr/geod/User/viet-bac.le/outils/>

<sup>6</sup> Institut National d'Astrophysique, Optique et Electronique (INAOE)

Un autre problème à résoudre pour développer un système de reconnaissance automatique de parole continue dans une langue peu dotée est la constitution d'un dictionnaire de prononciation (ou dictionnaire phonétique) qui est une ressource essentielle pour le système et pour enrichir un dictionnaire bilingue, permettant au locuteur étranger de connaître la prononciation du mot en langue cible. Cette tâche est difficile pour des langues peu dotées dont le système phonologique est parfois méconnu ou sujet à débats (langues peu ou mal décrites). Si nous excluons les méthodes manuelles de phonétisation qui, bien que donnant des dictionnaires de prononciation de meilleure qualité, sont longues et coûteuses, l'approche automatique la plus rapide que nous avons choisie pour notre méthodologie, utilise un système de reconnaissance phonémique appliqué sur des enregistrements de mots à phonétiser. On obtient ainsi un premier étiquetage automatique en phonèmes d'une liste de mots qui peut être alors révisée par un opérateur humain. Il faut pour cela disposer d'un système de reconnaissance automatique de phonèmes, qui provient généralement d'une langue source bien dotée (par exemple, un système de reconnaissance des phonèmes du français). L'efficacité de la méthode dépend de la couverture phonémique de la langue cible par le décodeur phonémique de la langue source, d'où la nécessité d'employer si possible des décodeurs phonémiques multilingues, couvrant un maximum de phonèmes dans l'alphabet phonétique international. Pour le vietnamien par exemple, lorsque la langue source utilisée est le français, le taux de couverture au départ est d'environ 63%, ce qui n'est évidemment pas optimal, mais conserve un intérêt comme "amorçage" de la méthode. Néanmoins, lorsqu'on part de modèles acoustiques multilingues (ce qui fut fait dans le cadre d'un travail en collaboration avec CMU, lors d'un séjour de V.B. Le à Pittsburgh, en avril 2005), obtenus à partir de 7 langues (chinois, croate, français, allemand, japonais, espagnol et turc) le taux de couverture du vietnamien atteint 87%. Ainsi, nous avons pu montrer pour le vietnamien qu'avec une quantité réduite de signaux en langue cible (2h), il est possible d'obtenir des performances acceptables (63.8% de mots correctement reconnus à partir d'un modèle français et 66.3% à partir d'un modèle multilingue), alors qu'il faudrait certainement 50 à 100 fois plus de signaux en langue cible pour obtenir des performances comparables, sans adaptation à partir de la langue source.

Cependant, afin d'améliorer les performances du système de reconnaissance automatique du vietnamien, un corpus de parole vietnamien est toujours en cours d'enregistrement à MICA. A ce jour, il contient 35 locuteurs, 16 femmes et 19 hommes, venant des régions nord, centre et sud du vietnam. Chaque locuteur a enregistré environ 1 heure de parole ce qui fait un total de 35 heures. Le corpus contient des séquences de lettres, de nombres et de mots isolés, mais aussi la lecture de phrases complètes et de paragraphes. Des détails supplémentaires sur les ressources collectées pour le vietnamien sont fournis dans ACT-53 (V.B. Le et al., 2004) et une description détaillée des traitements et des expérimentations associées pour l'adaptation rapide de modèles acoustiques au vietnamien est donnée dans ACT-68 (V.B. Le et al., 2005). De même, pour la reconnaissance de l'espagnol-mexicain (castillan), une base de données orales (5000 phrases, 50 locuteurs) est en cours d'enregistrement à l'INAOE (Puebla, Mexique) pour l'apprentissage des modèles acoustiques des phonèmes (22 phonèmes) afin d'améliorer les performances d'un système de reconnaissance automatique du castillan (ACT-64, L. Villaseñor-Pineda et al., 2004 et ACT-74, L. Villaseñor-Pineda et al., 2005).

D'autre part, afin de confronter un certain nombre de systèmes développés à GEOD au cadre expérimental qu'apportent les campagnes d'évaluation organisées par divers instituts nationaux ou internationaux, GEOD a étendu son axe de recherche en reconnaissance automatique de la parole dans le sens d'une "*transcription enrichie*" qui consiste à produire un document structuré très riche en informations à partir d'un signal audio brut. Cela inclut la recherche d'éléments non linguistiques dans un flux audio (segmentation en locuteurs, détection de zones d'intérêt, détection de "jingles" audio, etc...) qui fait appel à des techniques ou à des algorithmes exploitant des paramètres similaires à ceux utilisés en reconnaissance de parole. Les organisateurs de ces campagnes proposent un corpus commun et aussi des métriques d'évaluation communes à tous les participants. Les données proposées en quantité suffisamment importante couvrent aussi tous les types d'enregistrements : enregistrements téléphoniques, enregistrements de journaux télévisés et enregistrements de réunions. Une participation soutenue à de telles campagnes permet de publier des résultats expérimentaux sur des corpus connus de la communauté, apportant ainsi une visibilité nationale ou internationale aux travaux réalisés. Un autre intérêt est l'incitation à collaborer avec d'autres équipes de recherche en raison de la gestion lourde qu'implique la participation à de telles campagnes (quantité de données à traiter, réactivité nécessaire pour renvoyer les résultats dans les délais). Il faut, cependant, noter aussi un aspect négatif de ces campagnes qui ne laissent pas toujours le temps de proposer des systèmes très exploratoires d'un point de vue des méthodes. Enfin, les applications possibles de ces travaux sont l'indexation et la recherche d'informations par le contenu dans des documents audio ou audio-vidéo, avec souvent un contexte très fort de multimodalité. Plusieurs activités ont été conduites ou sont encore en cours à GEOD dans cet axe :

1). *Transcription enrichie*, dans le cadre de la participation à la campagne française ESTER<sup>7</sup> (Evaluation des Systèmes de Transcription Enrichie des émissions Radiophoniques) qui visait à l'évaluation des systèmes de reconnaissance automatique de la parole appliqués à la transcription d'émissions radiophoniques. L'objectif était aussi d'enrichir les transcriptions par un ensemble d'informations annexes comme le découpage en tours de parole ou la segmentation et le regroupement en locuteurs et le marquage d'entités nommées. Cette campagne a été organisée par l'Association Francophone de la Communication Parlée (AFCP) avec le concours de la Délégation Générale pour l'Armement (DGA) et de l'Agence Européenne de Ressources Linguistiques (ELRA/ELDA). Les travaux et les résultats obtenus sont présentés en détail dans ACT-92 (R. Lamy et al., 2004) et dans COM-9 (L. Besacier et al., 2005).

2). *Segmentation en locuteurs* (thèse de D. Moraru, 2004), travaux qui ont notamment permis à GEOD de participer avec succès à plusieurs campagnes d'évaluation (SpRec 2002, RT 2003 et RT 2004)<sup>8</sup> organisées par l'institut américain NIST (National Institute of Standards and Technology) sur diverses catégories de documents sonores : données téléphoniques, journaux télévisés, données de réunions et aussi à la campagne ESTER, citée au point précédent. Ces travaux ont été menés en coopération avec le Laboratoire d'Informatique d'Avignon (LIA).

3). *Biométrie* (recherche d'informations concernant le locuteur) dans le cadre de la participation à l'action européenne COST275<sup>9</sup> (*Biometric Person Authentication over the Internet*) de 2001 à 2003. Cette action a également servi de cadre expérimental pour diagnostiquer les performances des systèmes de vérification automatique du locuteur (et de reconnaissance automatique de la parole) dans des conditions de pertes de paquets sur réseaux (Voix sur IP) et de données compressées (thèse de P. Mayorga-Ortiz, 2005). Dans cette même catégorie de travaux, se situe également l'étude de la cohérence entre le signal de parole et les mouvements labiaux du locuteur (*biométrie labiale bimodale*), effectuée par N. Eveno (2004-2005), post-doctorant accueilli à GEOD dans le cadre d'une ACI Sécurité Informatique du CNRS, projet BIOMUL (2003-2006) mené en coopération avec le laboratoire LIA.

4). *Détection de "zones d'intérêt" intonatives* dans le signal de parole (thèse de Q. Vu-Minh, en cours) qui peuvent fournir des informations de nature sémantique (phrase interrogative, par exemple, que l'on pourrait exploiter dans un système de dialogue) ou renseigner sur l'état émotionnel du locuteur.

5). *Détection de "jingles" audio* ou de séquences sonores particulières dans un flux audio pouvant fournir des indices précieux de macrosegmentation d'un document audio ou audio-visuel. Ces travaux ont été menés dans un contexte multimodal (fusion de modalités audio et vidéo) en coopération avec l'équipe MRIM du CLIPS et avec le laboratoire LIS<sup>10</sup> de l'INPG (Institut National Polytechnique de Grenoble) pour la participation à des campagnes d'évaluation TRECVID<sup>11</sup> (Text REtrieval Conference VIDEO), organisées par le NIST en 2002, 2003 et 2004. Par exemple, les tâches abordées dans ces campagnes, nécessitant la fusion d'informations audio et vidéo, étaient la détection de changement de plan vidéo et la "segmentation en histoires" de documents audiovisuels. Le *tableau 1* présente une synthèse des résultats obtenus aux diverses campagnes d'évaluations auxquelles GEOD a participé.

6). *Reconnaissance de diverses classes d'événements sonores* sur des enregistrements issus d'espaces perceptifs. Les enregistrements obtenus dans ces conditions possèdent des spécificités particulières et les travaux de recherche correspondants seront présentés en détail dans le sous thème suivant (§ 2.1.2).

7). Pour terminer la description de ce sous thème de recherche, on citera enfin les travaux effectués sur l'étude de la "*parole audio-visuelle*" dans le cadre du projet TELMA (2004-2006), financé en 2005 par un BQR INPG et mené en coopération avec les laboratoires ICP<sup>12</sup> et LIS de l'INPG. Le projet a pour objectif technique précis d'exploiter la modalité visuelle de la parole (lecture labiale), associée aux gestes de la Langue Française Parlée Complétée (LPC) pour un développement algorithmique de fonctionnalités audio-visuelles à l'usage des personnes malentendantes. L'étude de faisabilité de leur intégration dans un terminal autonome de télécommunication téléphonique fait aussi partie du projet.

Tâches \ Année	2002	2003	2004	2005
<b>Segmentation en locuteurs</b>	NIST meeting 1/4 NIST BN 2/4 NIST Tel 3/4	Rich Transcription (RT) BN 2/8*	Rich Transcription (RT) meeting 1/3*	ESTER BN 4/5 Rich Transcription (RT) meeting 2/3*

<sup>7</sup> <http://www.afcp-parole.org/ester/index.html>

<sup>8</sup> <http://www.nist.gov/speech/>

<sup>9</sup> [http://www.fub.it/cost275/pages/\\_home\\_main/index.htm](http://www.fub.it/cost275/pages/_home_main/index.htm)

<sup>10</sup> <http://www.lis.inpg.fr/>

<sup>11</sup> <http://www-nlpir.nist.gov/projects/trecvid/>

<sup>12</sup> <http://www.icp.inpg.fr/>

Transcription				ESTER BN 6/8
<b>Recherche d'informations</b>	Extraction de plans Parole 7/13 Monologue 3/9	Extraction de plans Personne X 4/4	Segmentation en histoires 3/6	

Tableau 1. Bilan des différentes participations de GEOD à des campagnes d'évaluation (\*=collaboration avec le LIA ; BN=Broadcast News)

## 2.1.2. Environnements perceptifs : la parole et les sons comme composante de l'interaction multimodale

Participants à ce thème de recherche : **J.F. Serignat** (Resp.), L. Besacier, E. Castelli, S. Chaillol, D. Istrate, D. Tuffelli, M. Vacher.

Mots-clés : Traitement du signal, détection, segmentation, classification des sons, espace perceptif, interaction...

### a) Résumé

Dans ce thème sont décrits les travaux de recherche de GEOD sur la parole et sur les sons, dans le contexte applicatif des espaces perceptifs et plus particulièrement dans le cadre d'une coopération avec le laboratoire TIMC<sup>13</sup> pour l'*Habitat Intelligent pour la Santé* (HIS). Dans les locaux de TIMC, un appartement (30m<sup>2</sup>) a été équipé pour devenir un HIS prototype. Divers algorithmes de détection et de classification des sons de la vie courante ont été développés et validés pour la détection de situations de détresse d'un patient sous télésurveillance médicale. De même, un modèle de langage pour le système de reconnaissance de parole de GEOD a été adapté pour la reconnaissance des appels de détresse dans cet environnement. Quelques développements pour des applications en "smart room" (salle de réunion) sont également présentés.

### b) Description détaillée

Depuis plusieurs années s'est développé le concept général d'*espace perceptif* ou de *salle intelligente* dans lequel il s'agit de concevoir des salles (bureau d'étude, salle de réunion, cuisine, chambre hospitalière...) dotées de capteurs divers (microphones, caméras, détecteurs de présence infra-rouge, micro-capteurs mobiles de signaux biologiques, etc.) gérés par un système informatique. Ce système analyse les signaux en temps réel et répond de diverses façons (alarmes sonores, réponses vocales...) aux besoins, demandes, attentes des acteurs humains. Depuis l'année 2000, GEOD a considéré que ce domaine, abordé sous l'angle de l'interaction multimodale entre l'environnement et l'acteur humain, pouvait présenter un champ d'expérimentation intéressant pour des recherches menées sur la reconnaissance de la parole mais aussi sur la détection et la reconnaissance d'événements sonores de natures diverses. Ce contexte pose des problèmes spécifiques de traitement car les dispositifs d'acquisition sonore ne sont généralement pas situés près des personnes.

Les premières recherches en ce sens ont été initiées par une coopération avec l'équipe AFIRM<sup>14</sup> (N. Noury, V. Rialle) du laboratoire TIMC de l'UJF (Université Joseph Fourier) dans le cadre d'un projet de télésurveillance médicale pour personnes âgées ou en convalescence (RESIDE-HIS, achevé en 2001), subventionné par l'IMAG (Institut de Mathématiques Appliquées de Grenoble) et dans lequel GEOD était impliqué pour le traitement de la modalité sonore. Actuellement, les travaux dans ce domaine se poursuivent à GEOD avec deux orientations principales : une en direction de l'*Habitat Intelligent pour la Santé* (HIS), l'autre plus particulièrement tournée vers des applications type "*Smart Room*" destinée aux réunions de travail.

#### *Habitat Intelligent pour la Santé :*

L'*Habitat Intelligent* (parfois ramené au concept de *chambre intelligente*) en télémédecine a pour but de détecter des chutes, malaises, appels vocaux, etc., et de transmettre ces informations au personnel en charge des soins et de la sécurité. Cette approche consiste à installer au domicile de la personne des capteurs de divers types (sonores, physiologiques, actimétriques, de chute, environnementaux...) et à envoyer leurs signaux à un centre de télésurveillance, dont la fonction est de veiller à la sécurité de la personne et d'apporter une réponse rapide et appropriée en cas de situation critique. A plus long terme et au-delà de cette finalité de téléalarme, on peut envisager un véritable suivi en profondeur de l'évolution physique de la personne, comme la détection d'une altération de l'état général du patient, par exemple, grâce au calcul automatique d'indices de mobilité dans les activités quotidiennes.

<sup>13</sup> <http://www-timc.imag.fr/>

<sup>14</sup> <http://www-timc.imag.fr/AFIRM/>

Les travaux de GEOD dans ce domaine sont menés dans la continuité de la coopération scientifique avec l'équipe AFIRM du laboratoire TIMC. Ils ont bénéficié d'un soutien grâce à une ACI Santé du CNRS (projet DESDHIS, 2002-2004) et ont été réalisés à travers la thèse de Dan Istrate (2003). Dans le bâtiment Jean-Roger de la Faculté de Médecine de Grenoble, à l'intérieur du laboratoire TIMC, un local destiné à l'installation d'un Habitat Intelligent pour la Santé (HIS), a été équipé pour constituer un prototype d'appartement (figure 4) de type T1 (environ 30 m<sup>2</sup>), comprenant les zones d'habitat classiques que sont la chambre, le séjour, la cuisine, les toilettes, la douche et un couloir. Cet appartement a été équipé de capteurs sonores, physiologiques et de détecteurs de mouvement. La surveillance vidéo étant exclue, car elle peut être mal perçue par les patients, doit être remplacée par une surveillance sonore dont le contenu informatif est moindre, ce qui, en conséquence, augmente la complexité des algorithmes de prise de décision (fusion de données). Une zone technique attenante à l'appartement a été ajoutée afin de recevoir le système informatique d'expérimentation du projet. Une vitre sans teint séparant les deux parties autorise une surveillance de l'habitat depuis la zone technique.

L'infrastructure matérielle du système d'acquisition de données dans cet habitat avait été conçue et validée dans le cadre d'un projet précédent (RESIDE-HIS, déjà cité) et l'objectif général du projet DESDHIS (*DEtection de Situations de Détresse dans un Habitat Intelligent*)

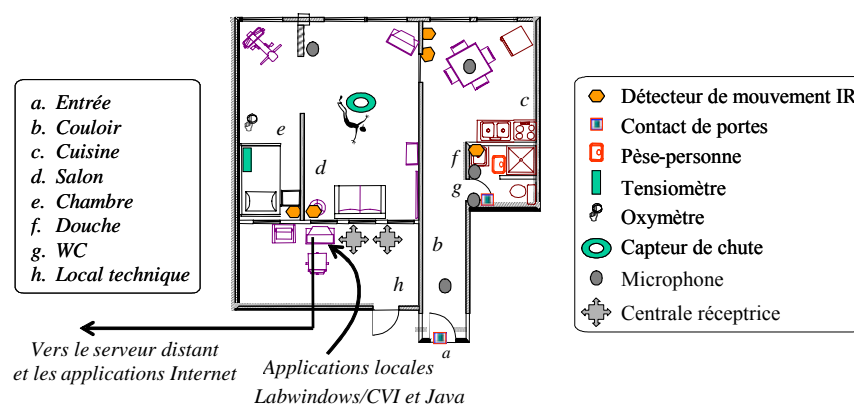


Fig. 4 : Habitat Intelligent Santé expérimental.  
(Salle 333 – TIMC-IMAG, Faculté de Médecine de Grenoble)

pour la Santé) était de développer et de valider l'algorithmique de traitement des signaux issus des divers capteurs, GEOD s'impliquant essentiellement dans celui des signaux issus des capteurs sonores. Le but du traitement, en plus de déterminer la pièce où est localisée la personne, est de reconnaître parmi les sons de la vie courante, d'éventuels appels de détresse clairement formulés (*au secours, appelez un médecin,...*) ou d'identifier un bruit spécifique pouvant être un indice d'une situation critique (chute, bris de verre, cri de douleur, etc...).

Le système d'analyse sonore exploite cinq entrées en parallèle et fonctionne en temps réel, mais pour des contraintes liées à la préservation de la sphère privée, le signal est acquis et traité en temps réel, sans enregistrement. Le schéma global de l'organigramme de traitement sonore est présenté figure 5. Après l'acquisition du signal et la détection puis la localisation d'un événement sonore, une segmentation du signal en parole/non-parole est opérée afin de diriger les signaux de parole sur le système RAPHAEL de reconnaissance automatique de parole et les sons (non-parole) sur un système de classification et de reconnaissance des sons de la vie courante.

Pour la détection des événements sonores, plusieurs algorithmes ont été proposés dans la thèse de Dan Istrate (2003). Un des algorithmes se fonde sur la transformée en ondelettes du signal. Cet algorithme présente de très bonnes performances dans l'environnement bruité réel : un taux d'erreur de détection de 5.6% pour un niveau de bruit élevé (RSB=0 dB). Il semble mieux adapté au contexte applicatif où les sons à détecter sont majoritairement impulsionnels (claquements de portes, bris de verres, chutes).

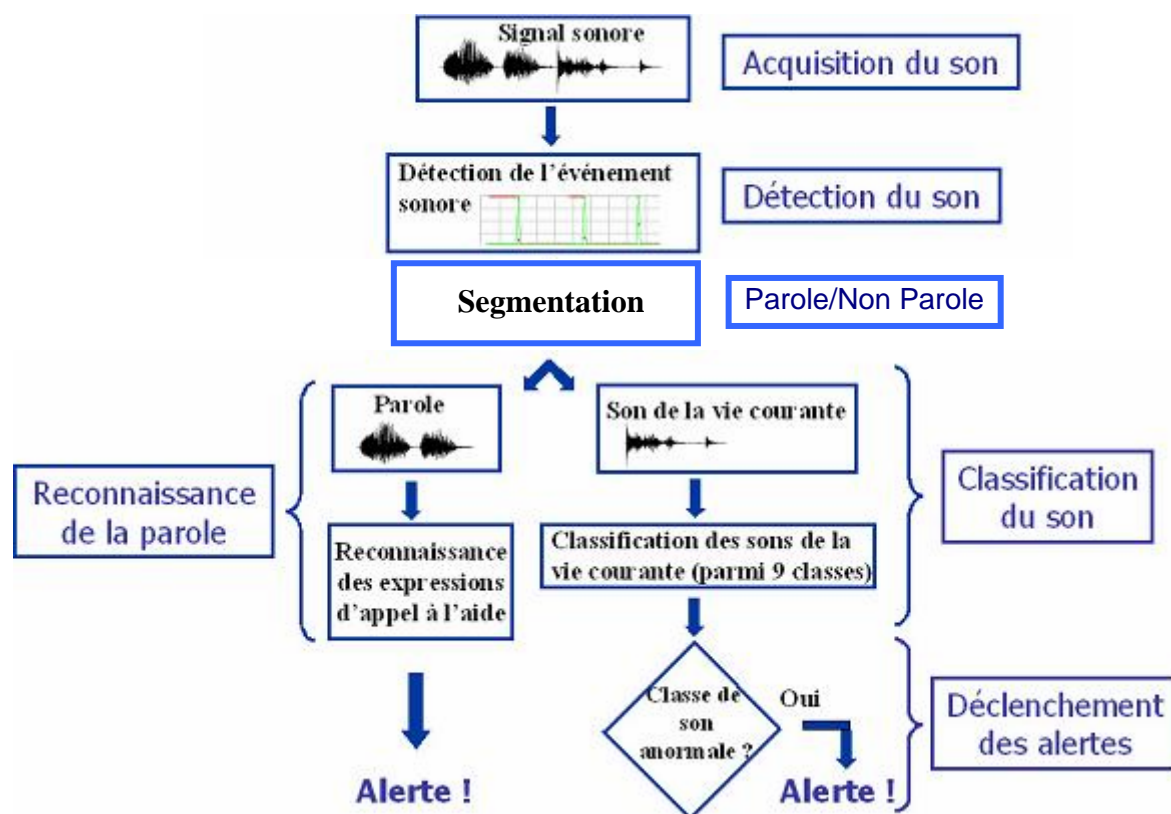


Fig. 5 : Synoptique global de l'algorithmique de traitement des signaux sonores.

Pour la classification des sons de la vie courante, l'algorithme de classification utilisé est fondé sur les mélanges de gaussiennes (GMM); il a été étudié dans diverses conditions de bruit ambiant et avec des valeurs du rapport signal à bruit comprises entre 0 dB et 40 dB. Comme la paramétrisation du signal est très importante pour une bonne classification des sons, l'étude des paramètres acoustiques adaptés aux sons de la vie courante a constitué l'axe principal de cette recherche. La méthodologie a consisté à évaluer par des méthodes statistiques la pertinence des paramètres utilisés habituellement pour la détection de signaux musicaux (nombre de passages par zéro, "roll-off point", barycentre spectral). Étant donné que la transformée en ondelettes est mieux adaptée à l'analyse de signaux impulsifs, la possibilité d'extraire des paramètres acoustiques à partir de cette transformée a été aussi étudiée. Le meilleur système présenté dans la thèse donne un taux d'erreur de classification de 7.1% pour 7 classes de sons. D'autres résultats plus récents sur l'utilisation des coefficients d'ondelettes pour la détection et la classification des sons de la vie courante sont donnés dans les communications ACT-62 et ACT-63 (M. Vacher et al., 2004).

Les problématiques liées au couplage entre la détection et la classification, ainsi que le problème de l'évaluation d'un tel système sont aussi abordées dans ce travail. Dans ce but, une base de données des sons de la vie courante a dû être enregistrée : l'ensemble de la base constituée correspond à 3354 signaux d'une durée d'environ une heure et 35 minutes. La méthodologie d'enregistrement et la description détaillée de cette base de sons de la vie courante sont présentées dans le chapitre 2 de la thèse de Dan Istrate. Une implémentation en temps réel des algorithmes proposés a été réalisée pour l'application de télésurveillance médicale et a été validée dans l'appartement test disponible pour le projet. On trouvera une synthèse détaillée de ces travaux dans le journal *IEEE Transactions on Information Technology in Biomedicine* (ACL-6, D. Istrate et al., 2005).

Dans les cas où le module de segmentation détecte des sons de parole, le système RAPHAEL de reconnaissance automatique de parole continue est utilisé avec un modèle de langage à vocabulaire limité (environ 15000 mots), adapté à un contexte de télésurveillance où des appels de détresse peuvent survenir. Un corpus spécifique de parole a été enregistré pour pouvoir faire des évaluations précises des performances du système. Il comporte 126 phrases courtes en français dont 64 correspondent à une situation normale (partie *anodine* : "Bonjour", "Où est le sel", etc...) et 64 sont des expressions de détresse (partie *appels de détresse* : "Au secours", "Un médecin, vite !", "Appelez les pompiers !", etc...). Ce corpus a été prononcé et enregistré par 21 locuteurs (11 hommes, 10 femmes), dans une tranche d'âge de 20 à 65 ans. La durée totale de parole enregistrée correspond à environ 38 minutes de parole et représente 2646 fichiers de signal. Quelques tests (intuitifs) du système de reconnaissance de parole avec son modèle de langage adapté à ce corpus ont été faits avec un locuteur parlant de manière spontanée dans le local et semblent prometteurs. Cependant une évaluation globale et rigoureuse sur l'ensemble du corpus enregistré reste à faire pour pouvoir donner des résultats chiffrés, également

en faisant varier les conditions de bruit ambiant. Le regroupement de tous les modules du système et leur intégration sur une même station de travail (PC-3GHz) est en cours et devrait nous permettre prochainement d'effectuer ces évaluations.

#### *Smart Room :*

Quelques travaux de recherche sur la parole comme composante de l'interaction multimodale ont été menés à GEOD pour une application de type "*smart room*" utilisée comme salle de réunion. Lorsqu'elle est issue d'une "*smart room*", la parole présente les spécificités suivantes:

- la parole de réunion est complètement spontanée, ceci conduit à la présence d'une grande quantité de phénomènes spécifiques à l'oral comme les hésitations, les faux départs (début d'un mot prononcé une ou plusieurs fois avant le mot complet), etc...
- on trouve souvent des recouvrements entre la parole de plusieurs locuteurs qui parlent en même temps,
- le bruit de fond peut être important (voix de fond, chuchotements, bruit de la ventilation d'un vidéoprojecteur, ...),
- il y a non plus un seul, mais potentiellement plusieurs enregistrements issus de multiples microphones disposés dans toute la salle
- dans le cas où les participants ne disposent pas d'un micro casque ou bouche, les locuteurs peuvent alors être à une distance importante des microphones placés au plafond ou sur la table; ceci a pour conséquence de diminuer la qualité des signaux enregistrés.

Cette recherche s'est concrétisée d'une part, dans la participation à un projet RNRT COUCOU (Conception participative Orientée Usage de services de Communication et d'objets Ubiquistes) entre 2002 et 2004 et, d'autre part, dans la participation à des campagnes d'évaluation NIST (SpeakerRec 2002, RT 2004 et RT 2005) consacrées aux enregistrements de réunions (meetings) (travaux déjà cités au §2.1.1)

Le projet COUCOU a été effectué en coopération avec l'équipe MULTICOM du CLIPS mais dans lequel intervenaient également le CNRS-MSH, le CEA-IDEAS LAB, France Télécom R&D et ST Microelectronics. Le but de ce projet était, entre autres, de spécifier et commencer à implémenter deux salles intelligentes de réunions (au CEA et au CLIPS) équipées d'outils avancés de prise de notes et de post-traitement des données enregistrées. L'implication de GEOD sur ce projet a concerné les aspects audio (AP-13, L. Besacier & A.C. Descalle, 2004). Ces spécifications ont permis à la "*smart room*" du CLIPS de voir le jour à partir d'octobre 2005 dans les nouveaux locaux du CTL<sup>15</sup> à Grenoble.

### **2.1.3. Développement de systèmes de dialogue H-M multimodaux**

Participants à ce thème de recherche : **J. Caelen** (Resp.), P. Dominguez, Y. Fouquet, S. Hollard, M. Kurdi, A. Lecomte, N.H. Nguyen, A. Ouayouch, V. Popescu, V.T. Tran, A. Xuereb.

Mots-clés : Dialogue HM, interaction HM, modèle de dialogue, négociation, acte de dialogue, attente en dialogue, évaluation automatique.

#### c) Résumé

Dans ce thème sont décrits les travaux de recherche de GEOD sur le dialogue homme-machine. Les principales avancées se sont concrétisées autour de la théorie des jeux et de la théorie de la représentation du dialogue (SDRT = Segmented Discourse Representation Theory). L'analyse et l'exploitation de corpus se sont poursuivies pour étudier les attentes des locuteurs, leurs modes de compréhension, leurs comportements face à des agents conversationnels expressifs. Pour cela diverses situations de dialogue ont été simulées notamment dans le cadre du projet ACE (Agent Conversationnel Expressif). Enfin la méthode DCR<sup>16</sup> a été approfondie pour obtenir une procédure validée en évaluation automatique de systèmes de dialogue. Toutes ces recherches ont été centrées sur des domaines applicatifs apportés par le projet PVE (Portal Vocal d'Entreprise) dont le but est de développer des services de dialogue en parole naturelle pour la vie sociale de l'entreprise (par exemple organisation de réunions, d'agendas personnels, etc.). Une ouverture vers le dialogue à plusieurs locuteurs a été amorcée, ce qui place l'équipe sur un terrain original.

#### d) Description détaillée

La description des apports des recherches est présentée selon quatre volets : (1) la théorie des jeux avec application à la gestion des tours de parole vue comme une résolution de conflits, (2) la modélisation du dialogue appréhendée comme un jeu stratégique, (3) l'interprétation pragmatique à travers la SDRT et (4) la

---

<sup>15</sup> Centre des Technologies du Logiciel (CTL)

<sup>16</sup> Les tests DCR (Demande Contrôle Référence) comportent une demande (D) de l'utilisateur, un énoncé de contrôle (C) et une référence (R), issue de la comparaison entre D et C (thèse Ahafhaf, 2004).

compréhension des énoncés sous le double aspect de la prise en compte des attentes dans le dialogue d'une part, et de l'évaluation automatique des systèmes d'autre part.

**(1) La résolution des conflits comme principe de jeu de dialogue dans la prise optimale des tours de parole.** L'usage de la langue naturelle pour les systèmes d'information automatisés dans les entreprises devrait connaître un certain essor. En effet, les personnels sont de plus en plus en situation de mobilité et souhaitent garder le contact, résoudre des problèmes urgents ou bénéficier des mêmes services qu'en situation de proximité : secrétariat, agendas partagés, accès à leurs dossiers, organisation de réunions, etc. Le téléphone mobile devient un vecteur de communication dans leurs déplacements. Dans le contexte du projet PVE<sup>17</sup>, l'analyse d'usage que nous avons faite dans des hôpitaux, des services d'administration universitaire, chez des professions libérales et des entreprises montre que les services vocaux sont très utiles dans les situations où l'opportunisme domine. Le dialogue oral dans ces situations s'apparente à un dialogue de résolution de problème en face à face : obtenir un accord, se coordonner pour une action, obtenir une information-clef pour débloquer une situation, etc. C'est à ce type de scénario que nous souhaitons apporter des solutions et des briques logicielles pour un système de dialogue homme-machine *orienté service*, et non plus seulement orienté tâche : il s'agirait d'un agent qui devrait avoir des capacités pour mémoriser les problèmes, pour rappeler plusieurs fois tous les interlocuteurs (les informer d'une réunion ou leur demander leurs disponibilités de dates puis leur confirmer la date et le lieu finalement choisis), pour collecter des demandes ou des contraintes, bref pour gérer des tâches multiples et de haut niveau impliquant plusieurs utilisateurs en gérant les conflits de ressources éventuels.

Nous avons simulé l'usage d'un tel service en *magicien d'Oz*. Il apparaît que le dialogue s'organise alors comme une suite de sessions entre deux interlocuteurs en face à face. Ces sessions sont généralement courtes et ne contiennent qu'un seul problème à traiter, mais ce problème peut être complexe.

L'enchaînement des sessions pose lui aussi des problèmes nouveaux si on le confie à une machine : garder le problème principal en mémoire, planifier les sessions, prendre des initiatives dans le dialogue et savoir négocier les appels. Ainsi, le modèle de dialogue doit-il fournir une description générale des diverses situations liées au service considéré et représenter les liens entre les diverses phases de négociation avec tous les intervenants.

L'exemple de service d'assistance à l'organisation d'une réunion permettra d'illustrer le type de dialogue participants/système que nous visons. Le dialogue se déroule en sessions durant lesquelles un seul participant est en conversation avec la machine pour communiquer ses dates de disponibilité ou accepter/décliner une invitation, etc. Dans ce cas, la gestion des sessions se fait par la résolution successive des conflits : lorsqu'un conflit apparaît on tente de le résoudre en ouvrant une(des) session(s) de dialogue avec le(tous les) participant(s) concerné(s). En fin de cycle on revient vers le demandeur pour l'informer de la situation ou lui demander de nouvelles instructions. Le système a été développé sur la base de la théorie des jeux. Il a été expérimenté pour une application de réservation de salle dans laquelle il y a de fréquents conflits de ressources (Thèse N.H. Nguyen, 2005). La ressource partagée est un agenda de salles de réunion. On a obtenu le type de dialogue suivant :

#### ***Session du demandeur***<sup>18</sup>

U : Ici Dupuis, réserve-moi la salle Lafayette pour demain, de 10h à midi
M : Désolé, cette salle est réservée demain par Monsieur Dupont de 9h à 11h et par Monsieur Dupin de 11h à 13h
U : Demande-leur de changer de salle
M : D'accord, je les appelle et je vous tiens au courant

#### ***Session 1***

M : Allô ? Monsieur Dupont ?
P1 : Oui, c'est moi
M : Monsieur Dupuis a une réunion importante demain de 10h à midi. Il vous demande si vous pouvez lui laisser la salle Lafayette demain matin
P1 : Attendez je regarde... oui c'est OK reportez ma réunion à mardi prochain
M : Merci, au revoir

#### ***Session 2***

M : Allô ? Monsieur Dupin ?
P2 : Oui, c'est pour quoi ?

<sup>17</sup> PVE : Portail Vocal d'Entreprise, projet RNRT [www.telecom.gouv.fr/rnrt/projets/res\\_01\\_5.htm](http://www.telecom.gouv.fr/rnrt/projets/res_01_5.htm)

<sup>18</sup> U désigne l'utilisateur initial, M le système (la machine), et Pn les autres participants.

M : Monsieur Dupuis a une réunion importante demain de 10h à midi. Il vous demande si vous pouvez lui laisser la salle Lafayette demain matin  
P2 : OK je peux changer de salle. Trouvez-m'en une autre  
M : La salle Aquarium est disponible demain de 11h à 13h, elle est un peu plus petite  
P2 : Entendu réservez-la-moi  
M : Merci de votre compréhension. Au revoir

### **Session compte-rendu**

M : Allô ? Monsieur Dupuis ?  
U : Oui, alors vous avez la solution ?  
M : Oui, Monsieur Dupont et Monsieur Dupuis ont accepté de laisser la salle Lafayette. Elle sera disponible demain pour vous de 9h à 11h. Je la réserve ?  
U : Oui, merci, prévenez aussi les membres du projet PVE  
M : Entendu, je le fais par e-mail, vous en aurez une copie. Au revoir.

On remarque dans le déroulement de ce dialogue multi-session que la machine n'a pas vocation à négocier elle-même en se substituant à l'un quelconque des intervenants, mais bien à choisir une tactique d'appel des correspondants puis à leur présenter le problème qu'ils sont tenus de résoudre eux-mêmes. L'exemple donné ici est simplifié dans la mesure où le conflit est résolu en explorant un seul chemin de l'arbre de conflit. Il est facile d'en imaginer de plus complexes.

Ce petit exemple ne montre pas non plus les effets de "ricochet" qui ne manquent pas de se produire souvent, lorsque par exemple Dupuis répond à la machine "n'y a-t-il pas une autre solution ?" ou "pouvez-vous demander à Monsieur Dupin d'abord ?", etc. D'autres effets en chaîne peuvent également se produire lorsque les intervenants répondent par "je suis d'accord à condition que..." ce qui nécessite de traiter la condition d'abord avant de revenir dans le dialogue. Le traitement de cette condition peut d'ailleurs être fort complexe et nécessiter à son tour l'intervention de plusieurs participants. Actuellement le système ne prend pas en compte ces difficultés. La machine collecte les informations et conditions et les présente au demandeur qui prend une décision lui-même, puis qui confie éventuellement un deuxième tour de négociation à la machine.

Le système a été programmé et un démonstrateur est opérationnel. C'est un premier pas vers un système de dialogue à plusieurs, à initiative partagée, car il ne s'agit encore que d'un système qui règle les tours de parole entre plusieurs intervenants et qui contrôle les contacts en cas de conflit. Le système est bien adapté aux portails vocaux dont l'accès aux ressources est concurrent. Il permet de gérer les ressources partagées comme les salles de réunion, les prises de rendez-vous, etc. Ce modèle de gestion de la négociation est relativement indépendant du modèle de tâche. Il travaille seulement sur l'état du but de dialogue et il permet donc de séparer le gestionnaire de dialogue du gestionnaire de la tâche, c'est-à-dire le *quoi* et le *comment*.

**(2) La modélisation du dialogue comme un jeu stratégique.** Les stratégies de dialogue sont des manières d'atteindre un but à travers le dialogue considéré comme une activité conjointe (Vernant, 1992)<sup>19</sup>. Même si, au départ d'un dialogue, le locuteur et l'allocutaire n'ont pas les mêmes objectifs, ils partagent au moins l'intention de maintenir la conversation et visent un certain *but conversationnel*. Dans tous ces cas, le dialogue se déroule dans un *cadre* qui se définit par les éléments suivants :

- (a) le but conversationnel (ou intentionnalité collective), qui donne la finalité de la conversation,
- (b) le déroulement,
- (c) le thème de la conversation et
- (d) l'arrière-plan de la conversation (situation, monde de la tâche s'il y en a une, rôles sociaux, etc.).

Il faut distinguer le but du dialogue qui est dans l'arrière-plan, du but conversationnel qui lui, est nécessairement partagé (s'il ne l'est pas il y a malentendu sur le type de dialogue). Le but conversationnel peut-être satisfait sans que le but d'arrière-plan le soit nécessairement. Dans un cadre donné, c'est-à-dire pour (a), (c) et (d) donnés, le déroulement va dépendre des stratégies utilisées par les interlocuteurs pour arriver à satisfaire le but conversationnel, et de manière réflexive, c'est la structure de déroulement qui va contraindre à son tour la stratégie que les interlocuteurs pourront utiliser. Le calcul d'une stratégie  $\delta$  se fait à l'aide de règles qui tiennent compte de la complétude d'un acte de dialogue (s'il est incomplet il faut peut-être tenter de le compléter en posant une question adéquate avant d'agir), de la possibilité d'atteindre un but, des conditions de réussite de l'acte (la situation est-elle favorable ?), de l'état du dialogue (un autre but est-il en attente ?), des attentes de l'utilisateur (quel but poursuit-il ?), de ses compétences (est-il expert ou non ?), de la stratégie précédemment

---

<sup>19</sup> D. Vernant, "Approche actionnelle et modèle projectif du dialogue informatif", in Recherches sur la philosophie et le Langage, n°. 14, *Du Dialogue*, 352 p., 1992.

adoptée (on a peut-être pas intérêt à changer trop souvent de stratégie pour ne pas égarer l'utilisateur), etc. Ces calculs s'appuient d'une part sur une règle de déclenchement au moment le plus opportun dans le dialogue (ou supposé tel), et d'autre part sur la logique d'enchaînement des tours de parole pour toutes les situations possibles. L'opportunité de déclencher telle ou telle stratégie dépend de la profondeur de l'échange, de l'attitude de l'utilisateur et du contenu de certains actes.

**Règle de déclenchement** : la stratégie devient réactive si la profondeur  $\pi$  dépasse un certain seuil. Ce mode est également activé en cas de choix de l'utilisateur ou pour conclure un dialogue {exemple de règle formalisée :  $(\pi > \pi_0) \vee F_U^S(\delta_{réactif}) \vee F_U^S(\text{clôture}) \supset \delta_{réactif}$  }.

**Comportement** : Les conditions de complétude, de vérité et de réussite sont supposées vérifiées, puisqu'on ne remet pas en question les demandes de l'utilisateur qui les assume. En cas d'incomplétude du contenu propositionnel  $p$  de l'acte de  $U$ , on le complète par défaut, en considérant le profil et les préférences de l'utilisateur. Cette procédure n'est pas explicitée ici. Les règles de comportement de la machine à développer pour cette stratégie sont uniquement celles pour lesquelles  $U$  a l'initiative (par définition  $M$  n'a jamais l'initiative pour cette stratégie). L'ensemble du formalisme est décrit dans la thèse de Y. Fouquet (2004).

Le dialogue homme machine (DHM) est traité ici comme une co-interaction verbale à propos d'une tâche. La logique utilisée s'appuie sur les croyances, les buts et les actes et contient implicitement une logique déontique (attentes "sociales" des partenaires). En faisant varier les stratégies dialogiques on aboutit à des dialogues relativement naturels. C'est donc à travers un cadre opératoire et actionnel que nous fondons notre démarche. Ce cadre permet de donner une assise plus générique à l'interaction homme machine : on imagine donc assez aisément que ce modèle peut s'adapter à toute forme d'interaction non langagière.

**(3) L'interprétation pragmatique.** Ce niveau de traitement dans le dialogue homme-machine reste le plus difficile. Notre but est de concevoir et de réaliser une application complète de dialogue homme-machine. Pour cela nous avons développé un système de dialogue modulaire (ACT-40, H. Nguyen & J. Caelen, 2003) de façon à rendre certains composants indépendants de la tâche et donc, à terme, portables dans d'autres types d'applications. Il comprend notamment un composant de reconnaissance vocale (développé par IBM), un composant de compréhension sémantique, un composant d'interprétation pragmatique, un contrôleur de dialogue, un gestionnaire de tâche, un générateur d'énoncés et un module de synthèse vocale.

On suppose pour la suite que le module de compréhension automatique fournit une représentation sémantique logique de l'énoncé d'entrée, qui peut être entachée d'erreurs. Le sens communiqué par l'énoncé, qui est différent du sens littéral de l'énonciation, n'est pas complètement codé dans l'énoncé (Sperber, Wilson, 1986)<sup>20</sup>. Ainsi, un processus d'interprétation pragmatique est nécessaire et ne peut se faire qu'en intégrant l'interprétation de l'énoncé dans le contexte et l'historique, en construisant la structure logique du dialogue, de manière à l'interpréter plus complètement :

- lier les indexicaux (déictiques) au contexte de l'énoncé (1) et résoudre les ellipses,
- résoudre les problèmes liés à l'anaphore (2) et à la cataphore,
- traiter les présuppositions et les implicatures conversationnelles (3) (dans l'exemple il faut présupposer qu'il s'agit de participants à une réunion et que prévenir implique envoyer un message).

La SDRT (*Segmented Discourse Representation Theory*) (Asher, Lascarides, 2003)<sup>21</sup> est un cadre dans lequel peut se formaliser l'interprétation dynamique de l'énoncé en contexte. Elle étend la structure de base de la DRT (*Discourse Representation Theory*) (Kamp, Reyle, 1993)<sup>22</sup> (référents, prédicats, accessibilité) en la complétant par des relations rhétoriques liant les segments du discours. Nous l'avons utilisée pour nos travaux en y intégrant plusieurs ajouts et adaptations dans le cadre spécifique du dialogue homme-machine finalisé : les six relations rhétoriques retenues sont identifiables par des critères calculables, une structure — la SDRS globale — est le contexte commun aux deux interlocuteurs (ce qui a été dit, mais aussi ce qui est implicite). Nous avons validé manuellement ce modèle sur le corpus : ces six relations rhétoriques permettent de résoudre les questions-réponses, en particulier les réponses partielles et inférentielles et les élaborations de questions. Nous avons vérifié la résolution des anaphores. Nous avons seulement esquissé le traitement des présuppositions, nous travaillons maintenant sur leur formalisation, ainsi que sur la modélisation des implicatures conversationnelles.

Nous avons aussi spécifié un prototype informatisé : les règles de calcul des relations rhétoriques sont basées sur la typologie des actes de langage et les propriétés sémantiques du contenu propositionnel. Nous

<sup>20</sup> D. Sperber & D. Wilson, *Relevance. Communication & Cognition*. Blackwell, 1986.

<sup>21</sup> N. Asher & A. Lascarides, *Logics of Conversation*. Cambridge University Press, 2003.

<sup>22</sup> H. Kamp & U. Reyle, *From discourse to Logic*, Kluwer Academic Publishers, 1993.

utilisons une logique monotone. Le moteur de l'interpréteur utilise un raisonnement hypothétique limité à un seul tour de parole : pour chaque tour de parole, après insertion de la DRS courante (force illocutoire et contenu propositionnel), les sites d'attachement disponibles sont calculés ainsi qu'une hypothèse de relation pour chaque nœud encore non étiqueté. Les inférences sont déclenchées sur la base de ces hypothèses pour tenter une résolution. L'hypothèse est alors acceptée ou refusée suivant le succès ou l'échec de cette résolution. Une hypothèse acceptée (le nœud est étiqueté) ne sera plus réévaluée au tour suivant.

Nous avons également travaillé sur la notion étendue de topique. La question du topique en SDRT classique est traitée en terme de relation rhétorique. Nous proposons de considérer le nœud topique non seulement comme élément structurant du discours (dialogue) mais également comme le réceptacle de représentations pragmatiques en cours de calcul. Nous avons étudié cette "extension" de la SDRT qui consiste à introduire plus systématiquement un nœud topique dans la SDRS (Discourse Representation Structure), en subsumant toute relation rhétorique pouvant contenir des présuppositions ou des implicatures conversationnelles ou des effets projectifs dans les énoncés liés (ACT-103, A. Xuereb & Caelen, 2005). La SDRT utilise la notion de relation rhétorique pour structurer le discours. Cette notion se fonde quelque peu sur la notion de paire adjacente issue des théories de la conversation (Goffman, 1967)<sup>23</sup> dans laquelle tout acte de langage tente de "fermer" une paire ouverte. Appliquée au dialogue, cette vision a tendance à l'enfermer dans un système de résolution des attentes. Il nous semble cependant que dans la perspective d'un modèle projectif (Vernant, 1997)<sup>24</sup> ou dans celle de la logique interlocutoire (Trognon, 1995)<sup>25</sup>, chaque acte est projeté vers le futur et prend sa signification dans un *interacte* construit de manière émergente entre les acteurs dans le dialogue. Il s'agit donc plutôt de "projeter" le dialogue en avant à chaque instant, chaque acteur prenant sa part dans l'action mais aussi en en déléguant une partie à autrui. La mesure des effets de ces actions devient alors primordiale pour la poursuite du dialogue et la coordination mutuelle.

Pour cela nous avons mis au point un mécanisme original. L'introduction d'un nouveau nœud topique est déclenchée par la détection d'un changement de thème. Les thèmes sont ceux du domaine, réunis dans une ontologie et organisés sous forme d'arbre. Ces thèmes sont eux-mêmes liés au modèle de tâche dont nous ne parlerons pas dans cette partie. L'arbre des tâches permet d'inférer les implications et de poser les hypothèses projectives qui sont reliées à la structure pragmatique. Les nœuds d'accroche sont les nœuds topiques (pointeurs entre structures). A titre d'exemple, l'arbre des thèmes pour un domaine restreint de PVE est donné figure 6.

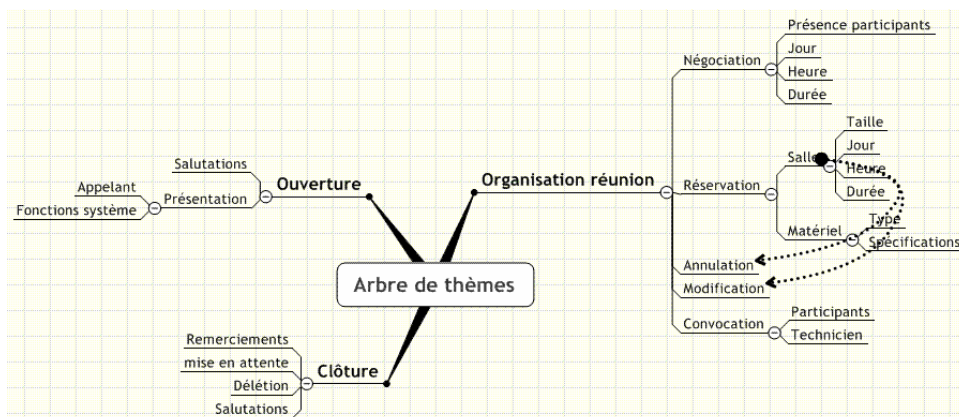


Fig. 6 : Arbre de thèmes pour la tâche "organisation de réunion" dans l'application PVE (Portail Vocal d'Entreprise).

D'autre part, nous cherchons à modéliser des phénomènes pragmatiques qui, jusqu'à présent, s'exprimaient dans une perspective "gricéenne" difficilement formalisable (ACT-86, A. Lecomte & A. Naït-Abdallah, 2003 et ACT-101, A. Lecomte & A. Naït-Abdallah, 2005). L'approche proposée repose sur l'idée que la grande majorité des dialogues se déroulent dans des situations d'*information partielle* : les réponses fournies à une demande de renseignement par exemple, ne sont jamais "exhaustives", elles laissent presque toujours la place à des demandes d'information supplémentaires et l'échange se clôt par l'accord des participants sur ce qu'il convient de considérer comme une réponse suffisamment informative.

<sup>23</sup> E. Goffman, *Interaction rituals : Essays on face-to face Behavior*, Anchor books, N. Y., 1967.

<sup>24</sup> D. Vernant, *Du discours à l'action, études pragmatiques*, PUF, Paris, 1997.

<sup>25</sup> A. Trognon, *Structures interlocutoires*, Les cahiers de linguistique française, n°17, 1995.

Ce type d'accord est celui qui se produit également dans des situations de recherche de solution négociée où, éventuellement, *une partie de l'information manque* à l'un des participants au dialogue parce que l'autre la lui cache volontairement. Chaque locuteur en présence "calcule" alors sa réplique en fonction de ses intérêts en fabriquant un modèle "de coût minimal" pour lui-même. La *logique de l'information partielle* d'Areski Naït-Abdallah (ACT-101, 2005) est alors un outil commode pour modéliser ces situations, on peut envisager de la combiner avec la *IF-logique (Independance-Friendly Logic)* de Hintikka<sup>26</sup>, qui possède une sémantique naturelle en termes de jeux. De telles entreprises reposent sur une perspective en logique autant basée sur une sémantique *actionnelle* (cf. la notion de jeu) que dénotationnelle.

**(4) la compréhension : problèmes liés aux attentes et évaluation automatique de systèmes.** Les attentes du locuteur semblent une alternative intéressante à la simple prédiction d'actes de dialogue dans le cadre du dialogue oral homme-homme (ACT-78, Y. Fouquet, 2002). Nous avons étudié ces attentes en dialogue homme-machine (dans le domaine d'un portail vocal d'entreprise PVE). Un système de dialogue homme-machine (décrit en détail dans COM-6, Y. Fouquet, 2003) a été développé, dans lequel s'intègrent les attentes des interactants. Une expérimentation à base de magicien d'Oz a permis de collecter des données linguistiques.

Le développement du système de dialogue oral PVE suit la méthodologie développée auparavant par Rouillard & Caelen en 1998 et décrite dans la thèse de J. Rouillard en 2000. Nous nous sommes appuyés sur l'enregistrement de dialogues réels entre des secrétaires et leurs interlocuteurs. Parmi les 800 dialogues enregistrés, nous avons repéré les dialogues les plus représentatifs des tâches considérées comme importantes pour un assistant virtuel : *joindre une personne, prendre un rendez-vous, réserver une salle, gérer un agenda partagé, recevoir une information et envoyer un document*. L'analyse de ces dialogues, première pierre de la méthodologie, nous a permis de développer un premier modèle de dialogue offrant 122 phrases typiques que le système devra être capable de dire.

Pendant, l'analyse de dialogues inter-humains est insuffisante pour créer un système de dialogue homme machine. Nous avons donc développé une plate-forme magicien d'Oz afin de simuler le comportement du système (contrôle du dialogue et de la tâche) et d'élaborer un corpus de dialogue oral homme-machine. L'analyse de ce corpus a montré des différences structurelles et dialogiques entre le dialogue oral inter-humain et le dialogue oral homme-machine pour les tâches ci-dessus. Dans cette approche, nous avons adopté une démarche stochastique pour prédire l'acte suivant à partir d'un historique plus ou moins grand. Notre corpus est composé de dialogues réels mais, comme usuellement dans de nombreuses situations réelles, il n'est pas possible de collecter un très grand nombre de données pour estimer correctement les probabilités. Il n'est donc pas raisonnable d'utiliser les techniques classiques. Nous devons trouver une solution pour deux problèmes importants. Le modèle de dialogue fondé sur les séquences de n-grams actes de dialogues ne peut pas être utilisé dans le cas d'une entrée non attendue. Les modèles à base de n-grams peuvent prédire plusieurs actes avec la même probabilité dans le cas d'un manque de données d'apprentissage. Le traitement de ces cas consiste, dans notre expérimentation, à utiliser le modèle (n-1)-gram et récursivement.

L'approche fondée sur la théorie des attentes met l'accent sur la compréhension par la machine des attentes de l'utilisateur, au-delà de la seule compréhension de l'énoncé en cours. La reconnaissance automatique des attentes est cependant un problème non évident. D'une part, il existe un grand nombre de manières de laisser transparaître une attente. D'autre part, les attentes sont souvent non marquées linguistiquement. Le modèle utilisé ici (ACT-78, Y. Fouquet, 2002) est dérivé de la théorie des actes de langages et utilise des statistiques de bigrammes. Les attentes sont une liste de réponses que l'utilisateur est susceptible d'attendre lorsqu'il formule un énoncé. L'identification du locuteur ainsi que le contenu propositionnel sont adjoints à cette notation de sorte que l'acte de dialogue résultant a pour forme :  $F_{\cup}^{fs}(p)$ .  $F^{fs}$  est l'acte de langage.  $\cup$  est l'identifiant de l'utilisateur,  $A$  pour agent, ici la machine et "(p)" est une représentation logico-sémantique du contenu propositionnel. Nos résultats confirment que l'approche de la gestion de la compréhension dans un système de dialogue à l'aide des attentes est plus performante.

## 2.2 Résultats majeurs

Les acquis de GEOD sont des *plates-formes matérielles et logicielles* qui ont été développées ou dont GEOD a acquis la maîtrise au cours de ses activités de recherche, mais aussi des *corpus* (écrits et oraux) produits ou recueillis au cours de divers travaux de recherche et réutilisables dans d'autres applications ou projets. Parmi les résultats principaux on peut citer :

---

<sup>26</sup> J. Hintikka, *The Principles of Mathematics revisited*, Cambridge University Press, N.Y., 1996.

• JANUS : Tous les systèmes de reconnaissance automatique de la parole du laboratoire ont été développés grâce à la partie *JANUS-Recognition-Toolkit (JRtk)* de la boîte à outils Janus III, développé dans le laboratoire ISL<sup>27</sup> (Interactive System Laboratory) des universités de Karlsruhe (Allemagne) et Carnegie Mellon (USA). La plate-forme Janus contient tous les composants nécessaires au développement d'un système de reconnaissance phonémique à base de Modèles de Markov Cachés. Janus offre un langage de programmation puissant qui permet de manipuler les structures de données internes et d'écrire des procédures complexes de haut niveau. Le langage de programmation de Janus utilise Tcl/Tk étendu avec des classes d'objets et leurs méthodes.

• RAPHAEL : système de "Reconnaissance Automatique de Phrases, d'Acronymes et d'Expressions Langagières" est le système de reconnaissance automatique de parole continue "*de référence*" du CLIPS (vocabulaires : de 6000 à 65000 mots), construit sur la plate-forme logicielle JANUS III. Il a été mis en œuvre avec succès dans divers contrats internationaux (CSTAR, NESPOLE) de dialogue homme-machine ou homme-homme médiatisé en contexte multilingue.

• ELISA : C'est une plate-forme commune de développement conçue par le consortium ELISA. Le consortium ELISA a été fondé en 1997 par le LIA (Avignon), l'ENST (Paris) et l'IRISA (Rennes), autofinancé par ses participants, avec pour objectif de faciliter les recherches coopératives en reconnaissance du locuteur ainsi que la participation des laboratoires francophones aux campagnes d'évaluations internationales telles que les campagnes NIST. Le laboratoire CLIPS (GEOD) a rejoint le consortium en 2001 et utilise cette plate-forme pour ses recherches depuis cette époque. La partie modélisation de la plate-forme ELISA permet d'estimer les paramètres de modèles de mélanges de gaussiennes (GMM) avec des matrices diagonales ou pleines. Plusieurs méthodes d'adaptation de modèles par maximum a posteriori (MAP) sont également disponibles avec cette plate-forme. Le module de décision permet quant à lui le calcul de vraisemblances associées à chaque modèle pour un corpus de test considéré. Actuellement, la plate-forme est en phase de ré-écriture sous la forme d'une boîte à outils (en licence GPL) nommée ALIZE<sup>28</sup> et une première version est déjà disponible sous forme de logiciel libre.

• EMACOP : "Environnement Multimédia pour l'Acquisition de CORpus de Parole", développé sur une architecture client-serveur, permet de conduire, de façon fiable et reproductible, de larges sessions d'enregistrement de corpus. Ce logiciel est utilisé dans toutes les campagnes d'enregistrement de corpus oraux nécessaires pour l'apprentissage ou l'évaluation des systèmes et des versions ont été adaptées à certaines langues étrangères (vietnamien, khmer).

• MELINA : Plate-forme d'intégration d'une infrastructure de gestion générique de dialogue dédié au service d'organisation de réunions, composée de sept modules principaux : reconnaissance de parole, compréhension sémantique, interpréteur pragmatique, contrôleur du dialogue intégrant un module *Expert* de gestion de négociations, contrôleur de la tâche, générateur textuel, et synthétiseur de parole. L'architecture possède les fonctionnalités permettant d'aborder la gestion de dialogue multi-sessions.

• CORPUS : Dans cette catégorie, en ne citant que les principaux, nous pouvons distinguer les *corpus produits* (BRAFI100, SONS de la vie courante, ANODIN/DETRESSE) et les *corpus recueillis* ou *achetés* ( BREF80, BREF120, AURORA, LE MONDE : années 1987-1991, 1992-1995, 1996-1999, 2000-2002).

### 3. ACTIVITES D'ENCADREMENT

#### 3.1 Thèses soutenues depuis le 01/10/01

Liste des thèses soutenues (du 01/10/2001 au 01/10/2005)

Nom Prénom	Directeur(s) de thèse	Date de soutenance	Mode de financement	Etablissement d'inscription
VAUFREYDAZ Dominique	Jean CAELEN	7/01/2002	A MENRT	UJF (0381838S)
NGUYEN Quoc-Cong	Eric CASTELLI	19/06/2002	BDI-CNRS	INPG (0381912X)
ISTRATE Dan	Eric CASTELLI Laurent BESACIER	15/12/2003	ACI-CNRS	INPG (0381912X)

<sup>27</sup> <http://www.is.cs.cmu.edu/js/>

<sup>28</sup> <http://www.lia.univ-avignon.fr/heberges/ALIZE>

KURDI Mohamed Zakaria	Jean CAELEN	18/04/2003	Vacat. Univ.	UJF (0381838S)
AHAFFHAF Mohamed	Jean CAELEN	30/09/2204	CDD privé	Stendhal-U3 Grenoble (0381840U)
FOUQUET Yannick	Jean CAELEN	26/10/2004	COLL TERR Région RA	UJF (0381838S)
MORARU Daniel	Eric CASTELLI Laurent BESACIER	20/12/2004	A MENRT	INPG (0381912X)
NGUYEN Ngoc-Hoa	Jean CAELEN	26/01/2005	Contrat RNRT (PVE)	UJF (0381838S)
MAYORGA-ORTIZ Pedro	Jean CAELEN Laurent BESACIER	8/02/2005	ETR CONACYT	INPG (0381912X)

### 3.2 Thèses et HDR en cours

*Thèses :*

CHAILLOL Stéphane, *Détection et classification de sons dans un habitat intelligent pour la santé*, Thèse CNAM, contrat formation Assedic, inscription 01/09/05.

DOMINGUEZ Patricia, *Modélisation du dialogue H-M, généralité*, ED MSTII, UJF, bourse CONACYT, première inscription 01/10/04.

LE Viet-Bac, *Portabilité rapide d'un système de reconnaissance automatique de la parole à des langues faiblement dotées en corpus*, ED MSTII, UJF, bourse d'excellence Eiffel, première inscription 01/10/02.

OUAYOUCH Anas, *Jeu de négociation en langage naturel*, ED MSTII, UJF, salarié, première inscription 01/10/05.

POPESCU Vladimir, *Dialogue pertinent Homme-Machine : génération des réponses orales en langage naturel*, ED MSTII, UJF, (Co-tutelle Fac. Electronique Bucarest), bourse MIRA, première inscription 01/10/05.

TAN Tien-Ping, *Reconnaissance du français pour des locuteurs non natifs*, ED MSTII, UJF, bourse du gouvernement Malais, première inscription 01/10/04.

TRAN Do-Dat, *Synthèse à partir du texte en vietnamien*, ED SIPT, INPG, bourse AUF, première inscription 01/10/03.

TRAN Vu-Truc, *Pragmatique computationnelle*, ED EDISCE, UPMF, (Co-tutelle INP-Hanoi), bourse MAE, Ambassade de France au Vietnam, première inscription 01/10/04.

VU-MINH Quang, *Détection de zones d'intérêt intonatives dans un flux audio*, ED MSTII, UJF, (Co-tutelle INP-Hanoi), bourse MAE, Ambassade de France au Vietnam, première inscription 01/10/03.

*HDR :*

BESACIER Laurent, *Transcription enrichie de documents dans un monde multilingue et multimodal*, ED MSTI, UJF, soutenance 2006.

## 4. COLLABORATION ET VALORISATION

### 4.1 Principales relations scientifiques hors contrats

GEOD entretient des relations scientifiques suivies depuis plusieurs années avec le centre MICA créé en 2001 à l'Institut Polytechnique de Hanoi (IPH). A l'origine, la création de cet Institut de Recherche fut un essaimage du laboratoire CLIPS sous l'impulsion et l'organisation d'Eric Castelli, en détachement CNRS et directeur-adjoint de ce centre depuis cette époque. Cette coopération se concrétise par l'accueil réciproque de chercheurs invités et par des thèses en co-tutelle .

Accueil de chercheurs de GEOD au centre MICA :

- Jean Caelen - visite le 12 mars 2002, puis 1 à 2 semaines chaque année.
- Dominique Vaufraydaz - du 24 août au 14 septembre 2002.
- Laurent Besacier - du 16 novembre au 29 novembre 2002.
- Viet Bac Le - du 1<sup>er</sup> juin au 31 juillet 2003.
- Laurent Besacier - Février 2005.
- Brigitte Bigi - du 14 octobre au 30 octobre 2005.

Accueil de chercheurs du centre MICA à GEOD : depuis la création du centre Mica à Hanoi, fin 2001, Eric Castelli effectue régulièrement des séjours à Grenoble, de une à deux semaines par an, pour coordonner la coopération scientifique entre les deux groupes. D'autres chercheurs de MICA ont également été accueillis à GEOD :

- *Eric Castelli et Dr Pham Thi Ngoc Yen - séjour d'une semaine en mars 2003.*
- *Dr Loan TRINH VAN - séjour d'une semaine en novembre 2004.*

Dans le cadre de la création par le CNRS du Laboratoire Franco-Mexicain d'Informatique (LAFMI), en coopération avec l'INAOE<sup>29</sup> à Puebla (Mexique) :

- *Dominique Vaufreydaz, séjour à l'INAOE en février 2002.*
- *Luis Villaseñor-Pineda, séjour à Grenoble en novembre 2002.*
- *Aurélio Lopez, séjour à Grenoble en octobre 2003.*
- *Jean Caelen, séjour à l'INAOE en octobre 2003.*
- *Luis Villaseñor-Pineda, séjour à Grenoble en juin-juillet-août 2004.*
- *Jean Caelen, séjour à l'INAOE en novembre 2004 et septembre 2005.*

Autres échanges nationaux ou internationaux :

- *Viet Bac Le, mission à CMU, Pittsburgh - du 28/02/05 au 24/03/05.*
- *Laurent Besacier, mission 1 semaine à l'ITC, Phnom Penh, Cambodge, février 2005.*
- *Tien-Ping Tan, mission au LPL, Aix-en-Provence, du 6/11/05 au 2/12/05.*
- *Jean Caelen, conférences invitées en Roumanie en 2003 et 2004, et co-tutelle de thèse en 2005.*
- *Laurent Besacier, mobilité internationale à IBM, New York, oct. 2005 à sept. 2006.*
- *Sethserey Sam, enseignant-chercheur de l'ITC, Phnom Penh, Cambodge, séjour à Grenoble, 1/10/05 au 23/03/06.*

Thèses en co-tutelle en cours entre GEOD (INP et UJF, Grenoble) et le centre MICA (INP Hanoi) :

- *Do-Dat TRAN, Synthèse à partir du texte en vietnamien, première inscription 01/10/03.*
- *Vu-Truc TRAN, Pragmatique computationnelle, première inscription 01/10/04.*
- *Quang VU-MINH, Détection de zones d'intérêt intonatives dans un flux audio, première inscription 01/10/03.*

GEOD accueille régulièrement des doctorants roumains grâce aux contacts privilégiés entretenus avec le Professeur Corneliu Burileanu, de la Faculté d'Electronique de Bucarest. Cette coopération scientifique s'est concrétisée par la soutenance de deux thèses effectuées à GEOD et une thèse en co-tutelle en cours :

- *Dan ISTRATE, Détection et reconnaissance des sons pour la surveillance médicale, Thèse INPG-SIPT, soutenue à Grenoble le 15 décembre 2003.*
- *Daniel MORARU, Segmentation en locuteurs de documents audios et audiovisuels : application à la recherche d'information multimedia, Thèse INPG-SIPT, soutenue à Grenoble le 20 décembre 2004.*
- *Vladimir POPESCU, Modèle de dialogue, co-tutelle en cours, première inscription 01/10/05.*

## **4.2 Contrats institutionnels**

GEOD a été impliqué dans plusieurs contrats institutionnels, nationaux et internationaux. Une présentation synthétique des contrats les plus marquants est donnée ci-après :

### **Contrats internationaux**

#### *Projet Européen NESPOLE & Consortium C-STAR*

Depuis 2000, GEOD s'est impliqué, en collaboration avec le GETA, dans deux projets de traduction automatique de parole dont le CLIPS était partenaire : CSTAR (*Consortium for Speech Translation Advanced Research*) et NESPOLE (*Negotiating Through SPOken Language in E-commerce*). Le but des recherches menées au sein de ces projets était la traduction automatique de parole spontanée avec tous les couples de langues possibles entre les différents partenaires. La contribution de GEOD était surtout de fournir l'étape de reconnaissance automatique de la parole pour le français, développé avec le système RAPHAEL.

#### *Action COST275 (Biometric Person Authentication over the Internet)*

Entre 2001 et 2004, GEOD a participé à l'action européenne COST 275 traitant de la biométrie. Sa contribution s'est située notamment au niveau du sous-groupe "évaluation" de l'action, qui a été co-animé par L. Besacier avec J-F Bonastre (LIA-Avignon).

---

<sup>29</sup> Institut National d'Astrophysique, Optique et Electronique (INAOE)

*Projets CORUS (Traitement de la parole en langue vietnamienne) et TALK (Traitement Automatique de la Langue Khmère)*

Ces deux projets, soutenus par le MAE (Ministère des Affaires Etrangères) ou par l'AUF (Agence Universitaire pour la Francophonie), concrétisent une collaboration qui existe depuis presque 5 ans entre le CLIPS et le centre MICA à Hanoi, et (plus récemment) l'Institut de Technologie du Cambodge (ITC) sur le domaine du traitement automatique du vietnamien et du khmer. L'implication de GEOD dans ces projets concerne d'une part, la reconnaissance automatique de la parole pour le vietnamien et le khmer et d'autre part, la synthèse du vietnamien.

### **Contrats nationaux**

*Projets RNRT PVE (Neurosoft, CLIPS, Novadis, IBM), 2002-03 et COUCOU (CNRS-MSH, CLIPS, CEA, FT-R&D, ST Microelectronics), 2002-04*

Le laboratoire CLIPS (équipes MULTICOM et GEOD), a participé à deux projets RNRT exploratoires, PVE (*Portail Vocal d'Entreprise*) entre 2002 et 2003 et COUCOU (*Conception participative Orientée Usage de services de Communication et d'Objets Ubiquistes*) entre 2002 et 2004. Dans le premier, la contribution de GEOD s'est située au niveau de la mise en œuvre de procédures d'expérimentation et d'évaluation ergonomique pour la qualification des performances linguistiques d'un système de dialogue vocal, notamment avec la méthode DQR. Cette expérimentation s'est faite dans le cadre de la réalisation d'un démonstrateur d'un système de dialogue générique pour une classe d'applications de communication en entreprise (interrogation de l'annuaire du personnel, gestion de l'agenda des groupes de travail et gestion personnelle du courrier électronique). Le projet COUCOU, auquel GEOD a participé en collaboration avec l'équipe MULTICOM du CLIPS, a permis de spécifier et commencer à équiper deux salles intelligentes de réunions (au CEA et au CLIPS) équipées d'outils avancés de prise de notes et de post-traitement des données enregistrées. La contribution de GEOD à ce projet a été focalisée principalement sur les aspects audio (équipement et traitement) d'une salle intelligente.

*Projet RNTL pré-compétitif ACE (As An Angel, La Cantoche, CLIPS), 2003-04*

Le projet ACE avait pour but de produire un prototype d'un agent conversationnel expressif (Angela 1.0) destiné à faciliter la communication avec l'utilisateur dans le cadre d'un dialogue homme-machine. La contribution de GEOD, en coopération avec l'équipe MULTICOM, se situe au niveau de la modélisation du système de dialogue, du recueil de corpus de dialogues finalisés dans le cadre d'une expérimentation avec magicien d'Oz.

*Projet Technolangue AGILE/ALIZE (CLIPS, LIA, DDL, IRISA, IRIT, ENST), 2003-04*

Le projet ALIZE consistait à réaliser une plateforme logiciel libre en vérification automatique du locuteur. L'objectif était de pérenniser le savoir-faire du consortium ELISA, acquis grâce à des participations continues aux campagnes d'évaluations NIST depuis 1998, aux entreprises et laboratoires académiques qui souhaitaient se lancer dans la vérification automatique du locuteur. La contribution de GEOD sur ce projet a concerné la segmentation en locuteurs : participation à des campagnes d'évaluation (NIST, ESTER) et ré-écriture du système existant à GEOD en utilisant la plateforme ALIZE.

*Projet ACI-Santé DESDHIS (CLIPS, TIMC), 2002-04*

Dans le cadre d'une ACI Technologies pour la Santé du CNRS, l'équipe GEOD a mené, en coopération avec le laboratoire TIMC de l'UJF, le projet DESDHIS (*DEtection de Situations de Détresse dans un Habitat Intelligent pour la Santé*) dont l'objectif global était la conception, la mise au point et l'expérimentation d'un dispositif de télé-surveillance médicale s'appuyant sur l'utilisation de capteurs sonores, de capteurs de déambulation et d'activité. Cette étude visait à apporter une aide sécuritaire au maintien à domicile de personnes fragiles (personnes âgées ou en convalescence). Dans ce projet, la contribution de GEOD a porté essentiellement sur le traitement du signal sonore (détection et segmentation parole/non-parole en milieu bruité, reconnaissance de classes de sons, reconnaissance de parole) dans le but de reconnaître des événements sonores susceptibles de caractériser une situation de détresse de la personne sous télésurveillance.

*Projet ACI-SI BIOMUL (CLIPS, EURECOM, LIA, INT), 2003-06*

Ce projet est financé par le CNRS sur l'ACI *Sécurité Informatique* (BIOMUL : Biométrie et Multimodalités). Dans ce cadre, GEOD a pu accueillir un post-doctorant au CLIPS (Nicolas Eveno, issu du laboratoire LIS/INPG) pour travailler sur la biométrie labiale bimodale. Au cours de ce projet, la problématique de reconnaissance du locuteur dans des environnements perceptifs a également été abordée.

*BQR INPG 2003 “ Vidéo-Sémantique ” (CLIPS, LIS, LSR)*

Ce projet financé par l’INPG avait pour but de fédérer les collaborations entre trois laboratoires grenoblois dans le domaine de l’indexation multimédia. La contribution de GEOD, en collaboration avec l’équipe MRIM du CLIPS, a concerné le traitement automatique de la bande son de documents vidéo et la participation annuelle aux campagnes d’évaluation TREC-Vidéo.

*BQR INPG 2004-2005 “ TELMA ” (CLIPS, LIS, ICP)*

Ce projet financé par l’INPG vise à l’étude et au développement algorithmique de fonctionnalités audio-visuelles originales à l’usage des personnes malentendantes, et à l’étude de faisabilité de leur intégration dans un terminal autonome de télécommunication téléphonique. Le projet a pour objectif technique précis d’exploiter la modalité visuelle de la parole, d’une part pour améliorer les techniques de débruitage du son de parole (la minimisation du bruit environnemental permettant une meilleure exploitation des “restes” auditifs des malentendants), et d’autre part, en mettant en œuvre des techniques d’analyse/synthèse de lecture labiale et de gestes de la Langue Française Parlée Complétée (LPC).

*Projet ACI-SI BIOMUL (CLIPS, EURECOM, LIA, INT), 2003-06*

Ce projet est financé par le CNRS sur l’ACI *Sécurité Informatique* (BIOMUL : Biométrie et Multimodalités). Dans ce cadre, GEOD a pu accueillir un post-doctorant au CLIPS (Nicolas Eveno, issu du laboratoire LIS/INPG) pour travailler sur la biométrie labiale bimodale. Au cours de ce projet, la problématique de reconnaissance du locuteur dans des environnements perceptifs a également été abordée.

<b>Contexte Catégorie</b>	<b>Thème</b>	<b>Coordonnateur</b>	<b>Nb de partenai res</b>	<b>Collaborations spécifiques<sup>30</sup></b>	<b>Début Fin</b>	<b>Soutien financier propre</b>
CSTAR III	Traduction multilingue pour le renseignement touristique	ETRI (Corée)	14	CMU (USA) Eq. GETA-CLIPS	01/02 06/06	0
NESPOLE	Dialogue multilingue pour le e-commerce	ITC-Irst (Italy)	4	CMU (USA) Eq. GETA-CLIPS	01/00 12/02	58K€
EU COST Programme COST 275	Biometrics-Based Recognition Over The Internet	University of Herfordshire (UK)	14 Countries (2 Partners/ Country)	J.F. Bonastre (LIA-Avignon)	06/01 05/05	0
MAE CORUS	Traitement de la parole en langue vietnamienne	E. Castelli MICA-INP Hanoi (Vietnam)	2	E. Castelli MICA-INP Hanoi (Vietnam)	12/03 11/06	17,9K€
AUF TALK	Traitement automatique de la langue Khmère	E. Castelli MICA-INP Hanoi (Vietnam)	3	E. Castelli MICA-INP Hanoi (Vietnam)	02/04 01/06	2K€
RNRT Exploratoire PVE	Portail Vocal d’Entreprise	O. Causse NEUROSOFT Paris	4	Eq. Multicom-CLIPS	01/02 12/03	75K€
RNRT Exploratoire COUCOU	Conception Participative Orientée Usage	P. Mallein CNRS-MSH Grenoble	5	Eq. Multicom-CLIPS	01/02 12/04	20K€
RNTL Précompétitif ACE	Agent Conversationnel Expressif	D. Noël As AN ANGEL (Paris)	3	Eq. Multicom-CLIPS	07/03 12/04	33K€

<sup>30</sup> Noms des partenaires avec lesquels sont menées plus particulièrement des recherches conjointes

Technolangu AGILE	Traitement de l'oral- segmentation en locuteurs	I. Magrin- Chagnolleau Lab. DDL (Lyon)	6	J.F. Bonastre (LIA- Avignon)	01/03 12/04	21K€
ACI Santé DESDHIS	Télésurveillance sonore Habitat Intelligent Santé	J.F. Serignat CLIPS-IMAG Eq. GEOD	2	V. Rialle, N. Noury Lab. TIMC-IMAG Eq. AFIRM	08/02 08/04	41K€
ACI Sécurité BIOMUL	Biométrie et Multimodalités	J.F. Bonastre (LIA-Avignon)	4	J.F. Bonastre (LIA- Avignon)	07/03 06/06	32K€
IMAG DELOC	Collecticiel pour réunions délocalisées	L. Besacier CLIPS-IMAG Grenoble	2	Eq. MRIM-CLIPS	01/02 12/02	10K€
CNRS- LAFMI IPHM	Interact. parlée H-M en espagnol- mexicain	L. Villaseñor- Pineda INAOE-Mexico (Mexique)	2	L. Villaseñor- Pineda NAOE-Mexico (Mexique)	09/02 08/04	10K€
INPG-BQR Vidéo Sémantique	Segmentation- Indexation	G. Quenot MRIM-CLIPS	3	Eq. MRIM-CLIPS	01/03 12/03	2K€
INPG-BQR TELMA	Téléphonie à l'usage des malentendants	L. Girin ICP-INPG	3	L. Girin ICP-INPG	06/04 05/05	10K€
MAE France- Québec DIALOGU E	Modélisation Dialogue	J. Caelen CLIPS-IMAG	2	Université Trois Rivières-Québec (Canada)	01/02 12/02	1,8K€

### 4.3 Contrats (co)financés par un industriel

Trois contrats financés par un industriel (Thomson Multimédia, Prosodie et Thalès) ont été réalisés par GEOD dans la période considérée. La nature de ces trois contrats était sensiblement la même. Le projet de convention portait sur une expertise et un développement menés conjointement par GEOD et le partenaire industriel (Thomson, Prosodie ou Thalès) dans le but de réaliser un système de reconnaissance automatique de la parole et de le comparer à un système de reconnaissance vocale de référence. La contribution de GEOD consistait à accompagner le développement du système au niveau de la réalisation d'un état de l'art sur la reconnaissance du point de vue des méthodes et des algorithmes, de l'apport de données d'apprentissage exploitables pour la modélisation acoustique (étiquetage complet d'une base d'apprentissage), de conseils techniques dans le développement (choix des paramètres acoustiques, modèles acoustiques, recherche des hypothèses et modèle de langage) et du suivi scientifique dans le développement du système de reconnaissance en prenant comme point d'appui les résultats du système RAPHAEL de GEOD pour comparer et valider les résultats à toutes les étapes de la reconnaissance. Le système propriétaire réalisé par Prosodie, avec l'aide de GEOD, est actuellement en exploitation dans quelques serveurs vocaux interactifs de la société.

Partenaire(s)	Thème	Début Fin	Montant
PROSODIE	Reconnaissance de parole téléphonique	11/00 01/02	44K€
THOMSON	Reconnaissance de parole continue-Expertise	01/01 12/02	34K€
THALES	Reconnaissance de parole continue-modèle de langage	10/02 12/03	15K€

## 5. PUBLICATIONS (01-05)

### Articles dans des revues avec comité de lecture internationales et nationales (ACL)

#### Revues internationales :

Année 2001

1. A. Lecomte, "Partial Proof-Nets, Hybrid Logics and Minimalist Representations", *Grammars*, vol. vol. 3, n° 2-3, pp. 243-257, 2001.

Année 2003

2. L. Besacier, J. F. Bonastre, P. Mayorga-Ortiz, C. Fredouille and S. Meignier, "Overview of compression and packet loss effects in speech biometrics", *IEE Proceedings Vision, Image & Signal Processing - Special Issue on Biometrics on the Internet*, vol. 150, n°6, pp. 372-376, December, 2003.

Année 2004

3. L. Besacier, M. Ariyaeinia, J.S. Mason, J.F. Bonastre, P. Mayorga, C. Fredouille, S. Meignier, J. Siau, N.W.D. Evans, R. Auckenthaler, R. Stapert, "Voice Biometrics Over the Internet in the Framework of COST Action 275", *EURASIP Journal on Signal Processing*, vol. 4, pp. 466-479, April 2004.
4. A. Lecomte, "Rebuilding MP on a Logical Ground", *Research on Language and Computation, Special Issue on Resource Logics and Minimalist Grammars*, vol. 2, n°1, pp. 27-35.

Année 2005

5. C. Fredouille, D. Moraru, S. Meignier, J. F. Bonastre and L. Besacier, "Step-by-Step and Integrated Approaches in Broadcast News Speaker Diarization", *To appear in Computer Speech and Language Journal (Elsevier), 2005*.
6. D. Istrate, E. Castelli, M. Vacher, L. Besacier and J. F. Serignat, "Medical Telemonitoring System Based on Sound Detection and Classification", *To appear in IEEE Transactions on Information Technology in Biomedicine, 2005*.
7. A. Xuereb and J. Caelen, "Actes de langage et relations rhétoriques dans le dialogue homme-machine", in *Revue de l'Université de Moncton, Canada, sous presse : 2005*.

#### Revues nationales :

Année 2001

8. J. Rouillard, "Dialogue et multimodalité", *Revue RIHM*, vol. 2, n°1, pp. 97-125, 2001.

Année 2005

9. A. Xuereb and J. Caelen, "Speech Acts and Rhetorical Relations in Man-Machine Dialogue", in *Presses Universitaires de Nancy, à paraître : 2005*.

### Articles dans des revues sans comité de lecture (SCL)

Année 2001

1. E. Castelli and G. Nguyen Trong, "Le Centre de Recherche International MICA", *Bulletin franco-vietnamien du P.F.I.E.V.*, vol. n° 4, pp. 10-12, novembre 2001.

Année 2002

2. J. Caelen and L. Rodet, "Internauts votre regard vous trahit", in *Le Figaro*, Paris, 20 février 2002.
3. J. Caelen, "Les laboratoires d'usage", in *CNRS Info*, n° 402, spécial Innovation-Valorisation, juin 2002.
4. J. Caelen, "Le CNRS crée le premier laboratoire d'usage", *Revue des Techniques de l'Ingénieur*, vol. n° 27, Sept./Oct./Nov. 2002.
5. J. Caelen and L. Rodet, "Les mouvements oculaires de l'internaute utilisés pour juger les sites Web", in *Le Monde Sciences*, Paris, 22 novembre 2002.
6. J. Caelen and P. Mallein, "Article collectif sur les objets communicants", *01 Informatique*, n° 1701, 22 novembre 2002.
7. C. P. Nguyen and E. Castelli, "Utilisation du son pour la surveillance dans un appartement intelligent en médecine", in *Automation Today (revue vietnamienne)*, vol. 24, juin 2002, pp. 10-11 (article en vietnamien).

Année 2003

8. J. Caelen, "CLIPS place la dimension humaine au centre de ses recherches", in *Revue Présences*,

*Année 2004*

9. J. Caelen, “La plate-forme Multicom”, in *La lettre de l'INPG*, n° 04-2, juin 2004.
10. J. Caelen, “La plate-forme Multicom, pour inventer les usages des objets communicants du futur”, in *Papyrus, le magazine de l'Université Joseph Fourier*, octobre 2004.

<b>Conférences invitées (INV)</b>
-----------------------------------

**Conférences d'audience internationale**

*Année 2002*

1. L. Besacier, R. Lamy, P. Mayorga-Ortiz and D. Vaufreydaz, “Remote Recognition: speech coding and packet loss problems”, Guilin (China), Ed.: C-STAR III Workshop, March 2002.
2. J. Caelen, “Stratégies de dialogue”, in *Colloque International Logique et Dialogue*, Grenoble, octobre 2002.
3. D. Vaufreydaz, “Tutorial at ESSLLI'02 (European Summer School in Logic, Language and Information)”, in *Recent Advances in Speech Translation Systems Workshop: ASR and Scalability*, Trento (Italy), August 2002.

*Année 2004*

4. A. Lecomte, “Derivations as Proofs”, in *Categorical Grammar CG'04*, Montpellier, 7-11 juin 2004.

**Conférences d'audience nationale**

*Année 2001*

5. J. Caelen, “L'ergonomie des documents électroniques : catalogues en ligne, documentation technique, e-learning”, Conférence France Telecom R&D, ZIRST, Meylan (France), octobre 2001.
6. J. Caelen, “Dialogue et Internet”, Conférence à l'Université du Québec, Trois-Rivières, Canada, octobre 2001.
7. J. Caelen, “Dialogue adaptatif”, Conférence à l'Université Laval, Québec, Canada, octobre 2001.
8. D. Vaufreydaz, “Méthodes automatiques de calcul et de repondération de modèles de langages statistiques par l'utilisation d'Internet”, *Laboratoire d'Informatique d'Avignon (LIA)*, Avignon, mai 2001.

*Année 2002*

9. S. Hollard, “Analyse de documents par oculométrie”, in *Journée Scientifique du Pôle Rhône-Alpes de Sciences Cognitives*, Archamps (France), 8 mars 2002.
10. D. Vaufreydaz, “Modélisation statistique du langage à partir d'Internet pour la reconnaissance automatique de la parole continue”, in *Laboratoire LORIA*, Nancy, Février 2002.
11. D. Vaufreydaz, “Reconnaissance de la parole spontanée continue au laboratoire CLIPS”, in *Laboratoire GRAVIR (Graphisme, Vision et Robotique)*, Montbonnot (France), Juin 2002.
12. D. Vaufreydaz, “Reconnaissance automatique de la parole grand vocabulaire”, in *Laboratoire MICA (Multimedia Information, Communication and Applications)*, Hanoi (Vietnam), Septembre 2002.
13. D. Vaufreydaz, “Modélisation statistique du langage à partir d'Internet pour la reconnaissance automatique de la parole continue”, in *Laboratoire MICA (Multimedia Information, Communication and Applications)*, Hanoi (Vietnam), Septembre 2002.

*Année 2003*

14. J. Caelen, “Modèle de dialogue générique et portail vocal”, in *Séminaire “Expression et communication en interface homme-machine”*, Paris, Ed.: Université Paris VIII, 11 mars 2003.
15. J. Caelen, “Dialogue homme-machine générique”, in *Séminaire du LIMSIS*, Orsay, 24 avril 2003.
16. Lecomte and C. Retoré, “Sémantique dans les grammaires minimalistes”, in *Journées Sémantique et Modélisation, GDR Sémantique CNRS*, Paris, 2003.

*Année 2004*

17. J. Caelen, “La conception participative : enjeux et méthodes”, in *2ème Conférence Internationale des Chercheurs Vietnamiens et Francophones en Informatique*, Hanoi, Vietnam, février 2004.
18. J. Caelen, “Conception participative”, in *Séminaire du LIMSIS*, Orsay, France, 4 mai 2004.
19. J. Caelen, “Evaluation des sites Internet par oculométrie”, in *Atelier CIDE*, La Rochelle, juin 2004.
20. J. Caelen, “Ergonomie des agents conversationnels expressifs”, in *Séminaire LIP6*, Paris, juin 2004.
21. J. Caelen, “Conception participative des objets interactifs : principes, méthodes et instrumentation”, in *Ecole d'été TIC et société, CNRS*, Carry le Rouet, 2004.
22. J. Caelen, “Modèles de dialogue pour l'interaction”, in *16ème Conférence francophone sur l'Interaction*

*Homme-Machine (IHM'04)*, Namur, Belgique, 30 août-3sept. 2004.

23. A. Lecomte, “*Pré-groupes et minimalisme*”, presented at Séminaire du LABRI, Lab. LABRI, Bordeaux, 18 février 2004,
24. A. Lecomte, “*Vers des Assistants-LEcteurs INtelligents (VALEIN)*”, in *Journée Programmes Interdisciplinaires PI-TCAN*, Centre National de la Recherche Scientifique (CNRS), Paris, 13 octobre 2004.

## Communications avec actes internationales et nationales (ACT)

### Conférences d'audience internationale

#### Année 2001

1. L. Besacier, C. Bergamini, D. Vaufréydaz and E. Castelli, “*The Effect of Speech and Audio Compression on Speech Recognition Performance*”, presented at MMSP 2001, IEEE Multimedia Signal Processing Workshop, Cannes, October 3-5, 2001, pp. 301-306.
2. L. Besacier, H. Blanchon, Y. Fouquet, J. P. Guilbaud, S. Helme, S. Mazenot, D. Moraru and D. Vaufréydaz, “*Speech Translation for French in the NESPOLE! European Project*”, presented at Eurospeech 2001, Aalborg, Danemark, September 3-7, 2001, pp. 1291-1294.
3. S. Burger, L. Besacier, P. Coletti, F. Metzger and C. Morel, “*The NESPOLE! VoIP Dialogue Database*”, presented at Eurospeech 2001, Aalborg, Danemark, September 3-7, 2001, pp. 2043-2046.
4. E. Castelli and D. Istrate, “*Multichannel Audio Acquisition for Medical Supervision in an Intelligent Habitat*”, presented at ECCTD 2001, 15th European Conference on Circuit Theory and Design, Helsinki, Finland, August 28-31, 2001, pp. II-1 II-4.
5. D. Istrate and E. Castelli, “*Multichannel Sound Acquisition with Stress Situations Determination for Medical Supervision in a Smart House*”, in *Text, Speech and Dialogue*, P. Faculty of Applied Sciences, Ed., Zelezna Ruda, Czech Republic: Springer-Verlag, september, 2001, pp. 266-272.
6. D. Istrate and E. Castelli, “*Everyday Life Sounds and Speech Analysis for a Medical Telemonitoring System*”, presented at Eurospeech 2001, Aalborg, Danemark, September 3-7, 2001, pp. E15 2417.
7. M. Z. Kurdi, “*A spoken language understanding approach which combines the parsing robustness with the interpretation deepness*”, presented at IC-AI01, International Conference on Artificial Intelligence 2001, Las Vegas, USA, June 25-28, 2001, pp. 1143-1149.
8. A. Lecomte and C. Retoré, “*Extending Lambek Grammars : a logical account of Minimalist Grammars*”, presented at 39th Annual Meeting of the Association for Computational Linguistics (ACL), Toulouse, France, July 9-11, 2001, pp. 354-362.
9. A. Lecomte, “*Semantic Representations in Logic for Minimalism*”, presented at Workshop on Logic and Language (JoLL'2001), Seville, Spain, Nov. 29-Dec.1, 2001, pp. 155-162.
10. Q. C. Nguyen, N. Y. Pham-Thi and E. Castelli, “*Shape vector characterization of Vietnamese tones and application to automatic recognition*”, presented at ASRU2001 (Automatic Speech Recognition and Understanding Workshop), Madonna di Campiglio, Trento, Italy, December 9-13, 2001, pp. 4 (on CDROM).
11. D. Vaufréydaz, L. Besacier, C. Bergamini and R. Lamy, “*>From generic to task-oriented speech recognition: French experience in the NESPOLE! European project*”, presented at ITRW Workshop on Adaptation Methods for Speech Recognition, Sophia-Antipolis, France, August, 29-30, 2001.
12. D. Vaufréydaz and M. Gery, “*Internet Evolution and Progress in Full Automatic French Language Modelling*”, presented at ASRU2001 (Automatic Speech Recognition and Understanding Workshop), Madonna di Campiglio, Trento, Italy, December 9-11, 2001, pp. 4 (on CDROM).

#### Année 2002

13. J. Y. Antoine, C. Bousquet-Vernhettes, J. Goulian, M. Z. Kurdi, S. Rosset, N. Vigouroux and J. Villaneau, “*Predictive and objective evaluation of speech understanding: the "challenge" evaluation campaign of the I3 speech workgroup of the French CNRS*”, presented at Third International Conference on Language Resources and Evaluation (LREC'02), Las Palmas (Spain), May 29 – 31, 2002, pp. 529 - 536.
14. L. Besacier, P. Mayorga-Ortiz, J. F. Bonastre and C. Fredouille, “*Methodology for Evaluating Speaker Verification Robustness over IP Networks*”, presented at COST275 Workshop on The Advent of Biometrics on the Internet, Rome, Italy, November 7-8, 2002.
15. M. Gery, M. H. Haddad and D. Vaufréydaz, “*Web as Huge Information Source for Noun Phrases Integration in the Information Retrieval Process*”, presented at International Conference on Information and Knowledge Engineering (IKE'02), Las Vegas - Nevada (USA), June 2002, pp. 72-77.
16. M. H. Haddad, M. Gery and D. Vaufréydaz, “*Le Web : une source d'information pour l'intégration de multi-termes dans un processus de Recherche d'Information*”, presented at Journées Francophones d'Accès Intelligent aux Documents Multimédias sur l'Internet (MediaNet 2002), Sousse (Tunisia), June 2002, pp. 257-268.
17. M. Z. Kurdi, “*Combining pattern matching and shallow parsing techniques for detecting and*

- correcting spoken language extragrammaticalities*”, presented at 2nd Workshop on ROBust Methods in Analysis of Natural language Data (ROMAND 2002), Frascati-Rome (Italy), July 17, 2002, pp. 1-9.
18. M. Z. Kurdi and M. Ahafhaf, “*Toward an objective and generic Method for Spoken Language Understanding Systems Evaluation: an extension of the DCR method*”, presented at Third International Conference on Language Ressources and Evaluation (LREC'02), Las Palmas (Spain), May 29-31, pp. 545-550.
  19. A. Lecomte, “*A minimal logic for minimalism*”, presented at 3rd and the 4th International Symposium on Language, Logic and Computation, Amsterdam (The Netherlands), 2002.
  20. A. Lecomte, “*Bi-grammars for the interface between syntax and semantics*”, presented at Formal Grammar 2002 Conference, Trento (Italy), August 2002, pp. 67-79.
  21. P. Mayorga-Ortiz, R. Lamy and L. Besacier, “*Recovering of packet loss for distributed speech recognition*”, presented at Eusipco 2002, Toulouse, septembre 2002.
  22. G. M. Quenot, D. Moraru, L. Besacier and P. Mulhem, “*CLIPS-IMAG at TREC-11 : Experiments in Video Retrieval*”, presented at 11th Text Retrieval Conference, Gaithersburg, MD, USA, November 19-22, 2002.
  23. S. Rossato, H. Blanchon and L. Besacier, “*Speech-to-speech translation system evaluation : results for French for the Nespole! Project first showcase*”, presented at ICSLP2002, Denver, USA, September 2002.
  24. L. Villaseñor-Pineda, M. Montes-y-Gomez, M. Perez-Coutino and D. Vaufreydaz, “*Comparación léxica de corpus para generación de modelos de lenguaje*”, presented at IBERAMIA workshop on Multilingual Information Access and Natural Language Processing, Sevilla (Spain), November 2002, pp. 6 pages.

#### Année 2003

25. B. Bigi, “*Using Kullback-Leibler Distance for Text Categorization*”, presented at 25th European Conference on Information Retrieval Research, Pisa (Italy), 2003, pp. 305-319.
26. J. Caelen, “*Strategies of Dialogue*”, presented at Speech Technology and Human-Computer Dialogue Conference, Bucarest, 10-11 avril 2003, pp. 27-42.
27. E. Castelli, M. Vacher, D. Istrate, L. Besacier and J. F. Serignat, “*Habitat Telemonitoring System Based on the Sound Surveillance*”, presented at ICICTH (International Conference on Information Communication Technologies in Health), Samos Island, Greece, July 13-15, 2003, pp. 141-146.
28. D. Istrate, G. Virone, M. Vacher, E. Castelli and J. F. Serignat, “*Communication between a Multichannel Audio Acquisition and an Information System in a Health Smart Home for Data Fusion*”, presented at IMSA 2003 : 7th IASTED International Conference on Internet and Multimedia Systems and Applications, Honolulu, Hawaii, August 13-15, 2003.
29. M. Z. Kurdi and M. Ahafhaf, “*A Grammar Based Method for Systematic and Generic Spoken Language Understanding Systems Evaluation*”, presented at NLPKE 2003 : International Conference on Natural Language Processing and Knowledge Engineering, Beijing Media Center (China), October 26-29, pp. 498-505.
30. R. Lamy and L. Besacier, “*Non-Linear Acoustical Pre-Processing for Multiple Sampling Rates ASR and ASR in Noisy Condition*”, presented at NOLISP'03 (ISCA tutorial and research workshop on non-linear speech processing), Le Croisic (France), May 20-23, 2003, pp. 57-60.
31. V. B. Le, B. Bigi, L. Besacier and E. Castelli, “*Using the Web for Fast Language Model Construction in Minority Languages*”, presented at Eurospeech'2003, Geneva, Switzerland, September 1-4, 2003.
32. A. Lecomte, “*Y a-t-il une logique de la classification ?*”, presented at Congrès de ISKO-France (International Society of Knowledge Organisation), Grenoble, juillet 2003.
33. A. Lecomte, “*A Computational Approach to Minimalism*”, presented at ICON'2003 (International Conference on Natural Language), Mysore, India, December 19-21, 2003.
34. H. Le-Xuan, L. Trinh-Van and E. Castelli, “*Automatic Synthesis of Vietnamese by concatenation of Diphones*”, presented at ICT-rda'03 Conference (Information Technology & Communication, Research, Development and Applications), Hanoi (Vietnam), February 22-23, 2003, pp. 317-322 (article en vietnamien).
35. N. Mana, S. Burger, R. Cattoni, L. Besacier, V. Mac Laren, J. McDonough and F. Metze, “*The NESPOLE! VoIP Multilingual Corpora in Tourism and Medical Domains*”, presented at Eurospeech'2003, Geneva, Switzerland, September 1-4, 2003.
36. P. Mayorga-Ortiz, L. Besacier, L. Lamy and J. F. Serignat, “*Audio Packet Loss over IP and Speech Recognition*”, presented at ASRU 2003 (Automatic Speech Recognition & Understanding), Virgin Islands, USA, Nov. 1-Dec. 4, 2003, pp. 607-612.
37. P. Mayorga-Ortiz, L. Besacier and R. Villa, “*Perdida de datos en conexiones IP y su influencia en reconocimiento de voz*”, presented at ELECTRO 2003 (XXV Congreso Internacional de Ingeniería Electronica), Chihuahua, Chi. (Mexico), Octubre 27-31, 2003.
38. D. Moraru and L. Besacier, “*Towards Conversational Model for Speaker Segmentation*”, presented at Speech Technology & Human-Computer Dialogue, Bucharest, Roumanie, April 10-11, 2003, pp. 69-78.

39. D. Moraru, S. Meignier, L. Besacier, J. F. Bonastre and I. Magrin-Chagnolleau, “*The ELISA Consortium Approaches in Speaker Segmentation during the NIST 2002 Speaker Recognition Evaluation*”, presented at ICASSP'2003, Hong-Kong, China, May 2003.
40. H. Nguyen and J. Caelen, “*Generic manager for spoken dialogue systems*”, presented at DIABRUCK 2003 : 7th Workshop on the Semantics and Pragmatics of Dialogue, Saarbrucken (Deutschland), September 4-6, 2003, pp. 201-202.
41. P. B. Nguyen, L. Trinh-Van and E. Castelli, “*Real Time System for Isolated Vietnamese Digits Recognition*”, presented at ICT-rda'03 Conference (Information Technology & Communication, Research, Development and Applications), Hanoi (Vietnam), February 22-23, 2003, pp. 310-316 (article en vietnamien).
42. M. Vacher, D. Istrate, L. Besacier, E. Castelli and J. F. Serignat, “*Smart Audio Sensor for Telemedicine*”, presented at Smart Objects Conference (SOC) 2003, Grenoble, 15-17 Mai, 2003, pp. 222-225.
43. M. Vacher, D. Istrate, L. Besacier, J. F. Serignat and E. Castelli, “*Life Sounds Extraction and Classification in Noisy Environment*”, presented at SIP 2003 : 5th IASTED International Conference on Signal and Image Processing, Honolulu, Hawaii (USA), August 13-15, 2003.
44. L. Villaseñor-Pineda, M. Montes and J. Caelen, “*Un Modelo para la Conversación Multimodal Hombre-Máquina: Integración del Habla y la Acción*”, presented at Congreso Latinoamericano de Interacción Humano-Computadora, Rio de Janeiro (Brasil), August 17-20, 2003.
45. L. Villaseñor-Pineda, M. Montes-y-Gomez, M. Perez-Coutino and D. Vaufreydaz, “*A Corpus Balancing Method for Language Model Construction*”, presented at Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2003), Mexico City (Mexico), February 16-22, 2003, pp. 393-401.
46. L. Villaseñor-Pineda, M. Montes-y-Gomez, D. Vaufreydaz and J. F. Serignat, “*Elaboracion de un Corpus Balanceado para el Calculo de Modelos Acusticos usando la Web*”, presented at XII Congreso Internacional de Computacion CIC'03, Mexico City (Mexico), October 13-17, 2003.
47. G. Virone, D. Istrate, M. Vacher, J. F. Serignat, N. Noury and J. Demongeot, “*First Steps in Data Fusion between a Multichannel Audio Acquisition and an Information System for Home Healthcare*”, presented at 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun (Mexico), September 17-21, 2003, pp. 1364-1367.

#### Année 2004

48. L. Besacier, G. Quenot, S. Ayache and D. Moraru, “*Video Story Segmentation with Multi-Modal Features: Experiments on TRECvid 2003*”, presented at 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, New York, USA, October 15-16, 2004.
49. B. Bigi, Y. Huang and R. De Mori, “*Vocabulary and Language Model Adaptation using Information Retrieval*”, presented at International Conference on Spoken Language Processing (ICSLP 2004), Jeju Island, Korea, October 4-8, 2004, 2, pp. 1361-1364.
50. C. Fredouille, D. Moraru, S. Meignier, L. Besacier and J. F. Bonastre, “*The NIST 2004 Spring Rich Transcription Evaluation : Two Axis Merging Strategy in the Context of Multiple Distance Microphone Based Meeting Speaker Segmentation*”, presented at RT2004 Spring Meeting Recognition Workshop, Montreal, Canada, May 17, 2004.
51. D. Istrate, M. Vacher, E. Castelli and C. P. Nguyen, “*Sound Processing for Health Smart Home*”, presented at 2nd International Conference on Smart Homes and Health Telematic, Singapore, September 15-17, 2004.
52. D. Istrate, M. Vacher and J. F. Serignat, “*Multichannel Smart Sound Sensor for Perceptive Spaces*”, presented at Complex Systems Intelligence and Modern Technology Applications (CSIMTA 2004), Cherbourg, France, September 19-22, 2004, pp. 691-696.
53. V. B. Le, D. D. Tran, E. Castelli, L. Besacier and J. F. Serignat, “*Spoken and Written Language Ressources for Vietnamese*”, presented at Fourth International Conference on Language Ressources and Evaluation (LREC 2004), Lisbonne, Portugal, May 26th-28th, 2004, 2/5, pp. 599-602.
54. S. Meignier, D. Moraru, C. Fredouille, L. Besacier and J. F. Bonastre, “*Benefit of Prior Acoustic Segmentation for Speaker Segmentation Systems*”, presented at International Conference on Acoustics Speech & Signal Processing (ICASSP 2004), Montreal, Canada, May, 2004.
55. D. Moraru, S. Meignier, C. Fredouille, L. Besacier and J. F. Bonastre, “*The ELISA Consortium Approaches in Broadcast News Speaker Segmentation during the NIST 2003 Rich Transcription Evaluation*”, presented at International Conference on Acoustics Speech & Signal Processing (ICASSP 2004), Montreal, Canada, May, 2004.
56. D. Moraru, L. Besacier and E. Castelli, “*Using a priori information for speaker diarization*”, presented at The Speaker and Language Recognition Workshop, Toledo, Spain, May 31-June 4th, 2004.
57. D. Moraru, S. Meignier, C. Fredouille, L. Besacier and J. F. Bonastre, “*ELISA NIST RT03 Broadcast News Speaker Diarization Experiments*”, presented at The Speaker and Language Recognition Workshop, Toledo, Spain, May 31-June 4th, 2004.

58. D. Moraru, L. Besacier, S. Meignier, C. Fredouille and J. F. Bonastre, “*Speaker Diarization in the ELISA Consortium over the past 4 Years*”, presented at RT2004 Fall Workshop, Montreal, Canada, November, 2004.
59. H. Nguyen and J. Caelen, “*Multi-Session Management in Spoken Dialogue System*”, presented at IBERAMIA 2004, Puebla, Mexico, Novembre 22-26, 2004.
60. M. Vacher, D. Istrate, L. Besacier and J. F. Serignat, “*Sound Detection and Classification for Medical Telesurvey*”, presented at 2nd Conference on Biomedical Engineering (BIOMED 2004), Innsbruck, Austria, February 16-18, 2004, pp. 395-398.
61. M. Vacher, D. Istrate and J. F. Serignat, “*Sound Detection and Classification through Transient Models using Wavelet Coefficient Trees*”, presented at 12th European Signal Processing Conference (EUSIPCO 2004), Vienna, Austria, September 6-10, 2004, pp. 1171-1174.
62. M. Vacher, D. Istrate and J. F. Serignat, “*Sound Detection through Transient Models using Wavelet Coefficient Trees*”, presented at Complex Systems Intelligence and Modern Technology Applications (CSIMTA 2004), Cherbourg, France, September 19-22, 2004, pp. 367-372.
63. L. Villaseñor-Pineda, M. Montes-y-Gómez, D. Vaufreydaz and J. F. Serignat, “*Experiments on the Construction of a Phonetically Balanced Corpus from the Web*”, presented at Conference on Intelligent Text Processing and Computational Linguistics CICLing-2004, Lecture Notes in Computer Science, Springer-Verlag, Seoul (Korea), February 2004, vol. 2945, 4.
64. Q. Vu-Minh, L. Besacier, E. Castelli, B. Bigi and H. Blanchon, “*Interchange format-based language model for automatic speech recognition in speech-to-speech translation*”, presented at RIVF'04 (Recherche Informatique Vietnam-Francophonie), Hanoi, Vietnam, February 2-5, 2004, 1/1, pp. 47-50.

#### Année 2005

65. N. Eveno and L. Besacier, “*A Speaker Independent "Liveness" Test for Audio-Visual Biometrics*”, presented at 9th European Conference on Speech Communication and Technology (INTERSPEECH 2005), Lisbon (Portugal), September 4-8, 2005, pp. 3081-3084.
66. D. Istrate, M. Vacher and J. F. Serignat, “*Détection et classification des sons : application aux sons de la vie courante et à la parole*”, presented at GRETSI Traitement du signal et des images, Louvain-la-Neuve, Belgique, 6-9 septembre 2005, pp. 485-488.
67. V. B. Le, D. D. Tran, L. Besacier, E. Castelli and J. F. Serignat, “*First Steps in Building a Large Vocabulary Continuous Speech Recognition System for Vietnamese*”, presented at 3ème Conférence Internationale en Informatique - Recherche, Innovation & Vision du Futur (RIVF2005), Université de Can Tho (Vietnam), 21-24 février 2005, pp. 330-333.
68. V. B. Le and L. Besacier, “*First Steps in Fast Acoustic Modeling for a New Target Language: Application to Vietnamese*”, presented at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'05), Philadelphia (USA), March 19-23, 2005, Vol. 1, pp. 821-824.
69. A. Lecomte and A. Naït-Abdallah, “*On Expressing Vague Quantification and Scalar Implicatures in the Logic of Partial Information*”, presented at Workshop Logical Aspects of Computational Linguistics (LACL'05), Bordeaux, April 2005.
70. B. Senechal, D. Pellerin, L. Besacier, I. Simand and S. Brès, “*Audio, Video and Audio-Visual Signatures for short Video Clip Detection: Experiments on Trecvid2003*”, presented at IEEE International Conference on Multimedia and Expo (ICME'05), Amsterdam (The Netherlands), July 6-8, 2005.
71. M. Tomokiyo, G. Chollet, S. Hollard, “*Semantic Representations of Emotions Considering Speech Acts of Utterances in Oral Dialogues*”, presented at Pacific Association for Computational Linguistics (PACLING2005), Tokyo, Japan, August 24-27, 2005.
72. D. D. Tran, E. Castelli, J. F. Serignat, V. L. Trinh and X. H. Le, “*Influence of F0 on Vietnamese Syllable Perception*”, presented at 9th European Conference on Speech Communication and Technology (INTERSPEECH 2005), Lisbon (Portugal), September 4-8, 2005, pp. 1697-1700.
73. M. Vacher, D. Istrate, J. F. Serignat and N. Gac, “*Detection and Speech/Sound Segmentation in a Smart Room Environment*”, presented at The 3rd Conference on Speech Technology and Human-Computer Dialogue (SpeD 2005), Cluj-Napoca (Romania), May 13-14, 2005, pp. 37-48.
74. L. Villaseñor-Pineda, V. B. Le, M. Montes-y-Gomez and M. Perez-Coutino, “*Toward Acoustic Models for Languages with Limited Linguistic Resources*”, presented at 6th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005), Mexico City (Mexico), February 13-19, 2005, Vol. 3406, pp. 433-436.
75. Q. Vu-Minh, L. Besacier, E. Castelli and N. Y. Pham, “*A Decision Tree-Based Method for Speech Classification into Question and NonQuestion*”, presented at 3ème Conférence Internationale en Informatique - Recherche, Innovation & Vision du Futur (RIVF2005), Université de Can Tho (Vietnam), 21-24 février 2005, pp. 306-309.

#### Conférences d'audience nationale

##### Année 2001

76. E. Castelli, D. Istrate, V. Rialle and N. Noury, “*Information Extraction from Speech in Stress Situation*”.

*Application to the Medical Supervision in a Smart House.*”, presented at ORAGE 2001, Oralité et Gestualité - Interactions et comportements multimodaux dans la communication., Aix-en-Provence, 18-22 juin 2001, pp. 362-371.

#### Année 2002

77. J. Y. Antoine, S. Letellier-Zarshenas, P. Nicolas, I. Schadle and J. Caelen, “*Corpus OTG et ECOLE MASSY : vers la constitution d'une collection de corpus francophones de dialogue oral diffusée librement*”, presented at 9ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN'2002), Nancy, France, 24-27 juin, 2002, pp. 6 pages.
78. Y. Fouquet, “*Un modèle de dialogue par les attentes du locuteur*”, presented at 9ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN'2002), Nancy, France, 24-27 juin 2002, pp. 371-376.
79. R. Lamy and L. Besacier, “*Adaptation spectrale par quantification vectorielle : exemple de la RAP à fréquences d'échantillonnage multiples*”, presented at XXIVèmes Journées d'Etude sur la Parole (JEP'2002), Nancy, 24-27 Juin 2002, pp. 321-324.
80. D. Moraru and L. Besacier, “*Segmentation en locuteurs de conversations sur IP*”, presented at XXIVèmes Journées d'Etude sur la Parole (JEP'2002), Nancy, Juin 2002.
81. Q. C. Nguyen and E. Castelli, “*Caractérisation et reconnaissance automatique des tons du vietnamien*”, presented at RFIA 2002 (Reconnaissance des Formes et Intelligence Artificielle), Angers, 8-10 janvier 2002.
82. Y. Pham-Thi-Ngoc, E. Castelli and Q. C. Nguyen, “*Gabarits des tons vietnamiens*”, presented at XXIVèmes Journées d'Etude sur la Parole (JEP'2002), Nancy, Juin 2002, pp. 23-26.
83. S. Rossato, H. Blanchon and L. Besacier, “*Évaluation du premier démonstrateur de traduction de parole dans le cadre du projet NESPOLE!*”, presented at Congrès TALN (Traitement Automatique du Langage Naturel), Nancy, Juin 2002.

#### Année 2003

84. J. Caelen, “*Stratégies de dialogue*”, presented at Conférence MFI'03 (Modèles Formels de l'Interaction), Lille, mai 2003.
85. J. Caelen, V. Eglin, S. Hollard and B. Meillon, “*Mouvements oculaires et évaluation de documents électroniques*”, presented at CIDE-6, 6ème Colloque sur le Document Electronique, Caen, 24-26 novembre 2003, pp. 107-135.
86. A. Nait-Abdallah and A. Lecomte, “*Un modèle de raisonnement avec propositions implicites*”, presented at Journées sur les Nouveaux Modèles de Raisonnement, Paris, 27-28 novembre 2003.
87. H. Nguyen, “*Vers une architecture générique de système de dialogue oral homme-machine*”, presented at RECITAL 2003 (Rencontre des Etudiants Chercheurs en Informatique pour le Traitement Automatique des Langues), Batz sur Mer, 11-14 juin 2003, pp. 539-545.

#### Année 2004

88. H. Blanchon and L. Besacier, “*Traduction de dialogue : résultat du projet NESPOLE! et pistes pour le domaine*”, presented at XIème Conférence sur le Traitement Automatique du Langage Naturel (TALN'04), Session Poster, Fès, Maroc, 19-21 avril 2004.
89. J. Caelen and H. Nguyen, “*Gestion de buts de dialogue*”, presented at XIème Conférence sur le Traitement Automatique du Langage Naturel (TALN'04), Fès, Maroc, 19-21 avril 2004, pp. 345-350.
90. J. Caelen and F. Jambon, “*Conception participative par "moments"*”, presented at 16ème Conférence Francophone sur l'Interaction Homme-Machine (IHM'04), Namur, Belgique, 30 août-3 sept. 2004, pp. 29-36.
91. Y. Fouquet, “*Prédiction d'actes et attentes en dialogue : expérience avec un assistant virtuel simulé*”, presented at XIème Conférence sur le Traitement Automatique du Langage Naturel (TALN'04), Fès, Maroc, 19-21 avril 2004, pp. 193-198.
92. R. Lamy, D. Moraru, B. Bigi and L. Besacier, “*Premiers pas du CLIPS sur les données d'évaluation ESTER*”, presented at 25èmes Journées d'Etude sur la Parole (JEP 2004), Fès, Maroc, 19-22 avril 2004, pp. 301-304.
93. A. Lecomte and M. Amblard, “*Syntax and Semantics Interacting in a Minimalist Theory*”, presented at Journée de Sémantique et Modélisation (JSM'04), ENS, Lyon, mars 2004.
94. D. Moraru, S. Meignier, C. Fredouille, L. Besacier and J. F. Bonastre, “*Segmentation selon le locuteur : les activités du Consortium ELISA dans le cadre de NIST RT03*”, presented at 25èmes Journées d'Etude sur la Parole (JEP 2004), Fès, Maroc, 19-22 avril 2004, pp. 389-392.
95. S. A. Selouani and J. Caelen, “*Un système hybride pour l'identification de traits phonétiques complexes de la langue arabe*”, presented at 25èmes Journées d'Etude sur la Parole (JEP 2004), Fès, Maroc, 19-22 avril 2004, pp. 449-452.
96. Q. Vu-Minh, L. Besacier, H. Blanchon and B. Bigi, “*Modèle de langage sémantique pour la reconnaissance automatique de parole dans un contexte de traduction*”, presented at XIème Conférence

sur le Traitement Automatique du Langage Naturel (TALN'04), Session Poster, Fès, Maroc, 19-21 avril 2004.

97. A. Xuereb and J. Caelen, “*Un modèle d'interprétation pragmatique en dialogue homme-machine basé sur la SDRT*”, presented at XIème Conférence sur le Traitement Automatique du Langage Naturel (TALN'04), Fès, Maroc, 19-21 avril 2004, pp. 505-514.

#### Année 2005

98. L. Besacier, V. B. Le, E. Castelli, S. Sethserey and L. Protin, “*Reconnaissance automatique de la parole pour les langues peu dotées : application au Vietnamien et au Khmer*”, presented at TALN 2005 : 12ème Conférence Annuelle sur le Traitement Automatique des Langues Naturelles, Workshop Traitement des Langues peu dotées, Dourdan (France), 6-10 juin 2005, Vol. 2, pp. 207-217.
99. J. Caelen and H. Nguyen, “*Vers un dialogue collectif avec la machine*”, presented at Troisièmes Journées Francophones Modèles Formels de l'Interaction (MFI'05), Caen (France), 25-27 mai 2005, pp. 211-216.
100. Y. Fouquet and J. Caelen, “*Deux plate-formes pour l'expérimentation d'un agent conversationnel animé*”, presented at WACA'01 : Premier Workshop francophone sur Agents Conversationnels Animés, Grenoble, 13-14 juin 2005.
101. A. Lecomte and A. Naït-Abdallah, “*Scalar Implications and Logic of Partial Information*”, presented at Journée de Sémantique et Modélisation (JSM'05), Paris, 18 mars 2005.
102. D. Moraru, L. Besacier, G. Quenot and S. Ayache, “*Speaker and Story Segmentation Using Audio-Video Information*”, presented at The 3rd Conference on Speech Technology and Human-Computer Dialogue (SpeD 2005), Cluj-Napoca (Romania), May 13-14, 2005.
103. A. Xuereb and J. Caelen, “*Topiques Dialogiques*”, presented at TALN 2005 : 12ème Conférence Annuelle sur le Traitement Automatique des Langues Naturelles, Dourdan (France), 6-10 juin 2005, Vol. 1, pp. 273-282.

### Communications sans actes (COM)

#### Année 2001

1. Y. Fouquet, “*De l'étude de dialogues oraux unilingues dans une langue non maternelle*”, in *Quatrièmes Rencontres Jeunes Chercheurs en Parole (RJC)*, Mons, Belgique, Ed.: Groupe Francophone de la Communication Parlée (GFCP), 11-14 septembre 2001.
2. D. Istrate and E. Castelli, “*Multichannel Sound Acquisition with Stress Situations Determination for Medical Supervision in a Smart House*”, in *TSD 2001, 4th International Conference Text, Speech and Dialogue*, Zelezná Ruda, Czech Republic, September 11-13, 2001.
3. D. Istrate and Q.-C. Nguyen, “*Room Echo Cancellation for Speech Recognition*”, in *ICASSP 2001, Student Forum*, Salt Lake City, Utah, USA, May 6-11, 2001.
4. R. Lamy, “*Adaptation de modèles acoustiques pour la reconnaissance automatique de la parole*”, in *Quatrièmes Rencontres Jeunes Chercheurs en Parole (RJC)*, Mons, Belgique, Ed.: Groupe Francophone de la Communication Parlée (GFCP), 11-14 septembre 2001.
5. Q.-C. Nguyen, D. Istrate and J. Barton, “*Blind Source Separation*”, in *ICASSP 2001, Student Forum*, Salt Lake City, Utah, USA, May 6-11, 2001.

#### Année 2003

6. Y. Fouquet, “*Le magicien d'Oz pour du dialogue oral : expérience avec un assistant virtuel en entreprise*”, presented at Cinquièmes Rencontres Jeunes Chercheurs en Parole, Grenoble, 23-26 septembre 2003.

#### Année 2004

7. M. Tomokiyo, S. Hollard, “*A Semantic Representation of Emotions Based on a Dialogue Corpus Analysis*” Papillon Workshop, IMAG Institute, Grenoble, August 30, 2004.
8. A. Xuereb and J. Caelen, “*L'interprétation pragmatique en dialogue homme-machine finalisé*”, presented at Journées JSM'04, Lyon, mars 2004.

#### Année 2005

9. L. Besacier, B. Bigi, D. Moraru, R. Lamy, “*Participation du CLIPS aux Evaluations ESTER : Tâches TRS et SRL*”, Atelier ESTER, Avignon, 30-31 mars 2005.
10. S. Hollard, M. Tomokiyo, D. Tuffelli, “*Une approche de l'expression orale des émotions. Etude d'un corpus réel*”, Communication affichée, 4<sup>èmes</sup> journées de linguistique de corpus, Lorient, France, 15-17 septembre 2005.
11. M. Tomokiyo, G. Chollet, S. Hollard, “*VoiceUNL : A Semantic Representation of Emotions Based on a Multimodal Dialogue Corpus Analysis*”. in *Un siècle de Phonétique expérimentale : histoire et*

## Ouvrages scientifiques (ou chapitres) (OS)

### Année 2001

1. A. Lecomte, "Categorial Minimalism", in *Logical Aspects of Computational Linguistics*, vol. LNAI n° 2014, M. Moortgat, Ed.: Springer Verlag, 2001, pp. 143-158.

### Année 2002

2. J. Caelen, "Modèles formels de dialogue", in *Actes des 2èmes assises du GdR I3, Information, Interaction Intelligence*, J. L. Maître, Ed.: Cépadués Editions, Toulouse, 2002, pp. 31-58.

### Année 2003

3. M. Amblard, A. Lecomte and C. Retoré, "Syntax and Semantics interacting in a minimalist theory", in *Prospects and Advances in the theory of Syntax Semantics Interface*, D. Duchier, Ed., Nancy, 2003, pp. 17-22.
4. L. Besacier and J. F. Bonastre, "Subband Approach for Automatic Speaker Recognition : Optimal division of the Frequency Domain", in *Audio and Video Biometric Person Authentication, Lectures Notes in Computer Science*, vol. 1206, B. al., Ed.: Springer Verlag, 2003, pp. 195-202.
5. J. Caelen, V. Eglin and S. Hollard, "Analyse de documents par oculométrie", in *Assistance intelligente à la recherche d'informations*, chapitre 9, Hermès, Ed., Paris: Gaussier E., Stefanini M.H., 2003, pp. 285-315.
6. J. Caelen, V. Eglin, S. Hollard and B. Meillon, "Mouvements oculaires et évaluation de documents électroniques", in *Document électronique dynamique, Actes du sixième colloque international sur le document électronique (CIDE-6)*, U. d. Caen/Europia, Ed., Caen: Faure C., Madelaine J., 2003, pp. 107-135.
7. J. Caelen, "Dialogue homme-machine et recherche d'information", in *Assistance intelligente à la recherche d'informations*, chapitre 7, Hermès, Ed., Paris: Gaussier E., Stefanini M.H., 2003, pp. 219-254.
8. J. Caelen, "Systèmes interactifs : essai de caractérisation", in *Communication homme-systèmes informationnels*, Chapitre 2, Hermès, Ed., Paris: B. Miège, 2003, pp. 41-58.

### Année 2004

9. J. Caelen, *Le consommateur au coeur de l'innovation*: CNRS Editions, Paris, 2004.
10. J. Caelen, "Préface de livre", in *VoiceXML*, de J. Rouillard, Ed.: Vuibert, Paris, 2004.

### Année 2005

11. A. Lecomte and A. Naït-Abdallah, "On Expressing Vague Quantification and Scalar Implicatures in the Logic of Partial Information", in *Logical Aspects of Computational Linguistics*, Ed.: FoLLI/LNAI, Springer-Verlag (LNCS), April 2005.
12. A. Lecomte, "Logical Models for Dialogue", in *Logic and Dialogue*, D. Vanderveken, D. Vernant, and J. Caelen, Eds., Ed.: Kluwer, Amsterdam, 2005.
13. A. Lecomte, "Categorial Grammar for Minimalism", in *Language and Grammar*, C. Casadio, P.J. Scott, R.A.G. Seely Eds., CSLI Publications, Stanford, pp. 163-189, 2005.
14. M. Tomokiyo, G. Chollet and S. Hollard, "Studies of Emotional Expressions in Oral Dialogues", in *Research on Computer Science, Universal Networking Language, Advances in Theory and Applications*, Mexico, Ed.: Cerdeñosa J., Gelbukh A., Tovar E., pp. 286-299, 2005.

## Directions d'ouvrages (DO)

### Edition de livres

1. J. Caelen, D. Vernant and D. Vanderveken, *Logic and Dialogue*, à paraître : Kluwer ed., Amsterdam, 2005.

## Autres publications (AP)

### Année 2001

1. C. Bergamini, J. F. Serignat, L. Besacier and D. Vaufraydaz, "French Database for Automatic Speech Recognition", Laboratoire CLIPS-IMAG, GEOD, Grenoble, Rapport de contrat avec ISL (Interactive Systems Laboratories) de l'Université de Karlsruhe, septembre 2001.
2. J. Caelen, "Rapport scientifique du laboratoire CLIPS 1998-2000", Laboratoire CLIPS-IMAG, Grenoble, septembre 2001.
3. E. Castelli and G. Nguyen Trong, "MICA : Centre de recherche international en Multimédia, GEOD - 30 -

Informations, Communication et Application”, International Research Center MICA, Hanoi, Vietnam, Rapport de demande de soutien au CNRS, 5 janvier, 2001.

4. E. Castelli, M. Gery and B. Guerin, “Masters en traitement du signal, 10ème mission. Cours Communications numériques et réseaux industriels, Multimédia, Reconnaissance des formes.”, Centre de Recherche International MICA, Hanoi, Vietnam, Rapport de Mission, 12 février-8mars 2001.
5. J. F. Serignat, “Rapport scientifique de l'équipe GEOD du CLIPS 1998-2000”, Laboratoire CLIPS-IMAG, Grenoble, septembre 2001.

#### Année 2002

6. L. Besacier, C. Bergamini, R. Lamy, D. Vaufreydaz, J. F. Serignat, P. Ged, F. Boiteux and L. Cochet, “Expertise en vue de la réalisation d'un système de reconnaissance de la parole téléphonique”, Laboratoire CLIPS-IMAG, Grenoble, Rapport de fin de contrat avec la Société PROSODIE SA, Mai 2002.
7. J. F. Serignat, E. Castelli and V. Rialle, “Reconnaissance de situations de détresse en Habitat Intelligent Santé”, Rapport final du projet IMAG RESIDE-HIS, Laboratoires CLIPS & TIMC, IMAG, Grenoble, 40 p., septembre 2002.

#### Année 2003

8. D. Istrate and M. Vacher, “Détection de signaux sonores noyés dans le bruit”, Laboratoire CLIPS-IMAG, Grenoble, 60 p., février 2003.
9. D. Istrate and M. Vacher, “Reconnaissance de classes de sons”, Laboratoire CLIPS-IMAG, Grenoble, 76 p., juin 2003.
10. D. Istrate and M. Vacher, “Application des ondelettes à la détection de signaux sonores noyés dans le bruit”, Laboratoire CLIPS-IMAG, Grenoble, 35 p., juillet 2003.
11. J.F. Serignat, “Interaction Parlée Homme-Machine”, Projet LAFMI-CNRS, Rapport d'avancement 1<sup>ère</sup> année 2002-2003, Laboratoire CLIPS, UJF-IMAG, Grenoble, 25p., septembre 2003.
12. A. Xuereb and J. Caelen, “PVE (Portail Vocal d'Entreprise), Architecture Générale : principes”, Rapport de contrat RNRT 2001, Laboratoire CLIPS-IMAG, Grenoble, 81 p., 18 septembre 2003.

#### Année 2004

13. L. Besacier, A.C. Descalle, “Analyse des dialogues et développement de logiciels de traitement”, Rapport Interne 3.4 – SP3.4, Projet RNRT COUCOU, Laboratoire CLIPS, UJF-IMAG, Grenoble, 2004.
14. J. Caelen, “PVE : Portail Vocal d'Entreprise”, Rapport final du contrat PVE, RNRT, Laboratoire CLIPS, UJF-IMAG, Grenoble, 30 p., 2004.
15. J. Caelen, “ACE : Agent Conversationnel Expressif”, Rapport final du contrat ACE, RIAM, Laboratoire CLIPS, UJF-IMAG, Grenoble, 30 p., 2004.
16. J.F. Serignat, L. Villaseñor-Pineda, J. Caelen, D. Vaufreydaz, “Interaction Parlée Homme-Machine”, Projet LAFMI-CNRS, Rapport final partie française 2002-2004, Laboratoire CLIPS, UJF-IMAG, Grenoble, 34p., octobre 2004.
17. J. F. Serignat, V. Rialle, M. Vacher, D. Istrate and G. Virone, “DESDHIS: Détection de Situations de Détresse dans un Habitat Intelligent Santé”, Rapport final du contrat DESDHIS, ACI Technologies pour la Santé (CNRS 2002-2004), Laboratoire CLIPS, UJF-IMAG, Grenoble, 25 p., novembre 2004.

#### Année 2005

18. L. Besacier, “Rapport de mission projet TALK (AUF)”, Mission ITC (Phnom-Penh) - MICA (Hanoi), Laboratoire CLIPS/GEOD, Grenoble, 3 p., mars 2005.
19. J. Caelen, “COUCOU : Conception participative Orientée Usage de services de Communication et d'Objets Ubiquistes”, Rapport final et livrables du contrat COUCOU, RNRT, Laboratoire CLIPS, UJF-IMAG, Grenoble, 20 p., 2005.
20. V. B. Le, “Rapid Acoustic Models Construction for Vietnamese using Multilingual Models”, CMU (Pittsburgh) Mission Report, CLIPS/GEOD Laboratory, Grenoble, 19 p., March 2005.
21. D. Tuffelli, “Tutorial pour le logiciel PRAAT : Manipulations élémentaires”, Rapport interne, Laboratoire CLIPS-IMAG, équipe GEOD, Grenoble, 95 p., 7 juin 2005.

### Thèses et habilitations

1. D. Vaufreydaz, “Modélisation statistique du langage à partir d'Internet pour la reconnaissance automatique de la parole continue”, DU Spécialité Informatique, Université Joseph Fourier (UJF), Grenoble, 226 p., 7 janvier 2002.
2. Q.C. Nguyen, “Reconnaissance de la parole en langue vietnamienne”, DU Spécialité Signal Image Parole Télécom (SIPT), Institut National Polytechnique (INP), Grenoble, 176 p., 19 juin 2002.
3. M. Z. Kurdi, “Contribution à l'analyse du langage oral spontané”, DU Spécialité Informatique, Université Joseph Fourier (UJF), Grenoble, 355 p., 18 avril 2003.
4. D. Istrate, “Détection et reconnaissance des sons pour la surveillance médicale”, DU Spécialité Signal

- Image Parole Télécom (SIPT), Institut National Polytechnique (INP), Grenoble, 183 p., 15 décembre 2003.
5. M. Ahafhaf, "Evaluation des systèmes de dialogues oral homme-machine : quelques éléments linguistiques appliqués au paradigme DCR", DU Spécialité Sciences du Langage, Université Stendhal, Grenoble, 30 septembre 2004.
  6. Y. Fouquet, "Modélisation des attentes en dialogue oral", DU Spécialité Informatique Système et Communication, Université Joseph Fourier (UJF), Grenoble, 220 p., 26 octobre 2004.
  7. D. Moraru, "Segmentation en locuteurs de documents audios et audiovisuels : application à la recherche d'information multimedia", DU Spécialité Signal Image Parole Télécom (SIPT), Institut National Polytechnique (INP), Grenoble, 217 p., 20 décembre 2004.
  8. N. H. Nguyen, "Dialogue Homme-Machine : modélisation de multisession", DU Spécialité Informatique Système et Communication, Université Joseph Fourier (UJF), Grenoble, 207 p., 26 janvier 2005.
  9. P. Mayorga-Ortiz, "Reconnaissance vocale dans un contexte de voix sur IP : diagnostic et propositions", DU Spécialité Signal Image Parole Télécom (SIPT), Institut National Polytechnique (INP), Grenoble, 177 p., 8 février 2005.

## 6. PRINCIPALES RESPONSABILITES SCIENTIFIQUES ET ADMINISTRATIVES

GEOD (R. Lamy, D. Vaufreydaz) a été l'organisateur des "Cinquièmes Rencontres Jeunes Chercheurs en Parole" (RJC'2003) à Grenoble du 23 au 25 septembre 2003, manifestation parrainée par l'Association Francophone de la Communication Parlée (AFCP) dont l'objectif est de fournir aux doctorants ou jeunes docteurs l'occasion de se rencontrer, de présenter leurs travaux et d'échanger leurs idées sur les divers domaines de la parole.

L. Besacier

Est membre élu du bureau de l'AFCP (Association Francophone de la Communication Parlée) depuis 2001.  
Est relecteur pour les revues *Computer Speech and Language*, *IEEE Transactions on Speech and Audio Processing*, *IEEE Signal Processing Letters*, *IEEE Transactions on Signal Processing*, *IEEE Transactions on Multimedia*, *Traitement du Signal*, *Acta Acustica*.

Est relecteur pour les conférences : *Eurospeech 2005*, *RECITAL 2005*, *Speaker Odyssey 2004 et 2006*, *Journées d'Etude sur la Parole 2002 2004 et 2006*, *ECCTD 2001*.

Est organisateur d'une session spéciale Biométrie<sup>31</sup> à la conférence *ISPA 2005*.

Est membre des commissions de spécialistes 27ème section de l'Université J. Fourier (depuis 2003) et de l'Université d'Avignon (2001-2004).

Est membre du conseil scientifique de l'université J. Fourier (depuis Janvier 2005).

J. Caelen

Est directeur du laboratoire CLIPS.

Est éditeur en chef de la revue RIHM (édité par Europa Productions, éd. Paris).

Est relecteur dans *SpeechCom*, *IJHS* et *Signal Processing* (revues internationales) et dans *Free Speech Journal* (revue électronique internationale).

Est membre du comité scientifique de congrès internationaux.

Est membre du conseil de direction du GdR I3.

Est membre de sociétés savantes de l'ACL, de l'ESCA, de la SFA.

A été animateur d'ateliers sur les objets communicants à l'Agence Rhône-Alpes Numérique en 2002-2003 et membre du comité scientifique SOC'2003.

Est coordonnateur du PPF Objets, Agents et Environnements Communicants au sein du GIS PEGASUS.

A été organisateur du colloque international *Dialogue et Logique*, Grenoble 2002.

A été organisateur du congrès international SOC (Smart Object Conference), Grenoble 2001 et 2003.

Est membre du comité scientifique du congrès MFI depuis 2003.

A été co-organisateur du workshop WACA, Grenoble 2005.

---

<sup>31</sup> <http://www.isispa.org/sssipb.html>

E. Castelli

Est Directeur Adjoint du Centre de Recherche International MICA (Multimédia, Informations, Communication et Applications) à Hanoi, sous la cotutelle de l'INP Grenoble et de l'Institut Polytechnique de Hanoi.

Est relecteur du J.A.S.A. (Journal of the Acoustic Society of America).

Est rédacteur en chef du Bulletin du PFIEV (Programme de Formation d'Ingénieurs d'Excellence au Vietnam).

S. Hollard

Est la correspondante officielle du laboratoire CLIPS auprès de l'ISDN (Institut des Sciences du Document Numérique).

Est relectrice de In Cognito (Revue de l'association de Recherche en Sciences cognitives).

A. Lecomte

Est membre du comité de rédaction de la revue TAL, revue de l'ATALA, éditée chez Hermès, avec le concours du CNRS,

Est membre du comité de programme des colloques LACL'01, LACL'05, MOL9, Formal Grammar 02, 05, ACL'01,

Est relecteur pour ACL'01, EACL'03, ARC 01,

Est membre associé de l'équipe SIGNES (INRIA et LaBRI, Bordeaux 3),

Est membre associé du Groupe de Recherches " Philosophie, Langages et Cognition " de l'université Grenoble 2,

Est membre du GDR " Sémantique Formelle et Modélisation " (dir. F. Corblin),

Est membre du CPER " Mathématisation " (Région Rhône-Alpes),

Est membre de la commission de spécialistes 27-72 de l'UPMF, de la CS 19 de l'UPMF, de la CS 72 de l'Université de Nantes,

Est membre du CNU, section 72,

Est ancien directeur de l'UFR " Sciences de l'Homme et de la Société " de l'UPMF,

Est co-directeur du Master " Information, Cognition et Apprentissages ", master de site co-organisé par l'UJF, l'UPMF et l'INPG

J.F. Serignat

Est membre de la SEE (Société des Electriciens Electroniciens) de France et membre du Comité Régional Dauphiné-Savoie.

Est Responsable du Groupe Relations Industrielles à l'ENSERG (Ecole Nationale Supérieure d'Electronique et de Radioélectricité de Grenoble).

## 7. PERSPECTIVES DE RECHERCHE 06-10

Les perspectives de recherche de GEOD s'inscrivent dans la continuité des axes de recherche actuels et son domaine de recherche demeurera sans doute, pour les prochaines années, centré sur l'interaction orale avec les deux axes principaux : *Reconnaissance* (au sens large) et *Dialogue*. Les projets en cours ou déposés et les encadrements en cours peuvent servir à illustrer les prochaines tendances :

### Direction 1 : Multilinguisme

Le but de cet axe, déjà en cours de développement (Thèse de Viet-Bac Le) est d'étudier et de proposer des solutions et des outils permettant d'accélérer la portabilité des systèmes de reconnaissance automatique de parole continue grand vocabulaire vers une nouvelle langue. Dans un premier temps, nous avons choisi de travailler sur le vietnamien, étant donné nos collaborations actuelles avec le laboratoire MICA<sup>32</sup>, situé à Hanoi. Mais nous envisageons dans un futur très proche d'appliquer ces travaux vers d'autres langues minoritaires (langues pour lesquelles très peu de corpus est disponible) comme le Khmer (parlé au Cambodge) et les langues des pays baltes (projet ANR JC déposé par L. Besacier).

Les projets CORUS (MAE) et TALK (AUF) qui se prolongent en 2006, marquent l'importance du

---

<sup>32</sup> <http://www.mica.edu.vn/Home/index.jsp>

multilinguisme (principalement en direction des langues peu dotées) dans les recherches en traitement automatique de la parole.

Un projet MULTILING (L. Besacier, B. Bigi), déposé à l'appel d'offres ANR jeunes chercheurs sur la reconnaissance automatique de parole multilingue renforce l'intérêt marqué pour l'aspect multilingue dans les préoccupations de recherche de GEOD, avec notamment l'ajout de nouvelles langues dans les perspectives de recherche pour montrer la portabilité des approches proposées par GEOD.

La thèse de Tien-Ping TAN, commencée en octobre 2004 (dirigée par L. Besacier) et dont le sujet de recherche est "*Reconnaissance du français pour des locuteurs non natifs*", bien que centrée sur la reconnaissance du français, relève en fait de l'aspect multilingue par l'étude de l'influence des accents étrangers sur les performances des systèmes de reconnaissance de parole.

## **Direction 2 : Multimodalité**

Bien que fortement positionnés dans le domaine du traitement automatique de documents audio au départ, nous envisageons de faire évoluer nos travaux de recherche vers un contexte multimodal où, quelque soit le domaine applicatif, on rencontre des flux de données provenant de multiples modalités (audio, vidéo et autres). Ceci s'est déjà concrétisé par une première collaboration avec le laboratoire LIS et l'équipe *MRIM* du CLIPS dans le domaine de la recherche de documents vidéos (expériences réalisées lors des campagnes d'évaluation TREC Vidéo 2003 et 2004). Dans ces premiers travaux, nous avons tenté de fusionner des modalités audio et vidéo, tant au niveau des données (*early integration*, recherche par similarité de vidéos-clips dans une base de vidéos) qu'au niveau des décisions (*late integration*, segmentation de vidéos en histoires) (ACT-48, Besacier, 2004).

Dans les années à venir, nous envisageons notamment de travailler sur des données audiovisuelles en renforçant l'intégration des deux modalités. A ce sujet, un travail a déjà été réalisé avec un post-doctorant (Nicolas Eveno) issu du LIS, et financé par le CNRS sur l'ACI *Sécurité Informatique* (projet BIOMUL déjà mentionné). Le potentiel des travaux de Nicolas Eveno sur la détection de contours labiaux à partir d'une vidéo nous permet d'envisager d'appliquer un traitement bimodal (audio et lèvres) dans les domaines suivants :

- détection automatique de playback (par mesure d'asynchronie entre deux sources audio et vidéo) pour des applications biométriques (projet BIOMUL)
- traitement bi-labial de la parole audiovisuelle (projet TELMA qui implique les laboratoires grenoblois CLIPS, LIS et ICP)
- recherche d'information multimodale par utilisation d'indices audio/vidéo.

Par ailleurs, le co-encadrement d'un doctorat avec le laboratoire ICP est en cours (bourse de thèse fléchée, projet BQR TELMA); ainsi que le co-encadrement de Masters Recherche avec le laboratoire LIS (par l'intermédiaire de D. Pellerin). Une prolongation du projet TELMA financé par l'ANR (appel RNTS) vient par ailleurs d'être acceptée.

Enfin, ces travaux trouveront un cadre expérimental par l'intermédiaire des *campagnes d'évaluations internationales TREC Vidéo* organisées par le NIST aux USA (pour l'axe multimodalité) et au niveau national, par l'intermédiaire des *programmes Technolangue (pour l'axe multilinguisme) et Technovision (pour l'axe multimodalité) du ministère de la recherche* dans lesquels le CLIPS est impliqué.

## **Direction 3 : Espaces perceptifs**

Dans le domaine des espaces perceptifs, les recherches se poursuivent dans la continuité de la coopération (projet DESDHIS) entre l'équipe AFIRM du laboratoire TIMC et l'équipe GEOD du CLIPS. Une thèse (Anthony FLEURY) co-encadrée par N. Noury de TIMC et M. Vacher du CLIPS a démarré en octobre 2005 sur le thème : "*Détection de motifs temporels dans les environnements multi-perceptifs – Application à la détection automatique des activités de la vie quotidienne d'une personne suivie à domicile par télémédecine*". Cette thèse entre dans le cadre du projet SIEMAD (**S**ystème d'**I**nformation **E**mbarqué pour le **M**onitoring **A**mbulatoire de la personne **D**épendante), supporté par la Région Rhône-Alpes et déposé par l'équipe AFIRM du laboratoire TIMC, dans lequel GEOD est impliqué pour le traitement du signal sonore.

De même, un nouveau travail a commencé en octobre 2005 (Stéphane CHAILLOL, Mémoire CNAM) sur la "*Détection et la classification des sons de la vie courante à l'aide de chaînes de Markov cachées (HMM)*", dans le but d'améliorer les performances (taux de reconnaissance, robustesse au bruit) du système actuel à base de mélanges de modèles gaussiens (GMM) qui ne prend pas suffisamment en compte l'aspect évolutif temporel des sons dans un environnement perceptif.

#### Direction 4 : Dialogue

En dialogue, deux nouvelles thèses commencent en 2005 : Vladimir POPESCU en co-tutelle avec la Roumanie (Direction : J. Caelen, C. Burileanu) sur le “*Dialogue pertinent Homme-Machine : génération des réponses orales en langage naturel*” et Anas OUAYOUCH (Dir. J. Caelen) sur “*Jeu de négociation en langage naturel*” qui marquent la volonté de GEOD de continuer dans la modélisation des systèmes de dialogues mieux adaptés et toujours plus proches du comportement d’un opérateur humain. Les problèmes importants qu’il faut continuer à approfondir sont :

- L’interprétation pragmatique,
- La génération d’énoncés naturels et expressifs,
- La prise en compte du processus de négociation entre plusieurs locuteurs.

Depuis maintenant une dizaine d’années, des tentatives pour ajouter de l’émotion à la parole de synthèse ont été entreprises. Cependant, la parole émotionnelle synthétique sonne toujours peu impliquée et, dans un contexte de dialogue/interaction homme-machine, ce point est particulièrement crucial. En effet, un système de synthèse (*text-to-speech*) ne produit des réponses orales qu’en mettant en entrée des énoncés écrits. Or on ne dialogue pas comme on lit, on ne dialogue pas non plus comme on écrit : les formes linguistiques sont différentes (au niveau syntaxique et au niveau prosodique, mais aussi par un emploi plus fréquent d’ellipses, de formes emphatiques, etc.), et surtout ces formes véhiculent une composante perlocutoire évidente. L’adéquation des réponses de la machine vis-à-vis des interventions de l’utilisateur peut alors être contestable d’une part et les échanges peuvent paraître très artificiels d’autre part. Plus que la recherche d’un dialogue *naturel* qui n’est peut-être pas une fin en soi (exceptée dans le cadre de recherches fondamentales en linguistique), il faut viser l’intelligibilité et l’adéquation ou la justesse d’une réponse qui procède de sa pertinence ; on sait également qu’une réponse peu pertinente n’est pas compréhensible. Au niveau des interfaces homme-machine actuelles il y a donc nécessité de contrôler cette pertinence pour deux raisons : (a) pour diminuer le coût cognitif de l’utilisateur (avec des retombées évidentes sur la performance d’exécution de la tâche, la diminution de la fatigue, etc.) et, (b) pour respecter le principe ergonomique *d’observabilité* qui stipule que l’utilisateur doit avoir une conscience claire des véritables capacités de la machine.

De plus, en conclusion, on peut aussi constater une influence grandissante du *multilinguisme* (bien qu’il soit déjà présent dans les activités de recherche de GEOD, notamment en direction des langues peu dotées) dans de nombreuses facettes de ces recherches. Le multilinguisme ouvre en effet des opportunités de coopération avec d’autres équipes au sein de projets internationaux et contribue aussi au rayonnement international de l’équipe. Le passé récent a montré aussi que GEOD, bien que n’ayant travaillé jusqu’alors que sur la reconnaissance du français, a trouvé avantage à coopérer avec l’équipe GETA au sein de projets internationaux de dialogue homme-homme médiatisé (CSTAR et NESPOLE) où l’aspect multilingue était une caractéristique fondamentale de ces projets. En effet, la mise en commun des spécificités complémentaires de chacune des deux équipes (“traduction multilingue” pour le GETA et “traitement de l’oral” pour GEOD), s’est révélée judicieuse et source d’efficacité pour une coopération au sein de ce type de projets. Aussi, est-il envisagé dans un proche avenir d’effectuer un regroupement des deux équipes GEOD et GETA au sein d’une même entité, dont le nom n’est pas encore vraiment fixé, mais qui pourrait se rapprocher de **GETALP**, pour “Groupe d’Etude en Traduction/Traitement Automatique des Langues et de la Parole”.