

**RAPPORT D'ACTIVITÉS DE LA THÉMATIQUE
LANGAGE DU LIA
NOVEMBRE 2011**

**LABORATOIRE INFORMATIQUE D'AVIGNON
EA 4128**

Université d'Avignon et des Pays de Vaucluse



1.1.1 Présentation générale

1.1.1.1 Positionnement et bilan synthétique

Les activités du LIA dans le domaine du traitement automatique du langage ont débuté dès la création du laboratoire, par les travaux menés sous la direction de Henri Méloni en traitement de la parole. Le spectre thématique s'est ensuite élargi avec l'arrivée de Marc El-Bèze, en traitement de l'écrit puis avec celle de Renato De Mori, en modélisation du dialogue homme-machine. En juin 2011, la thématique langage regroupait 11 enseignants-chercheurs permanents. Sa couverture thématique sur le domaine du traitement automatique du langage est large, les recherches réalisées portant sur une grande variété d'objets linguistiques (parole lue ou spontanée, dialogues, texte mono ou multilingues...) des niveaux et des paradigmes de représentation.

Cette diversité des sujets d'études est abordée avec une culture scientifique et technologique commune, principalement basée sur les statistiques bayésiennes, l'apprentissage et la classification automatique, la théorie de l'information, le traitement du signal et l'analyse de données.

Enfin, les méthodologies mises en œuvre impliquent le plus souvent des expérimentations sur corpus, le développement de systèmes et leur engagement dans des campagnes d'évaluation. Globalement, la thématique Langage s'organise en trois sous-thématiques qui concernent le traitement de l'écrit, de l'oral et la modélisation du dialogue. Ces travaux s'articulent autour de sujets porteurs, souvent interdisciplinaires, qui adressent les problèmes d'extraction d'information, de structuration, d'indexation, de ré-éditorialisation des contenus linguistiques.

Ces dernières années, la thématique a augmenté significativement sa production scientifique (x2.5 en ACL, x2 C-ACTI par rapport à [2002,2005]) et sa capacité d'encadrement (6 HDR et 23 thèses ont été soutenues). L'activité contractuelle a, elle aussi, sensiblement progressé, avec 18 projets ANR et 4 européens acceptés.

Par ailleurs, l'équipe a confirmé son rôle de producteur de ressources pour la communauté scientifique en continuant le développement de plate-formes logicielles et de corpus, telles que CORTEX (résumé automatique), SPEERAL (transcription enrichie), ou ALIZE qui est utilisée par plus de 40 industriels et laboratoires dans le monde (dont le CMU, MIT, Univ. de Brisbane, etc.).

Ce dynamisme a permis de développer des relations nationales et internationales qui se sont traduites par de nombreuses visites de chercheurs étrangers, des séjours des membres de la thématique dans des laboratoires réputés (Cambridge, Drexel University, Université de Mexico), des thèses en co-tutelles, l'organisation de conférences nationales et internationales (EACL 2012 et InterSpeech 2013 seront organisées et co-organisées par le LIA).

1.1.1.2 Effectifs

En Juin 2011, 11 enseignants-chercheurs permanents (2 professeurs, 5 Mdc HDR, 4 Mdc) et un Professeur Emérite (Renato De Mori) exerçaient leur activité dans la thématique Langage. F. Béchet et P. Bellot étaient tout deux actifs dans cette thématique avant leurs promotions sur des emplois de PR respectivement au LIF en 2009 et au LSIS en 2011. Enfin, Georges Linarès a été promu PR au LIA, en septembre 2011.

Le personnel non permanent est composé de 18 doctorants et de 11 ingénieurs, ATER ou post-doctorants, 23 thèses ayant été soutenues depuis 2006. Les post-doctorants, ingénieurs et la majeure partie des doctorants (13 sur les 18 en cours) sont financés sur projets de recherche industriels, régionaux, nationaux ou internationaux.

Nom	Prénom	Année de naissance	Corps grade	Section CNU	HDR	Date d'arrivée dans l'unité
BELLOT *	Patrice	1972	MCF	27	X	sept.-00
BONASTRE	Jean - François	1967	PR2	27	X	sept.-94
BECHET *	Frédéric	1967	MCF	27	X	sept-94
DE MORI	Renato	1941	PR Émérite	27	X	janv.-97
EL-BÈZE	Marc	1950	PR1	27	X	sept.-93
FREDOUILLE	Corinne	1973	MCF	27		sept.-03

HUET	Stephane	1983	MCF	27		Sept-10
JOURLIN	Pierre	1971	MCF	27		sept.-00
LEFEVRE	Fabrice	1970	MCF	27	X	fév.-05
LINARES	Georges	1968	MCF	27	X	sept.-98
MATROUF	Driss	1967	MCF	27	X	sept.-98
NOCERA	Pascal	1963	MCF	27		oct.-92
SANJUAN	Éric	1971	MCF	27		juill.-06
TORRES MORENO	Juan Manuel	1964	MCF	27	X	sept.-03

Table T1.1 : Enseignants-chercheurs permanents actifs dans la thématique Langage entre 2006 et Juin 2011

1.1.1.3 Animation scientifique, Rayonnement

La thématique a augmenté sa visibilité nationale et internationale par l'organisation ou la co-organisation de conférences, en s'investissant dans les sociétés savantes du domaine et en multipliant des collaborations qui permettent de maintenir ou de développer des liens avec la communauté scientifique.

Organisation de conférences :

- co-organisation de MMUA : MultiModal User Authentication, Toulouse 2006
- organisation JEP/TALN 2008 (<http://lia.univ-avignon.fr>, plus de 300 participants) à l'UAPV
- organisation de MajecSTIC (<http://majecstic2009.univ-avignon.fr/>) et des Rencontres des Jeunes Chercheurs en Parole (RJCP, <http://rjcp2009.univ-avignon.fr/>) en 2009 (UAPV, plus de 150 participants)
- co-organisation des Journées de Phonétique Clinique (JPC'03) en 2009 (Aix-en-Provence),
- co-organisation de la session spéciale Robust Speech Recognition à Acoustics'08, Paris 2008.
- organisation de CORIA 2011 (Conference en Recherche d'Information et Applications, <http://coria11.univ-avignon.fr>) à l'UAPV
- en 2010, le LIA a été retenu pour :
 - organiser **EACL 2012 (+de 600 participants) à l'UAPV** (<http://eacl2012.org>)
 - **co-organiser InterSpeech 2013 (+ de 1000 participants)**

Animation scientifique :

Nous participons à l'administration d'un certain nombre de sociétés savantes à différents niveaux : 2 membres dans le comité Speech and Langage Technical Comitee d'IEEE (J.F. Bonastre et F. Béchet), 2 membres au CA de l'AFCP, vice-présidence de l'ISCA, membre du CA de l'ATALA,... Enfin, nous participons très largement aux activités de relecture et aux comités scientifiques ou éditoriaux (revue Traitement Automatique des Langues, IEEE Transaction on Multimedia, IEEE Transaction on Audio, Speech and Langage Processing, Speech Communication, Computer Speech and Langage,...).

1.1.1.4 Projets, Collaborations et valorisation

La majeure partie des ressources de la thématique provient de projets et de contrats industriels, qui nous permettent notamment de financer des doctorants (10 des 18 doctorats en cours sont financés sur projet ANR, européens ou conventions CIFRE). En 2011, Patrice Bellot a obtenu un Google [Award](#) (50 000\$) pour conduire des travaux en extraction d'information pour les bibliothèques numériques (programme Digital Humanities).

Cette activité contractuelle s'appuie sur des plate-formes logicielles qui sont développées au LIA et sur une visibilité entretenue par nos activités de recherche et de valorisation.

Projets Nationaux et européens

La thématique dans 5 projets européens et 18 projets ANR dont la liste figure dans le tableau T1.2.

Intitulé	Sujet	Responsable scientifique	Type	Co-contractants ou nature du partenaire	Début /Fin	Sub. LIA
MISTRAL	Plateforme open source d'authentification biométrique	Jean-Francois Bonastre	ANR	LIUM, THALES, Calistel, CLIPS/IMAG, Eurecom	2006/2009	106K€
EPAC	Caractérisation et reconnaissance de la parole spontanée	Frederic Bechet	ANR	IRIT, LI, LIUM	2006/2010	35K€
AVISON	Indexation audio multilingue de vidéos chirurgicales	Georges Linarès	ANR	Xtensive Technologies, IRCAD	2008/2011	263K€
PIITHIE	Plagiat et Impact de l'Information Textuelle recherche	Patrice Bellot	ANR	Sinequa, LINA, Advestigo	2006/2008	152K€
RP2M	Resume Plurimedia, Multi-documents et Multi-opinions	Georges Linarès	ANR	Eurecom, Sinequa, Syllabs, Wikio	2008/2011	182K€
DesPhoAPADy	Analyse Phonético-Acoustique de la parole dysarthrique	Corinne Fredouille	ANR	LPP, LPL	2009/2012	81K€
SEQUOIA	Analyse syntaxique probabiliste du français	Patrice Bellot	ANR	LIF, INRIA, LALIC	2009/2012	153K€
PORT-MEDIA	Robustesse et portabilité multilingue-Multi-domaine des systèmes de compréhension de la parole	Fabrice Lefevre	ANR	ELRA, LIG, LIUM, LORIA	2009/2012	117K€
PI	Reconnaissance de la parole pour des langues peu dotées	Pascal Nocera	ANR	LIG, MICA	2009/2012	120K€
ASH	Combinaison de systèmes de reconnaissance de la parole	Georges Linarès	ANR	LIUM, IRISA	2009/2012	95K€
DECODA	Depouillement de conversations issues de centres d'appels	Marc El-Bèze	ANR	LIF, Sonear, RATP	2009/2012	154K€
BIOBIMO	Biométrie multimodale (audio/vidéo)	Corinne Fredouille	ANR	Eurecom, E2V	2007/2010	149K€
SIGMUND	Surveillance de flux audio	Pascal Nocera	ANR	INA, CNAM,	2006/2009	220K€
SUMACC	Détection et suivi de concepts multimédias	Georges Linarès	ANR	Eurecom, Wikio, Syllabs	2011/2013	250K€
OTMedia	Observatoire TransMedia	Pascal Nocera	ANR	INA, AFP, INRIA, Syllabs, Univ. Paris III	2011/2013	150K€
PERCOL	Identification de personnes dans des flux multimédias	Corinne Fredouille	Défi ANR	Orange, LIF, Univ. Lille	2011/2013	139K€
CAAS	Recherche d'information contextualisée	Patrice Bellot	ANR	Univ P. Sabatier CNRS Délégation régionale Midi-Pyrénées	2011/2013	150K€
DIVINES	Etude diagnostique des systèmes de transcription automatique	Renato de Mori	EU	Orange, Telecom Italia, Univ Oldenburg, Multitel, Babel, Pol. di Torino	2005/2007	120K€
HARTES	Approche holistique pour le design d'applications temps réel, embarquées, reconfigurables	Jean-François Bonastre	EU: IP/VI PCRD	Atmel, FAITAL, FI, INRIA, Thomson, Thallès, Univ. di Ferrara, Univ. Politecnica delle Marche, TU Delft, Segula, Scaleo Chip, Pol. di Milan, Pol. Di Bari	2006/2009	220K€
LUNA	Compréhension de la parole dans des systèmes de communication multilingues	Renato De Mori	EU: STRE PS/VII PCRD	Loquendo, Univ. Trento, RWTH, Orange, IPAN, CSIP	2006/2009	362K€
MOBIO	Mobile BioMetry	Driss Matrouf	EU: ICT-FP7	IDIAP, BUT, Univ. Oulu, EvePmedia, IdeArk, Visidon, Univ. Surrey, Univ. Manchester	2007/2010	210K€
BioSpeak	Autentification biométrique	J.F. Bonastre	Prog. EuroStars	Calistel, Univ. de Swansea, Validsoft Ireland Ltd, Acapela Goup France, Validsoft	2010/2013	76K€

Table T1.2 : Projets ANR et européens dans la thématique Langage

Projets industriels et associatifs

Contrats CIFRE : EDF, Thalès, France Télécom, ETT, Xtensive, DGA, Semantia, Akator, Projet NuSpecs avec la société italienne Leaff, contrat de collaboration avec la société Zenvia (Montpellier), projet VML (2009) financé par l'association « Vaincre les Maladies Lysosomales ».

Plates-formes, corpus, support et valorisation

La thématique développe des composants logiciels, des outils et des environnements dont l'intérêt dépasse le cadre des recherches qui y sont menées. Notre politique est de distribuer ces outils non seulement à la communauté scientifique, mais aussi vers les industriels, en contractualisant les développements scientifiques que ces échanges génèrent.

ALIZE (projet MISTRAL) : plateforme d'authentification biométrique, LGPL. Quelques utilisateurs : Thalès (FR), MIT (USA), CMU (USA), Berkeley (ICSI, USA), LIMSI (FR), QUT (Australie), IIT/Kampur (Inde), ENST (FR), University of Swansea (UK), ...

SPEERAL : système de transcription automatique diffusé librement sous licence LGPL. Quelques utilisateurs : ETT (Cifre), Xtensive (Cifre), Zenvia, Voice Insight, LPP (Paris 3)...

CORTEX : Système de résumé automatique de documents. Quelques utilisateurs : École Polytechnique de Montréal (Canada) ; Institut Universitaire de Linguistique Appliqué (Barcelone) ; Vanguard Engineering Puebla S.A. (Mexique, en voie de tests) ; Thalès (France, en test).

Campagnes d'évaluation

Nous engageons régulièrement les systèmes développés au LIA dans **des campagnes d'évaluation** nationales et internationales. Ceci nous permet de confronter nos plate-formes aux systèmes état de l'art mais aussi d'obtenir des ressources et des environnements d'évaluation de référence.

NIST : vérification du locuteur (depuis 2006), segmentation en locuteurs (2006,7,9,2010,11), identification des langues (2009, 2011), résumé automatique (**DOC 06-07/TAC**), questions/réponses (**INEX 2008-10**), recherche d'information (TREC 2007, 2010)

ESTER 2005, 2008 : transcription enrichie, segmentation et suivi de locuteur, détection des entités nommées

DEFT : fouille de textes (2007/09)

CLEF : questions-réponses (2006)

Mobio, Evalita : vérification du locuteur (2010)

WMT : traduction automatique (2011)

MediaEval : structuration, et indexation multimédia (2011)

Collaborations internationales

Ces dernières années, l'équipe a multiplié les collaborations internationales dans le cadre de conventions de collaborations bilatérales, de co-encadrement de thèses, d'accueil de visiteurs ou de séjours dans des laboratoires étrangers.

Collaborations contractualisées avec : Maribor University (accord bilatéral, Slovénie), Brno University of Technology (BUT, accord bilatéral, Brno, république Tchèque), Queensland University of Technology (QUT, Brisbane, Australie, ARC Linkage), projet NuSpecs (avec la société italienne LEAFF), CENATAV/Cuba (visites, programme de doctorats en co-tutelle)

Séjours longs dans des laboratoires : Drexel University, USA (É. Sanjuan, 1 an, 2008/2009), University of Cambridge, UK (F. Lefèvre, 1 an, 2009/2010), Université de Mexico (JM Torres-Moreno, 2010/2011).

Séjours courts de chercheurs permanents : Universidad Autónoma Metropolitana (Mexique, Juan Manuel Torres-Moreno), USP Brasil & Universidad de Barcelona (I. da Cunha), Université de Maribor (Slovénie, G. Linarès), Université de Dalat (Vietnam, Patrice Bellot), African HTL society Juan-Manuel Torres Moreno, Pascal, Nocera), Université Pompéu Fabra (J.M. Torres-Moreno), CENTATAV / Cuba (J.F. Bonastre).

Accueil de chercheurs étrangers (seniors) : M^a Teresa Cabré, IULA (avril 2010), Horacio Saggion Univ. of Sheffield (décembre 2009) Jorge Vivaldi, IULA (juillet 2008), Gerardo Sierra, Universidad Nacional Autónoma de México (juillet 2008), Guy Lapalme, Univ. de Montréal (2005 & 2007), Thiago Pardo, USP Brésil (mars 2010) - RST, Brendan Baker (QUT, Australie), M. Wagner (University of

Cambara, Australie).

1.1.1.5 Thèmes de recherche

Traitement automatique de la langue naturelle écrite

La sous-thématique Traitement automatique de la langue naturelle écrite (TALNE) fédère un ensemble de sujets de recherche sur le traitement de l'information textuelle.

Extraction d'information

Questions Réponses (QR) et Recherche d'Information (RI). Nous avons élaboré et testé des systèmes QR/RI dans différentes campagnes d'évaluation. Nous avons développé des composants logiciels optimisés pour la RI dans les grandes BD textuelles et ils ont été testés dans les campagnes NIST TREC Enterprise'08/TREC ou Entity'10. Les composants du moteur QR développé (classification de questions, étiquetage EN, recherche documentaire, segmentation thématique, extracteur de réponses,...), ont été testés à CLEF'06. 2 thèses ont été soutenues sur ce sujet (L. Gillard et C. Vidrequin). Lors de notre participation à la campagne ESTER2 EN, nous avons obtenu les meilleurs résultats pour 3 tâches sur 4, ce qui a placé le système du LIA en première place. La thèse de B. Favre soutenue en 2007 était centrée sur l'application et l'adaptation des méthodes de RI aux données orales. La sous-thématique a participé (2007-9) au projet ANR PIITHIE sur la détection de plagiat et suivi informationnel. Ce projet a donné lieu à des travaux sur l'identification de citations, la détection de similarités (copies verbatim), la segmentation thématique et la structuration automatique de documents. Ils ont conduit au développement de services Web et des applications réelles (Advestigo). Citons aussi l'étude de la prise en compte de meta-informations dans l'indexation, l'utilisation de connaissances disponibles ou extraites automatiquement et leur enrichissement automatique, thèse soutenue en 2009 (C. Vidrequin), ou l'adaptation des méthodes pour tenir compte du handicap langagier (dyslexie) d'un utilisateur, thèse soutenue en 2007 (L. Sitbon).

À partir de 2008, nous avons étendu ces recherches au cadre des campagnes INEX (documents structurés XML) dont nous avons co-organisé la tâche Question/réponse pour l'édition 2009. Nous menons aussi des travaux sur les modèles mathématiques qui ont abouti à la proposition d'une logique multivaluée pour l'IR (Discrete App. Math.) unifiant approches booléennes, floues et probabilistes.

Assistance d'utilisateur en recherche d'informations (AU). Des recherches autour des pathologies de l'écrit ont été menées dans le cadre d'une thèse co-encadrée (financement STIC/SHS/CNRS, Parole & Langage, Aix-Marseille-I). L'activité s'est d'abord concentrée sur la caractérisation de pathologies telles que la dyslexie grâce à des expériences menées avec des psycholinguistes. De celles-ci ont été dérivés des critères évaluant les capacités de lecture et d'écriture d'un utilisateur mais aussi la complexité intrinsèque d'un document. Ces analyses ont été intégrées dans le processus de RI grâce à une assistance ad hoc de l'utilisateur, des méthodes d'apprentissage automatique et la proposition de mesures de similarité intégrant des critères non uniquement informationnels (TAL'07 et Document Numérique'10, SIGIR-AND'08, LREC'08, Interspeech'07, Coria'08).

Fouille et Analyse de Documents

Classification, segmentation thématique et Détection d'Opinion. Le LIA a participé au projet Technolanguage Agile/Oural sur la mise à disposition d'outils de traitement linguistique (segmentation thématique de textes). Les dernières recherches ont permis de proposer une approche nouvelle à base de chaînes lexicales pondérées à champ variable. À la suite des expériences concluantes menées en 2005, nous avons participé à DEFT'07 (Détection d'opinion) et 08. Les méthodes proposées par le LIA (chaînes de Markov, modélisation probabiliste...) nous ont permis d'obtenir la 1^{ère} place dans la compétition. Lors de DEFT'09, nous avons été invités à présenter un algorithme de vote combinant plusieurs méthodes de classification. Généralement, nous avons étudié la possibilité de ne pas dissocier la classification des autres problématiques traitées par les différents composants des applications de l'équipe. La plupart des problèmes de segmentation, de RI et de routage pouvant être vus sous l'angle de la classification de textes. En outre, nous avons proposé des adaptations au domaine des ressources humaines : comparaison entre offres d'emploi et candidatures, segmentation automatique de CV et d'offres (thèse de R. Kessler soutenue en 2009). En Détection d'Opinion, une thèse a démarré (R. Lavalley, convention Cifre avec EDF), fin 2008.

Résumé automatique de documents (RA). Abordé majoritairement via une approche numérique, avec l'introduction de nouvelles heuristiques (multilangue, multidoc) et la génération de résumés personnalisés, le système Cortex est devenu multidocument (thèse de F. Boudin). Ceci a servi pour développer un algorithme de détection de nouveauté (RA multi-doc/guidé par une thématique). Nous avons participé aux campagnes NIST DUC 06-07/TAC. Pour la campagne DUC, le couplage de systèmes de RA (SRA) plus un algorithme sophistiqué de vote a permis de nous placer dans le peloton de tête des 5 premières équipes sur 35 participants. Lors de TAC'08, nous avons développé un système pionnier de détection de la nouveauté qui s'est bien positionné. Une HDR (J-M. Torres) et une thèse (F. Boudin) ont été soutenues en 2007-8 sur ce thème.

Le projet ANR RPM2 (Résumé Pluri-média Multi-document) nous a permis de développer ce thème et de renforcer la collaboration entre écrit/oral. Une collaboration (2007) avec l'IULA Barcelone a permis de combiner les approches linguistiques et numériques. En 2007-9, nous avons développé un SRA hybride intégrant systèmes numériques (Cortex/Enertex) et symbolique (Disicosum).

SummTerm est un SRA pour des textes spécialisés (extracteur terminologique YATE+EuroWordNet) ; articles (LREC'10 et journal Terminàlia). Avec l'UAM-Mexico, nous avons développé REG, système de RA basé sur la théorie des graphes (JADT'10 et African HLT'10). Nous avons revisité une approche pour le RA fondée sur la théorie des graphes dans TALN'08-09 et SFC'09. Nous avons développé des méthodes pour la génération de RA combinant la compression de phrases. Une thèse (A. Molina) démarrée fin 2009 étudie ces algorithmes pour augmenter l'information dans un espace réduit. Les résultats préliminaires sont très encourageants (Aesla'10).

Analyse discursive automatique et génération de texte (GT). Dans la thèse de da Cunha (2008), l'analyse du discours dans le RA a été implantée partiellement (faute d'analyseurs discursifs pour l'espagnol). Nous avons développé un analyseur pour cette langue. En collaboration avec le NILC (Sao Paulo), nous réalisons l'adaptation de l'analyseur discursif pour le français. D'autres articles sur l'analyse discursive ont été publiés (STIL'09 et Discourse Studies). Nous avons effectué des recherches en GT par réécriture de phrases (thèse d'E. Charton).

Multilinguisme, ressources linguistiques et terminologie. Le LIA a été porteur du projet Carmel ce qui a donné lieu à 2 publications TALC/TALN'06, dans la période évaluée. Nous avons développé des méthodes pour l'enrichissement des corpus permettant l'annotation linguistique (syntaxique, morphologique, sémantique, pragmatique, etc.). Ces corpus sont utilisés pour entraîner les modèles employés dans les autres thèmes de recherche. Avec Lyon 3 et Drexel (Philadelphie), nous travaillons à l'analyse de la littérature scientifique par extraction, classification et visualisation de graphes terminologiques avec TermWatch (IEEE-VAST' 06, ECIR-ESAIR'08, ACL-CIKM'08, ACL-JCDL'09). Ces travaux sont étendus à la génération d'hypothèses à partir de la littérature scientifique (JADT'10).

Activités en Dialogue

Les activités de recherche de l'équipe Modèles de Langage Appliqués au Dialogue (ci-après désignée uniquement par Dialogue) se concentrent principalement sur les modèles de langage et de compréhension dans les applications de dialogue téléphonique homme-machine, ainsi que sur la fouille de données audio (extraction d'entités nommées, d'opinions...). Le type particulier d'interaction impliquée par le dialogue homme-machine permet d'une part d'étudier la langue naturelle orale spontanée et d'autre part de se situer dans un cadre applicatif réaliste.

Ainsi les activités du LIA autour des systèmes de dialogue continuent de s'inscrire dans un double objectif :

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

L'équipe a commencé à élargir ses activités depuis septembre 2010, notamment à la traduction automatique de la parole, en s'inscrivant dans une logique de développement qui est explicitée dans la partie projet du rapport.

Principaux sujets d'études

Stratégie de décodage et étude diagnostique des sorties de reconnaissances de parole

Dans le cadre du projet Européen DIVINES, l'équipe Dialogue était engagée dans l'étude diagnostique des sorties de reconnaissances de parole. Ce projet Européen a financé la thèse de Loïc Barrault, soutenue en 2008.

Extraction d'entités dans des messages vocaux

Ce travail sur l'extraction de "type d'appels" et sur la détection d'opinions à partir de corpus de sondages téléphoniques a été réalisé, à travers une convention de recherche avec France Télécom R&D (2004-7). Cette convention a financé la thèse de Nathalie Camelin, soutenue en 2007.

Compréhension de la parole spontanée dans un cadre de dialogue homme-machine

Le cœur des activités de la thématique Dialogue du LIA se situe autour du problème de la compréhension de la parole. Après avoir proposé des modèles de représentation des connaissances de base du processus de compréhension et des algorithmes de décodage conceptuel intégrés au processus de reconnaissance automatique de parole, nous nous sommes préoccupés de développer des modèles de compréhension plus riches permettant d'avoir une représentation sémantique complète d'un message, dans un cadre de dialogue homme-machine.

Ces travaux ont été menés dans le cadre du projet Européen LUNA (spoken Language UNderstanding in multilingAl communication systems, <http://www.ist-luna.eu/>), démarré en septembre 2006 et achevé en novembre 2009. Deux étudiants en thèse, Marie-Jean Meurs (soutenue en 2009) et F. Duvert (soutenue en 2010), ont été financés sur ce sujet depuis septembre 2006.

Un autre aspect très important des recherches sur les systèmes de dialogue oraux est la nécessité de disposer de corpus. Dans cette perspective, le LIA est coordinateur du projet ANR PortMedia impliquant 3 groupes de recherche (LIUM, LIG et LORIA) et un partenaire industriel (ELDA) dont l'objectif est de collecter de nouvelles données de dialogue et de les compléter avec des méta-informations pertinentes (transcriptions, annotations sémantiques, résultats d'évaluation...). Ce projet a démarré en mars 2009 et dure 36 mois.

Le travail du consortium durant les deux premières années a porté principalement sur les spécifications nécessaires pour le projet (définition de la tâche, collecte des données, annotations, outils d'annotations, évaluations) et la collecte des données. La thèse de Bassam Jabaian, lancée dans le cadre de ce projet en septembre 2009 en co-direction avec L. Besacier (LIG) porte sur l'étude des interactions entre les processus automatiques de traduction et de compréhension de la parole afin d'améliorer les deux aspects et de favoriser le développement de systèmes de compréhension multilingue.

Gestion du dialogue

Une problématique nouvelle a été abordée depuis 2007. Il s'agit de l'étude des modèles stochastiques pour la gestion du dialogue, constituant une alternative aux modèles classiques basés sur des automates fixes de dialogue. Cette étude a débuté dans le cadre du projet LUNA, à partir des corpus de dialogue collectés lors du projet, et dans le sous-projet WP4 dédié à l'ajout de connaissances contextuelles relatives à l'état du dialogue dans le processus de compréhension d'un message. La thèse de Florian Pinault, démarrée en novembre 2007 sur financement ministériel, est consacrée à ce thème. Par ailleurs, l'équipe a initié une collaboration avec le groupe du Prof. Steve Young de l'Université de Cambridge (Dialogue Systems Group du Machine intelligence Lab, Engineering Dpt). Après une visite d'un an dans le groupe d'un membre du LIA en 2009, la collaboration s'est poursuivie, notamment autour du développement d'un système de dialogue multilingue.

Apprentissage automatique et extraction d'informations à partir de corpus de dialogues

Un frein principal au développement d'une nouvelle application de serveur vocal interactif est le coût lié à la construction et à la maintenance des modèles de langage et de dialogue permettant de gérer l'interaction entre l'utilisateur et le serveur. Il est alors utile d'étudier des processus d'apprentissage automatique permettant d'automatiser tout ou partie de cette phase de développement et de maintenance de modèles de langage et de dialogue. La thèse de Christophe Servan, portant sur ce thème, financée par le conseil général, a été soutenue en 2008.

En complément des activités portant sur l'analyse de données et l'extraction d'informations appliquées à des corpus de dialogue oraux (Speech Mining) sont en cours. Ce travail s'est inscrit dans le cadre du projet ANR Masse De Données – Connaissances Ambiantes (projet EPAC, 2006-9), centré sur l'analyse de la parole conversationnelle dans des corpus radiophoniques (débat, interviews). Ce projet implique à la fois des chercheurs du LIA actifs en Reconnaissance de la Parole pour la segmentation et la caractérisation en locuteurs des corpus, et des chercheurs de la thématique

Dialogue pour la caractérisation de l'interaction. La thèse de Pierre Gotab, financée par une bourse Région/Entreprise avec Semantia depuis 2008, s'inscrit aussi dans ce cadre.

Activité en Traitement Automatique de la Parole

Les travaux du LIA en traitement automatique de la parole concernent essentiellement les problèmes d'extraction et d'analyse de l'information linguistique et extra-linguistique du signal de parole. Ces dernières années, l'activité s'est consolidée autour des thèmes centraux initiés auparavant (identification du locuteur, transcription enrichie, indexation audio), s'est étendue à des domaines connexes et s'est ouverte à une évolution plus générale de l'environnement des technologies vocales vers le multimédia, le multilinguisme, les grandes bases audiovisuelles et le WEB. Dans tous ces domaines, nous avons travaillé à la fois à la recherche de paradigmes alternatifs, à l'optimisation ou au croisement des technologies et aux interactions des composants de traitement de la parole avec leur environnement applicatif.

*En **reconnaissance du locuteur**, nous avons conforté notre positionnement international en contribuant significativement aux avancées scientifiques récentes dans le domaine, par exemple sur le rôle structurant du modèle UBM et les modèles de décomposition par analyse factorielle (thèse de N. Scheffer, soutenue en 2006), sur l'adaptation, en ligne des modèles (thèse d' A. Preti, soutenue en 2006), la biométrie multimodale (thèse de A. Larcher, projets ANR BioBimo, européen FP7 Mobio). Plus récemment, nous avons entamé une collaboration avec le LIG (Université J. Fourier, Grenoble) sur l'introduction de connaissances phonétiques dans les systèmes, et, réciproquement, sur les systèmes comme outils d'interprétation phonétique et phonologique (thèse de J. Kahn).*

*L'**identification des langues** est un thème qui a fait, lui aussi, l'objet d'un certain nombre d'études récentes dans le cadre de la thèse de F. Verdet, (co-tutelle avec l'Université de Fribourg, Suisse). Ces travaux ont permis une participation aux évaluations NIST LRE 2009. En parallèle, le développement de la plateforme ALIZE s'est poursuivi (ANR MISTRAL). Elle est devenue un standard du domaine, avec plus de 30 laboratoires et industriels contributeurs. Elle a été le support de plusieurs soumissions aux évaluations NIST ces dernières années.*

*Nos travaux dans ce domaine se sont étendus aux problèmes d'**analyse de la voix et parole pathologique** (projet ANR blanc DesphoApady, projet VML, thèse de G. Pouchoulin), avec des collaborations largement interdisciplinaires (phonétique clinique, psycholinguistique, phonologie,...) sur le plan national (Laboratoire Parole et Langage d'Aix-en-Provence, Laboratoire de Phonétique et Phonologie de Paris) et international (Université Polytechnique de Madrid).*

*En **transcription automatique**, les recherches sur la technologie "cœur" des systèmes se sont poursuivies à plusieurs niveaux, tirant profit de l'investissement lourd qui a été réalisé les années précédentes pour le développement de la plate-forme de transcription enrichie SPEERAL : modélisation acoustique (modèles compacts, discriminants, thèses CIFRE de C. Lévy et de M. Bouallègue), stratégies de décodage (projet ANR Blanc ASH, thèse de B. Lecouteux sur la combinaison de systèmes, systèmes temps réel, décodage interactif (thèse de G. Senay), sur l'introduction d'information haut niveau dans le moteur de reconnaissance de la parole, l'utilisation du Web à la fois comme ressource et comme objet d'étude (thèses de S. Oger et M. Rouvier). Nous avons développé une forte activité sur le traitement des langues peu dotées (thèses de Q. Nguyen et A. Nimaan), tonales et agglutinatives. Ces travaux ont été réalisés en collaboration avec l'institut Polytechnique de Hanoï et l'Institut des nouvelles Technologies de Djibouti ; ils trouvent leur prolongement dans le cadre du projet ANR BLANC PI.*

*Nos travaux en **indexation audio** se sont concentrés sur le traitement des données issues du WEB et des grandes bases numériques : modèles de langage Web (thèse de S. Oger), identification du genre vidéo (thèse de M. Rouvier), résumé automatique (thèse de B. Favre), segmentation de vidéos Web (thèse de P. Clément). Ces travaux ont souvent été réalisés dans le cadre de projets collaboratifs (projets ANR Sigmund, RPM2, Biobimo, AVISON). Enfin, des travaux sur l'intégration de la reconnaissance et la traduction ont été amorcés (thèse de R. Rubino, co-encadrée avec P. Langlais, RALI, Université de Montréal).*

1.1.2 Bilan synthétique

Depuis 2006, 6 HDR et 23 thèses ont été soutenues dans la thématique Langage. 18 thèses sont en cours. 8 des 12 enseignants-chercheurs de la thématique sont titulaires de la PEDR ou de la PES.

Le volume et la qualité de la production scientifique a significativement évolué ces dernières années : le nombre de publications a progressé dans toutes les catégories (cf. Tab. T1.3), alors que l'effectif des permanents est resté stable. Nos étudiants ont obtenu des prix pour la qualité de leurs articles (N. Scheffer : IBM Best paper Award, Odyssey 2008 ; M. MacLaren : Best Student paper Award, Interspeech 2008 ; Bonastre & Matrouf : Best Scientific Paper Award, International Association for Pattern Recognition).

Cette progression accompagne le développement des travaux transversaux aux 3 équipes du précédent contrat (écrit, oral et dialogue) qui se sont regroupées. Ce regroupement dans un ensemble cohérent nous permet maintenant de mettre en avant la couverture et la densité des sujets maîtrisés par le groupe. Il mène à un enrichissement mutuel qui est scientifiquement productif et qui nous place dans une position très favorable pour la réponse aux appels à projets nationaux et européens (18 projets ANR et 5 projets européens dans cette période). Ce dynamisme s'appuie sur des plate-formes logicielles largement diffusées vers des partenaires académiques ou industriels.

	ACL	ACLN	INV	ACTI	ACTN	OS+DO
2002-2006	17	3	5	104	45	6
2006-2010	44	6	14	213	99	19

Table T1.3 : Évolution des publications de la Thématique Langage selon les catégories AERES

Outre l'activité de publication, notre visibilité internationale s'intensifie avec l'organisation de conférences pour la période passée et celle à venir (le LIA sera l'organisateur d'EACL 2012 et co-organisateur d'Interspeech 2013, de CORIA 2011). Cela s'explique aussi par le développement de collaborations et de séjours/invitations dans des laboratoires reconnus et par des participations aux conseils d'administration de sociétés savantes majeures du domaine : 2 membres au Speech and Language Technology Comitee d'IEEE (F. Béchet et J.F. Bonastre), un membre IEEE Fellow (R. De Mori), co-présidence de l'ISCA (International society of Speech Communication Association), dont J.F. Bonastre sera le président à partir de Sept. 2011. Le laboratoire a aussi été présent dans les conseils d'administration de l'AFCP et de l'ATALA.

Par ailleurs, nous contribuons significativement à l'administration de l'enseignement et de la recherche de l'établissement : responsabilité pédagogique de la licence d'informatique depuis 2008, du Master depuis 2006, direction du LIA, vice-présidence du conseil d'administration de l'UAPV depuis 2008, co-direction de la Structure Fédérative de Recherche Agorantic, membres élus au CS et au CEVU...

1.2 Autoévaluation

Le bilan que nous avons présenté fait apparaître les forces et les faiblesses de la thématique ³ :

- **Forces :**

- **couverture/densité des compétences en traitement automatique** par des méthodes statistiques : une grande partie des sujets liés à l'extraction des contenus, à leur caractérisation et à leur interprétation est abordée. Cette densité repose sur une complémentarité de compétences qui nous semble un point fort de la thématique.
- **activité contractuelle et niveau de publication** sont soutenus et en très forte progression, que ce soit en termes de revue internationales référencées ou dans les grandes conférences qui sont les tribunes des communautés scientifiques concernées.
- les communautés scientifiques du TAL sont larges mais notre **visibilité internationale** sur des sujets spécialisés est très bonne (identification du locuteur, résumé automatique, dialogue, indexation audio). Elle est soutenue par un engagement régulier dans l'animation scientifique (organisation d'EACL 2012, présidence de l'ISCA) de nombreuses visites et séjours de ses membres dans des laboratoires étrangers (Cambridge, Dextrel Univ., Univ. Mexico).

- **Faiblesses :**

³Le volet risques/opportunités de l'analyse SWOT étant liée au projet, il a été inclus dans la partie Projet du dossier (section spécifique à la thématique).

- *le manque de support : le TAL est un domaine où la **logistique expérimentale peut être très lourde** ; nos besoins sont forts mais le manque de personnels IATOS limite notre capacité de transfert/valorisation des résultats de notre recherche.*
- *le **départ de « cadres »** pour raison de promotion ou de retraite expose la thématique à une certaine instabilité. L'organisation de groupes de recherche, la complémentarité des compétences qui se développent, les liens que les chercheurs seniors tissent dans leur communauté scientifique sont généralement portés par l'activité de l'ensemble de l'équipe ; le départ de cadres affecte cet équilibre.*
- *la thématique est très impliquée dans l'administration de la recherche et de l'enseignement (vice présidence du CA de l'UAPV, direction laboratoire, de la SFR, responsabilité du Master, responsabilité des études) : ces **charges administratives** sont utiles à la collectivité, mais elles ne sont compensées que par un engagement individuel très fort ; cette situation fragilise la thématique.*

1.2.1 Liste des Publications

ACRONYMES

- CORIA** : COnférence en Recherche d Informations et Applications
JADT : Journées internationales d'Analyse statistique des Données
 Textuelles/International Conference on the Statistical Analysis of Textual Data
JEP : Journées d'Etude sur la Parole
LNCS : Lecture Notes in Computer Science
TALN : Traitement Automatique des Langues Naturelles

DO (Direction d'ouvrages)

- T1.495** P. Bellot, *Recherche d'information contextuelle, assistée et personnalisée*, 2011, Hermès (collection Recherche d'Information et Web)
- T1.494** Agata Jackiewicz, Susan Hunston, M. El Bèze, Co-direction du n° spécial "Opinions, sentiments et jugements d'évaluation", 2010, *Revue TAL* Vol 51/3, pp.7-17, 2010
- T1.493** P. Bellot, *Vers une recherche d'information contextuelle, assistée et personnalisée*, Hermès Science Publications – 2010
- T1.492** K Berkling, J.F. Bonastre and Joe Campbell, *Special Section on Speaker and Language Recognition*, 2007, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 15, Issue 7, September

OS (Ouvrages Scientifiques, ou chapitres de ces ouvrages)

- T1.484** Juan-Manuel Torres-Moreno. *Résumé automatique de documents – Une approche statistique*. Hermès Lavoisier, 2011, ISBN 978-2-7462-3212-9
- T1.483** Juan-Manuel Torres-Moreno, Marc El-Bèze, Patrice Bellot, Frédéric Béchet, *Peut on voir la détection d'opinions comme un problème de classification thématique ? Modèles statistiques pour l'accès à l'information textuelle*, Eric Gausier, François Yvon (eds.), 2011. Hermès Lavoisier, ISBN 978-2746224971
- T1.482** P. Bellot, *Vers une prise en compte de certains handicaps langagiers dans les processus de recherche d'information*, 2011, in "Vers une recherche d'information contextuelle, assistée et personnalisée" sous la direction de P. Bellot, chapitre 7, Hermes, 2011.
- T1.481** R. Fraile, N. Saenz-Lechon, J. I. Godino-Llorente, V. Osma-Ruiz, C. Fredouille, *Effect of a Simulated Analogue Telephone Channel on the Performance of a Remote Automatic System for the Detection of Pathologies in Voice Impact of Linear Distortions on Cepstrum-based Assessment: Band Limitation, Frequency Response and Additive Noise*, 2010, Springer Computer Science book, A. Fred, J. Filipe, and H. Gamboa (Eds.): BIOSTEC 2009, CCIS 52.
- T1.480** A. El Hannani, J-F. Bonastre, et al., *Text-independent speaker recognition*, in *Biometric Reference Systems and Evaluation Framework*, 2009, in « Text-independent speaker recognition, in Biometric Reference Systems and Evaluation Framework », Springer, ISBN: 978-1-84800-291-3
- T1.479** D. Matrouf, J-F. Bonastre, *Session Effects on Speaker modelling*, 2009, in *Encyclopedia of Biometrics*, Springer, ISBN: 978-0-387-73003-5
- T1.478** F. Boudin, J-M. Torres, *A Maximization-Minimization Approach for Update Summarization*, 2009, Book chapter in *Current Issues in Linguistic Theory: Recent Advances in Natural Language Processing V*,
- T1.477** J-F. Bonastre, D. Matrouf, *Speaker matching*, 2009, in *Encyclopedia of Biometrics*, Springer, ISBN: 978-0-387-73003-5
- T1.475** J-F. Bonastre, et al., *Évaluation des Systèmes de Transcription enrichie d'Émissions Radiophoniques*, 2008, in *L'évaluation des technologies de traitement de la langue : les campagnes*

Technolanguage, Chapitre 7, Hermes/Lavoisier, (Traité IC2, série Cognition et traitement de l'information)

T1.474 H. Bonneau-Maynard, A. Denis, F. Béchet, L. Devillers, F. Lefèvre, M. Quignard, S. Rosset, J. Villaneau, Chapitre 9. MEDIA : évaluation de la compréhension dans les systèmes de dialogue, 2008, L'évaluation des technologies de traitement de la langue : les campagnes Technolanguage, Hermes Science Publishing Ltd

T1.473 P. Bellot, M. Boughanem, Recherche d'information et systèmes de questions-réponses, 2008 in " La recherche d'informations précises : traitement automatique de la langue, apprentissage et connaissances pour les systèmes de question-réponse (Traité IC2, série Informatique et systèmes d'information)", sous la direction de B.Grau, Hermès-Lavoisier, chapitre 1, p. 5-35

T1.472 C. Lévy, G. Linarès, P. Nocera, J-F. Bonastre, Embedded mobile phone digit-recognition, 2007, Chapter 7 in Digital Signal Processing for In-Vehicle and Mobile Systems 2. Springer Science, H. Abut, J.H.L. Hansen and K. Takeda edition, 2007

T1.471 Gillard, L. Sitbon, E. Blaudez, P. Bellot, M. El-Bèze, Relevance Measures for Question Answering, The LIA at QA@CLEF-2006, 2007, LNCS, 4730/2007 " Evaluation of Multilingual and Multi-modal Information Retrieval", pp.440-449.

T1.470 M. El-Bèze, Systèmes de questions-réponses, 2006, in Compréhension des Langues et interaction, ouvrage collectif sous la direction de Gérard Sabah, Chapitre 10, pp. 277-297, Hermès, Lavoisier.

ACL (Articles dans des revues nationales ou internationales référencées)

T1.461 Iria da Cunha, Eric SanJuan, Juan Manuel Torres-Moreno, Marina Lloberes, Irene Castellon, DiSeg 1.0 : The First System for Spanish Discourse Segmentation, Expert Systems With Applications, 2011, DOI : 10.1016/j.eswa.2011.06.058.

T1.460 M. McLaren, D. Matrouf, R. Vogt, J-F. Bonastre, Applying SVMs and Weight-based Factor Analysis to Unsupervised Adaptation for Speaker Verification, **2011**, Computer Speech & Language, Volume 25, Issue 2, April 2011, Pages 327-340

T1.459 E. San-Juan, P. Bellot, Overview of the INEX 2010 Question Answering Track (QA@INEX), **2011**, Lecture Notes in Computer Science LNCS - Elsevier (à paraître)

T1.458 N. Evans, S. Bozonnet, D. Wang, C. Fredouille, R. Troncy, A comparative study of bottom-up and top-down approaches to speaker diarization, **2011**, Transactions on Audio, Speech and Language Processing, special issue on "New Frontiers in Rich Transcription" (TASLP)

T1.457 D. Matrouf, F. Verdet, M. Rouvier, J-F. Bonastre, G. Linarès, Modeling Nuisance Variabilities with Factor Analysis for GMM-based Audio Pattern Classification, 2011, Computer Speech and Language, vol. 25, Issue 3, pp 481-498

T1.456 R. Lavalley, Clavel Chloé, P. Bellot, Extraction probabiliste de chaînes de mots relatives à une opinion, 2011, Traitement Automatique des Langues (TAL), p. 101 à 130, vol. 51-3

T1.455 X. Anguera, S. Bozonnet, N. W. D. Evans, C. Fredouille, O. Friedland, O. Vinyals, Speaker diarization : A review of recent research, 2011, Accepted for publication in "IEEE Transactions On Acoustics Speech and Language Processing" (TASLP), special issue on "New Frontiers in Rich Transcription"

T1.454 A. Ghio, Pouchoulin Gilles, B. Teston, S. Pinto, C. Fredouille, C. De Looze, D. Robert, F. Viallet, A. Giovanni, How to manage sound, physiological and clinical data of 2500 dysphonic and dysarthric speakers?", 2011, Accepted for publication in Speech Communication, Special issue on Advanced Voice function Assessment

T1.453 B. Lecouteux, G. Linarès, S. Oger, Integrating imperfect transcripts into speech recognition systems for building high-quality corpora", 2011, Computer Speech and Language (CSL)

T1.451 S. Hahn, M. Dinarelli, C. Raymond, F. Lefèvre, P. Lehen, R. De Mori, H. Ney, G. Riccardi "Comparing Stochastic Approaches to Spoken Language Understanding in Multiple Languages", 2010 IEEE Transactions on Audio, Speech and Language Processing, 14 pages, Volume 19, Issue 6, pp 1569-1583, Nov. 2010.

T1.450 Sitbon Laurianne, P. Bellot, Philippe Blache, Vers une recherche d'informations adaptée aux capacités de lecture des utilisateurs - Recherche d'informations et résumé automatique pour des

personnes dyslexiques", 2010, *Revue des Sciences et Technologies de l'Information, série Document numérique*, volume 13, 1-2010, p. 161-186

T1.449 J-M. Torres, Automatic grouping in families and Lexematization language independent, 2010, *Inteligencia Artificial*, 14(47):38-53.

T1.447 Nathalie Camelin, Frédéric Béchet, Géraldine Damnati, Renato De Mori Interpretation of opinion expressions in spoken surveys, 2010, *IEEE Transactions on Speech and Audio Processing*, vol., 18, no. 2, pp. 369–381

T1.446 M. McLaren, D. Matrouf, R. Vogt, J-F. Bonastre, Applying SVMs and Weight-based Factor Analysis to Unsupervised Adaptation for Speaker Verification, 2010, *Computer Speech & Language*

T1.445 E. Charton, J-M. Torres, Modélisation automatique de connecteurs logiques par analyse statistique du contexte, 2010, *Canadian Journal of Information and Library Science (in print)*

T1.444 V. Popescu, J. Caelen, C. Burileanu, A Constraint Satisfaction Approach to Context-Sensitive Utterance Generation in Multi-Party Dialogue Systems, *International Journal of Speech Technology* 12(2-3):95-112, 2009

T1.443 C. Fredouille, G. Pouchoulin, A. Ghio, J. Revis, J-F. Bonastre, A. Giovanni, Back-and-forth methodology for objective voice quality assessment: from/to expert knowledge to/from automatic classification of dysphonia, 2009, (on-line: <http://www.hindawi.com/journals/asp/2009/982102.html>) *Eurasip Journal on Advances in Signal Processing*, vol. 2009, Article ID 982102

T1.442 R. Fraile, N. Saenz-Lechon, J.I. Godino-Llorente, V. Osma-Ruiz, C. Fredouille, Automatic detection of laryngeal pathologies in records of sustained vowels by means of MFCC parameters and differentiation of patients by sex, 2009, *Folia Phoniatica et Logopaedia, International Journal of Phoniatrics, Speech Therapy and Communication Pathology, Special issue: COST Action 2103 - A Joint European Project for Advanced Voice Assessment, Vol. 61, N°3*

T1.441 M. Rouvier, G. Linarès, B. Lecouteux, Query driven strategy for on-the-fly term spotting in spontaneous speech, 2009, *Special Issue on Scalable Audio Content Analysis, EURASIP Journal on Audio, Speech, and Music Processing, Volume 2010 (2010), Article ID 326578, doi:10.1155/2010/326578*

T1.440 C. Lévy, G. Linarès and J. Bonastre, Compact acoustic models for embedded speech recognition, *EURASIP Journal on Audio, Speech, and Music Processing, Volume 2009 (2009), Article ID 806186, doi:10.1155/2009/806186*

T1.439 Campbell J.P, Shen W., Campbell W.M., Schwartz R., J-F. Bonastre, D. Matrouf, Forensic Speaker Recognition: A need for caution, 2009, *IEEE Signal Processing Magazine*, March 2009, vol 26 pp 95-103

T1.438 E. SanJuan, Heyting Algebras with Boolean Operators for Rough Sets and Information Retrieval applications, 2008, *Discrete Applied Mathematics*, 156(6), pp.967-983.

T1.437 L. Sitbon, P. Bellot, P. Blache, Éléments pour adapter les systèmes de recherche d'information aux dyslexiques, 2008, *Traitement Automatique des Langues (TAL)*, n° 48-2, 2008

T1.436 P. Zweigenbaum, B. Grau, A.-L. Ligozat, I. Robba, S. Rosset, X. Tannier, A. Vilnat, P. Bellot, Apports de la linguistique dans les systèmes de recherche d'informations précises, 2008, *RFLA (Revue Française de Linguistique Appliquée)*, XIII (1),

T1.435 T. Bazillon, V. Jousse, F. Béchet, Y. Estève, G. Linarès, D. Luzzati, La parole spontanée: transcription et traitement, 2008, *Revue Traitement Automatique des Langues - TAL- volume 49*

T1.434 Marie-Laure Guenot, Penser global, agir local : Les coordinations antithétiques dans une grammaire de construction multidimensionnelle, 2008, *Traitement Automatique des Langues*, 49 (1) pp.89-113.

T1.433 A. Bürki, C. Gendrot, G. Gravier, G. Linarès, C. Fougeron, Alignement automatique et analyse phonétique: comparaison de différents systèmes pour l'analyse du schwa, 2008, *Revue Traitement Automatique des Langues - TAL - volume 49 numéro 3 – pp. 672-708*

T1.432 R. De Mori, F. Béchet, F. Hakkani-Tur, D. McTear, M. Riccardi, G. Tur, Spoken language understanding, 2008, *IEEE Signal Processing Magazine*, 25 (3) pp.50-58.

T1.431 R. Gemello, F. Mana, S. Scanzio, Pietro Laface, R. De Mori, Linear hidden transformations for adaptation of hybrid ANN/HMM models, 2007, *Speech Communication* 49 (2007) 827–835

T1.430 A. Berry, M. Pouzet, E. SanJuan, special volume on knowledge discovery and discrete mathematics and a tribute to the memory of Peter L. Hammer, 2007, *Journal Annals of Mathematics and Artificial Intelligence (AMAI) Volume 49, Numbers 1-4*, pp. 1-312

T1.429 A. Berry, E. SanJuan, M. Pouzet, M. C. Golumbic, Introduction to the special volume on knowledge discovery and discrete mathematics and a tribute to the memory of Peter L. Hammer, 2007, *Annals of Mathematics and Artificial Intelligence, Volume 49, Numbers 1-4 / April, 2007*, pp. 1-4

T1.428 B. Fauve, D. Matrouf, N. Scheffer, J-F. Bonastre, J.S.D. Mason, State-of-the-Art Performance in Text-Independent Speaker Verification Through Open-Source Software, 2007, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 15, Issue 7, pp 1960-1968, September

T1.427 C. Raymond, F. Béchet, N. Camelin, R. De Mori, G. Damnati, Sequential decision strategies for machine interpretation of speech, 2007, *IEEE transactions on audio, speech and language processing vol. 15, no. 1*, pp. 162-171

T1.426 M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, C. Wellekens, Automatic speech recognition and speech variability: A review, 2007, *Speech Communication 49 (2007) 763–786*

T1.425 I. Da Cunha Fanego, Wanner, L., Cabré, M.T., Summarization of specialized discourse: The case of medical articles in Spanish, 2007, *Terminology 13 (2)*. Amsterdam: John Benjamins. 249-286. ISSN 0929-9971.

T1.424 R. Gemello, F. Mana, D. Albesano, R. De Mori, Multiple resolution analysis for robust automatic speech recognition, 2006, *Computer Speech and Language, 20(1)*, pp. 2-21

T1.423 R. Gemello, F. Mana, R. De Mori, Automatic speech recognition with a modified ephraim-malah rule, 2006, *IEEE Signal Processing Letters, 13(1)*:p. 52-55, January 2006

T1.422 C. Raymond, F. Béchet, R. De Mori, G. Damnati, On the use of finite state transducers for semantic interpretation, 2006, *Speech Communication, Special Issue on Spoken Language Understanding, Elsevier, 48(3)*: 288-304, March 2006.

T1.421 S. Meignier, D. Moraru, C. Fredouille, J-F. Bonastre, L. Besacier, Step-by-step and integrated approaches in Broadcast news speaker diarization, *Computer Speech and Language (CSL), Vol. 20/2-3 pp. 303-330, April/July 2006*

T1.420 Gillard, L. Sitbon, P. Bellot, M. El-Bèze, Dernières évolutions de SQuALIA, le système de Questions/Réponses du LIA, 2006, *Traitement Automatique des Langues (TAL), Hermès, vol. 46-3, p. 41-47*

T1.419 Dilek Hakkani-Tur, F. Béchet, G. Riccardi, Gokhan Tur, Beyond ASR 1-Best: Using Word Confusion Networks for Spoken Language Understanding, *Computer Speech and Language, Elsevier, vol.20-4, October 2006*, pp. 495-514

T1.418 S. Matsoukas, J.-L. Gauvain, G. Adda, T. Colthurst, Chia-Lin Kao, O. Kimball, L. Lamel, F. Lefèvre, J.Z. Ma, J. Makhoul, L. Nguyen, R. Prasad, Advances in transcription of broadcast news and conversational telephone speech within the combined EARS BBN/LIMSI system, 2006, *IEEE Transactions on Audio, Speech and Language Processing, Volume 4, Issue 5, Sept. 2006*, pp. 1541-

ACLN (Articles dans des revues internationales ou nationales avec comité de lecture, non référencées)

T1.406 Florian Boudin, Stéphane Huet and Juan Manuel Torres-Moreno, A Graph-Based Approach to Cross-Language Multi-Document Summarization, *Polibits, Research journal on Computer science and computer engineering with applications, 2011, 43 :113-119, ISSN 1870-9044, 2011. 3*

T1.405 Iria da Cunha, Maria Teresa Cabré, Eric SanJuan, Juan Manuel Torres Moreno and Jorge Vivaldi, Automatic specialized vs. non-specialized text differentiation : the usability of grammatical features in a Latin multilingual context. *TISLID monographic, volume, 2011*

T1.404 M.-J. Meurs, F. Lefèvre, R. De Mori, Approche bayésienne de la composition sémantique dans les systèmes de dialogue, 2010, *Ingénierie des Systèmes d'Information, RSTI série ISI, Vol 15 :2, Hermes-Lavoisier, 20p*

T1.403 Da Cunha Fanego, J-M. Torres, Velázquez, P., Vivaldi, J., Un algoritmo lingüístico-estadístico para resumen automático de textos especializados, 2009, *Linguamática 2. 67-79. ISSN 1647-0818.*

T1.402 El-Bèze, J-M. Torres, F. Béchet, Un duel probabiliste pour départager deux Présidents, 2007,

RNTI E-10 ISBN10 2-85428-776-2

T1.401 R. Kessler, J-M. Torres, M. El-Bèze, *Classification automatique de courriers électroniques par des méthodes mixtes d'apprentissage*, 2006, RSTI-ISI Vol 11 No 2. Pag 93-112

C-ACTI (Communications avec actes et comité de sélection dans un congrès international)

T1.385 Ludovic Bonnefoy, Patrice Bellot, Michel Benoit : *The Web as a Source of Evidence for Filtering Candidate Answers to Natural Language Questions*, 2011 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2011), August 22 – 27, 2011, Lyon

T1.384 Thierry Bazillon, Benjamin Maza, Mickael Rouvier, Frederic Bechet et Alexis Nasr, *Speaker Role Recognition using question detection and characterization* International conference of speech Conference of the Speech Communication Association, InterSpeech 2011

T1.383 F. Duvert et R. de Mori : *A conditional model for triggering understanding actions in a speech understanding system*. ICASSP 2011 Prague : 5620-5623

T1.382 J.F. Bonastre, P.M. Bousquet, D. Matrouf, x. m. Anguera *Discriminant binary data representation for speaker recognition*, IEEE International conference on speech audio and signal processing, ICASSP, 2011, Prague, Czech republic

T1.376 R. Deveaud, Boudin Florian, P. Bellot, LIA at INEX 2010 Book Track, 2011, Lecture Notes in Computer Science LNCS - Elsevier

T1.372 Alejandro Molina, Juan-Manuel Torres-Moreno, Iria Da Cunha, Eric Sanjuan, Gerardo Sierra and Patricia Velazquez-Morales, *Contextual Discourse Segmentation for Sentence Compression*, MICAI'11, Puebla (Mexique), nov 26-déc 4 2011

T1.372 Fatima Itzel Ronquillo, Concepcion Perez de Celis, Gerardo Sierra, Iria da Cunha, Juan-Manuel Torres-Moreno, *Automatic classification of biomedical texts : experiments with a hearing loss corpus*. The 4th International Congress on Image and Signal Processing and the 4th International Conference on BioMedical Engineering and Informatics, Shanghai

T1.371 Brenda Gabriela Castro Rolon, Gerardo Sierra, Juan-Manuel Torres-Moreno, Iria da Cunha. *Estructuras discursivas y calculo de similitud semantica como recursos para la deteccion de similitud textual*. III Workshop "A RST e os Estudos do Texto - STIL'11", Mato Grosso Brésil 2011.

T1.370 Iria da Cunha, Juan-Manuel Torres-Moreno, Gerardo Sierra, Luis Adrián Cabrera-Diego, Brenda Gabriela Castro Rolon and Juan Miguel Rolland Bartilotti. *The RST Spanish Treebank On-line Interface*. RANLP'11, 12-14 septembre 2011, Hissar, Bulgarie.

T1.369 Iria da Cunha, Juan-Manuel Torres-Moreno and Gerardo Sierra. *Aplicaciones linguisticas del analisis discursivo automatico*. XII Simposio Internacional de Comunicacion Social en el Siglo XXI, Vol. II, pp. 919-923 (Santiago de Cuba), 2011.

T1.368 E. San-Juan, P. Bellot, *Report on INEX 2009*", 2010, ACM SIGIR Forum 44, 1 (August 2010), 38-5

T1.367 Juan Manuel Torres-Moreno and Michel Gagnon, *The Cortex automatic summarization system at the QA@INEX track 2010*, INEX@QA'10, Lecture Notes in Computer Science Springer, 2011

T1.366 Iria da Cunha, Maria Teresa Cabré, Eric SanJuan, Juan Manuel Torres-Moreno, Jorge Vivaldi and Gerardo Sierra, *Automatic Specialized vs. Non-Specialized Sentence Differentiation*, CILing'11, Lecture Notes in Computer Science Volume 6608/2011 Springer 0302-9743 (Print) 1611-3349 (Online), pp 266–276, 2011

T1.365 I. Da Cunha Fanego, A. Molina, *Optimizacion de resumen automatico mediante compresion de frases*, 2011, AESLA 2010, Vigo, Espagne.

T1.364 Rojas-Barahona L.M., Bazillon T., Quignard M., F. Lefèvre, *Using MMIL for the High Level Semantic Annotation of the French MEDIA Dialogue Corpus*", 2011, International Conference on Computational Semantics, Oxford, UK

T1.363 R. Rubino, G. Linarès, *A Multi-view approach for Term Translation Spotting*, 2011, Conference on Intelligent Text Processing and Computational Linguistics, CiCling, 2011, Tokyo

T1.362 M. Bouallegue, D. Matrouf, G. Linarès, *A simplified subspace gaussian mixture to compact acoustic model for speech recognition*, 2011, IEEE International Conference on Audio, Speech and Signal Processing, ICASSP, 2011, Prague, Czech Republic

T1.361 G. Senay, G. Linarès, B. Lecouteux, *A segment-level confidence measure for spoken*

document retrieval, 2011, *IEEE International Conference on Speech Audio and Signal Processing, ICASSP, 2011, Prague, Czech Republic*

T1.360 Anguera Xavier, J-F. Bonastre, *Fast speaker diarization based on binary keys*, 2011 ICASSP 2011, May 2011, Prague

T1.359 J. Kahn, Aubibert Nicolas, Rossato Solange, J-F. Bonastre, *Speaker verification by inexperienced and experienced listeners vs. speaker verification system*, 2011, ICASSP 2011, May 2011, Prague

T1.358 Bonnefoy Ludovic, P. Bellot, Benoit Michel, *LIA-iSmart at TREC 2010 : A Web-oriented Language Modeling Approach for Question Related Entity Finding*, 2011, Text REtrieval Conference, NIST Special publication, 2011

T1.357 P. Clement, T. Bazillon, C. Fredouille, *Speaker Diarization of Heterogeneous Web Video Files: a Preliminary Study*, 2011, ICASSP'2011, May 2011, Prague

T1.356 C. Fredouille, Pouchoulin Gilles, *Automatic Detection of Abnormal Zones in Pathological Speech*, 2011, ICPHS'2011, August 2011, Hong-Kong

T1.355 J-F. Bonastre, X. Anguera, G. Sierra, P-M. Bousquet, *Speaker modeling using local binary decisions*, 2011, Interspeech 2011, Florence

T1.354 P-M. Bousquet, J-F. Bonastre, D. Matrouf, *Intersession compensation and scoring methods in the i-vectors space for speaker recognition*, Interspeech 2011, Florence

T1.353 B. Maza, M. El Bèze, G. Linarès, R. De Mori, *On the use of linguistic features in an automatic system of speech analytics of telephone conversations*, InterSpeech 2011, Florence, It, August 2011

T1.352 S. Huet, P. Langlais, *Identifying the Translations of Idiomatic Expressions using TransSearch*, 2011, 8th International Workshop on Natural Language Processing and Cognitive Science (NLPCS), Copenhagen (Denmark)

T1.351 S. Huet, F. Lefèvre, *Unsupervised Alignment for Segmental-based Language Understanding*, 2011, EMNLP 1st Workshop on Unsupervised Learning in NLP (UNSUP), Edinburgh (UK)

T1.350 N. Camelin, B. Detienne, S. Huet, D. Quadri, F. Lefèvre, *Unsupervised Concept Annotation using Latent Dirichlet Allocation and Segmental Methods*, 2011, EMNLP 1st Workshop on Unsupervised Learning in NLP (UNSUP), Edinburgh (UK)

T1.349 M. Potet, R. Rubino, B. Lecouteux, S. Huet, L. Besacier, H. Blanchon, F. Lefèvre, *The LIGA (LIG/LIA) Machine Translation System for WMT 2011*, 2011, EMNLP 6th Workshop on Statistical Machine Translation (WMT), Edinburgh (UK)

T1.348 J. Yingbo, B. Merialdo, M. Rouvier, G. Linarès, *Static and Dynamic Video Summaries*, 2011, ACM Multimedia Conference, Scottsdale, Arizona, USA

T1.347 G. Senay, S. Oger, R. Rubino, G. Linarès, Parent Thomas, *Audio Indexing on a Medical Video Database: the AVISON Project*, 2011, International Conference on BioMedical Engineering and Informatics (BMEI'11), Shanghai, China

T1.346 N. Camelin, B. Detienne, S. Huet, D. Quadri, F. Lefèvre, *Concept Discovery for Language Understanding in an Information-Query Dialogue System*, 2011, International Conference on Knowledge Discovery and Information Retrieval, KDIR 2011, Paris (France)

T1.345 F. Pinault, F. Lefèvre, *Semantic graph clustering for POMDP-based spoken dialog systems*, 2011, ISCA International Conference of the Speech Communication Association, InterSpeech'09, Brighton

T1.349 F. Pinault, F. Lefèvre, *Unsupervised clustering of probability distributions of semantic frame graphs for POMDP-based spoken dialogue systems with summary space*, 2011, IJCAI/7th Workshop on knowledge and reasoning in practical dialogue systems, Barcelona

T1.348 B. Jabaian, L. Besacier, F. Lefèvre, *Combination of stochastic understanding and machine translation systems for language portability of dialogue systems*, 2011, IEEE International Conference on Acoustic, Speech and Signal Processing, May 2011, Prague

T1.347 F. Béchet, C. Raymond, F. Duvert, et R. de Mori, 2010. *Frame based interpretation of conversational speech*, à paraître. IEEE Workshop on Speech and Language Technologies, SLT, 2010

T1.346 F. Duvert et R. de Mori, 2010. *Conditional models for detecting lambda-functions in a spoken language understanding system*, 2010 Japon Makuhari. Interspeech 2010, Makuhari, Japan

T1.379 B. Jabaian, L. Besacier, F. Lefèvre, *Investigating multiple approaches for SLU portability to a new language*, 2010, ISCA Interspeech, Tokyo

- T1.345** Mickael Rouvier, Richard Dufour, Georges Linarès et Yannick Estève, A Language identification inspired method for spontaneous speech detection International conference of speech Conference of the Speech Communication Association, InterSpeech 2010
- T1.344** E. San-Juan, P. Bellot, Report on INEX 2010, 2011, ACM SIGIR Forum
- T1.343** J.Kahn, N. Audibert, S. Rossato, J.F. Bonastre, Intra-speaker variability effects on Speaker Verification performance, 2010, Speaker Odyssey, Brno, République Tcheque
- T1.342** P. Tomassek, C. Fredouille, D. Matrouf, Factor analysis-based approaches applied to the speaker diarization task of meetings : a preliminary study, 2010, Speaker Odyssey, Brno, République Tchèque
- T1.341** F. Verdet, D. Matrouf, JF Bonastre, J. Hennebert, Copying two different transmission channels in language recognition 2010, Speaker Odyssey, Brno, République Tcheque
- T1.340** C. Servan, N. Camelin, C. Raymond, F. Béchet, R. De Mori, On the use of machine translation for spoken language, understanding portability, 2010, In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Mars, Dallas, USA
- T1.339** F. Verdet, D. Matrouf, J-F. Bonastre, J. Hennebert, Coping with Two Different Transmission Channels in Language Recognition, Speaker Odyssey, The Speaker and Language Recognition Workshop 2010 (in print)
- T1.338** R. Lavalley, C. Clavel, M. El-Bèze, P. Bellot, Finding topic-specific strings in text categorization and opinion mining contexts, 2010, International Conference on Data Mining 2010, DMIN'10, Las Vegas, USA
- T1.337** T. Beckers, P. Bellot, G. Demartini, L. Denoyer, C. M. De Vries, A. Doucet, K-N Fachry, N. Fuhr, P. Gallinari, S. Geva, Wei-Che Huang, T. Iofciu, Jaap Kamps, G. Kazai, M. Koolen, S. Kutty, M. Landoni, M. Lehtonen, V. Moriceau, R. Nayak, R. Nordlie, N. Pharo, E. SanJuan, R. Schenkel, X. Tannier, M. Theobald, J. A. Thom, A. Trotman, A. P. de Vries, "Report on INEX 2009", 2010, ACM SIGIR, Genève
- T1.336** V. Moriceau, E. SanJuan, X. Tannier, P. Bellot, Overview of the 2009 QA Track: Towards a common task for QA, focused IR and automatic summarization systems, 2010, Proceedings of 8th Workshop of the Initiative for the Evaluation of XML retrieval (INEX 2009) - LNCS - Elsevier
- T1.335** S. Bozonnet, N. Evans, C. Fredouille, The LIA-EURECOM RT'09 Speaker Diarization System: Enhancements in Speaker Modelling and Cluster Purification, 2010, ICASSP'10, Dallas, USA, March 2010
- T1.334** E. Charton, A. Larcher, C. Lévy, J-F. Bonastre, Mistral : open source biometric platform, 2010, 25th Symposium on Applied Computing (ACM), Sierre, Switzerland
- T1.333** A. Larcher, J-F. Bonastre, J. S.D. Mason, Constrained Viterbi decoding for embedded user-customised password speaker recognition, 2010, 25th Symposium on Applied Computing (ACM)
- T1.332** S. Fernandez, P. Jourlin, E. SanJuan, Unsupervised Mining of Knowledge Gaps in Scientific Literature, JADT 2010, in press.
- T1.331** S. Oger, M. Rouvier, G. Linarès, Transcription-based video genre classification, 2010, IEEE International Conference on Speech and Signal Processing, ICASSP 2010, Dallas, USA
- T1.330** F. Béchet, E. Charton, Unsupervised knowledge acquisition for extracting Named Entities from Speech, 2010, IEEE International Conference on Speech Audio and Signal Processing, ICASSP 2010, Dallas, USA
- T1.329** M. Rouvier, G. Linarès, D. Matrouf, On-the-fly Video genre classification by combination of audio features, 2010, IEEE International Conference on Speech Audio and Signal Processing, ICASSP 2010, Dallas, USA
- T1.328** C. Fredouille, C. Fougeron, O. Panseri, N. Lévèque, Automatic speech processing for the detection of abnormal speech zones in dysarthric speech, 2010, 13th Meeting of the International Clinical Phonetics and Linguistics Association (ICPLA), Oslo, 23-26 jun, 2010 (sur résumé)
- T1.327** E. Charton, J-M. Torres, NLGbAse: a free linguistic resource for Natural Language Processing systems, 2010, LREC 2010
- T1.326** B. Lecouteux, Nocéra Pascal, G. Linarès, Semantic cache model driven speech recognition, 2010 IEEE International Conference on Speech Audio and Signal Processing, ICASSP 2010, Dallas, USA
- T1.325** C. Fougeron, L. Crevier-Buchman, C. Fredouille, A. Ghio, C. Meunier, J-F. Bonastre, G. Pouchoulin, et al., Developing an acoustic-phonetic characterization of dysarthric speech in French, 2010 International conference on Language Resources and Evaluation (LREC)

- T1.324** S. Bonnoz, N. W. D. Evans, C. Fredouille, *The LIA-EURECOM RT'09 speaker diarization system : Enhancements in speaker modelling and cluster purification*, 2010, ICASSP'10, Dallas, US, March
- T1.323** C. de Loupy, S. Seng, C. Ayache, J-M. Torres, *A French Human Reference Corpus for multi-documents summarization and sentence compression*, 2010, LREC 2010
- T1.322** Jorge Vivaldi, I. Da Cunha Fanego, J-M. Torres, P. Velazquez-Morales, *Automatic Summarization Using Terminological and Semantic Resources*, 2010, LREC 2010
- T1.321** J-M. Torres, J. Ramirez, *REG : un algorithme glouton appliqué au résumé automatique de texte*, 2010, JADT'10
- T1.320** A. Molina, G. Sierra, J-M. Torres, *La energía textual como medida de distancia en agrupamiento de definiciones*, 2010, JADT'10
- T1.319** G. Senay, G. Linarès, B. Lecouteux, S. Oger, T. Michel, *Transcriber driving strategies for transcription aid system*, 2010, LREC 2010
- T1.318** E. SanJuan, F. Ibekwe, *Multi Word Term queries for focused Information Retrieval*, 2010, 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2010), Romania, March 21-27, LNCS, vol 6008, to appear march 10 2010, 12p.
- T1.317** F. Ibekwe, E. SanJuan, *Knowledge Organization Research in the last two decades: 1988-2008*, 2010, 11th International Conference of the International Society for Knowledge Organization (ISKO), Paradigms and conceptual systems in KO, Sapienza University of Rome, feb. 23-26 2010, pp 115-121.
- T1.316** J. Kahn, S. Rossato, J-F. Bonastre, *Beyond Doddington menagerie, a first step towards*, IEEE International Conference on Speech Audio and Signal Processing, ICASSP 2010, Dallas, USA
- T1.315** V. Popescu, J. Caelen, *The Non-Individuation Constraint Revisited: When to Produce Free Choice Items in Multi-Party Dialogue*, The 13th SEMDIAL Workshop on the Semantics and Pragmatics of Dialogue (DiaHolmia), Stockholm, 2009, pp. 91-98
- T1.314** V. Popescu, J. Caelen, C. Burileanu, *Optimizing a Discourse Structuring Component for Utterance Generation in Human--Computer Dialogue*, The 5th Conference on Speech Technology and Human--Computer Dialogue (SpeD), Constanța, 2009,
- T1.313** F. Piazza, S. Cecchi, L. Palestini, A. Lattanzi, F. Bettarelli, F. Capman, S. Thabuteau, C. Lévy, J-F. Bonastre, R. Toppi, *The hArtes Carlab: hardware implementation and algorithm development*, 2009, AES 36th International Conference, Dearborn - USA
- T1.312** R. Fraile, N. Saenz-Lechon, J. I. Godinoa-Llorente, V. Osma-Ruiz, C. Fredouille, *MFCC-based remote pathology detection on speech transmitted through the telephone channel*, 2009, International Conference on Bio-inspired Systems and Signal Processing, BIOSIGNAL'09
- T1.311** S. Oger, V. Popescu, G. Linarès, *Using the Word Wide Web for learning new words in Continuous Speech recognition tasks : two case studies*, 2009, In Proceedings of Speech And Computer, SPECOM 2009, St Petersburg, Russia
- T1.310** R. Fernandez, J-F. Bonastre, D. Matrouf, J. R. Calvo, *Feature Selection Based on Information Theory for Speaker Verification*, 2009, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, LNCS 5856, Springer
- T1.309** R. Rubino *Exploring context variations and lexical coverage in projection-based approach for term translation*, RANLP 2009, Bulgaria
- T1.308** R. Dufour, V. Jousse, Y. Estève, F. Béchet, G. Linarès, *Spontaneous Speech Characterization and Detection in Large Audio Database*, 2009, In proceedings of Speech and Computer, SPECOM'09, St Petersburg, Russia
- T1.307** R. Dufour, V. Jousse, Y. Estève, F. Béchet, G. Linarès, *Spontaneous Speech Characterization and Detection*, 2009, First Young Researchers Workshop on Speech Processing, Dublin, Ireland
- T1.306** N. Camelin, R. De Mori, F. Béchet, G. Damnati, *Error correction of proportions in spoken opinion surveys*. In Proceedings of the Interspeech conference, 2009, Brighton, UK
- T1.305** R. Kessler, N. Bechet, M. Roche, M. El-Bèze, J-M. Torres, *Job Offer Management: How to Improve the Ranking of Candidates*, 2009, International Symposium on Methodologies for Intelligent Systems (ISMIS2009), The Springer LNCS Volume 5722/2009, p 431-441
- T1.304** Evans, C. Fredouille, J-F. Bonastre, *speaker diarization using unsupervised discriminant analysis of inter-channel delay features*, 2009, ICASSP'09, Taipei, March 2009
- T1.303** M-J. Meurs, F. Lefèvre, R. De Mori, *Spoken Language Interpretation: On the Use of Dynamic Bayesian Networks for Semantic Composition*, 2009, IEEE International Conference on Acoustic,

Speech and Signal Processing

T1.302 M-J. Meurs, F. Lefèvre, R. De Mori, Learning Bayesian Networks for Semantic Frame Composition in a Spoken Dialog System, 2009, HLT NAACL, Boulder, Co USA

T1.301 M. Rouvier, G. Linarès, D. Matrouf, Robust Audio-based Classification of Video Genre, 2009, International conference of the Speech Communication Association, ISCA, InterSpeech'09, Brighton, UK

T1.300 M. Rouvier, D. Matrouf, G. Linarès, Factor Analysis for Audio-based Video Genre Classification, 2009, International conference of the Speech Communication Association, ISCA, InterSpeech'09, Brighton, UK

T1.299 B. Lecouteux, G. Linarès, B. Favre, Combined low level and high level features for Out-Of-Vocabulary Word detection, 2009, International conference of the Speech Communication Association, ISCA, InterSpeech'09, Brighton, UK

T1.298 S. Oger, V. Popescu, G. Linarès, Probabilistic and Possibilistic Language Models Based on the World Wide Web, 2009, International conference of the Speech Communication Association, ISCA, InterSpeech'09, Brighton, UK

T1.297 F. Verdet, D. Matrouf, J-F. Bonastre, J. Hennebert, Factor Analysis and SVM for Language Recognition, 2009, International Conference on Speech Communication and Technology (Interspeech) 2009

T1.296 P. Gotab, F. Béchet, G. Damnati, Active Learning for rule-based and corpus-based Spoken Language Understanding models, 2009, IEEE Workshop on Automatic Speech Recognition & Understanding (ASRU), Merano, Italie

T1.295 F. Pinault, F. Lefèvre, R. De Mori, Feature-based Summary Space for Stochastic Dialogue Modeling with Hierarchical Semantic Frames, 2009, International conference of the Speech Communication Association, ISCA, InterSpeech'09, Brighton, UK

T1.294 I. Da Cunha Fanego, Primera aproximación para la elaboración de un nuevo diccionario electrónico de términos inmobiliarios en español, 2009, Proceedings of the XXVII AESLA International Conference: Ways and Modes of Human Communication. Ciudad Real: Universidad de Castilla-La Mancha.

T1.293 Cabré, M.T., Bach, C., I. Da Cunha Fanego, Morales, A., Vivaldi, J., Comparación de algunas características lingüísticas del discurso especializado frente al discurso general: el caso del discurso económico, 2009, Proceedings of the XXVII AESLA International Conference: Ways and Modes of Human Communication. Ciudad Real: Universidad de Castilla-La Mancha.

T1.292 E. SanJuan, The landscape of information science 1996-2008, 2009, JCDL 2009, Austin, Texas, USA, pp. 401-402. The landscape of Information Science: 1996-2008, ACM/IEEE Joint Conference on Digital libraries. Designing tomorrow, preserving the past - today (JCDL09), Austin, Texas, 15-19 June, 2009

T1.291 M. Gasic, F. Lefèvre, F. Jurcicek, S. Keizer, F. Mairesse, B. Thomson, K. Yu, S. Young, Back-off Action Selection in Summary Space-Based POMDP Dialogue Systems, 2009, IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), Merano, Italy

T1.290 F. Lefèvre, M. Gasic, F. Jurcicek, S. Keizer, F. Mairesse, B. Thomson, K. Yu, S. Young, k-Nearest Neighbor Monte-Carlo Control Algorithm for POMDP-based Dialogue Systems, 2009, ACL/ISCA SIGDIAL, London

T1.289 J. Kahn, S. Rossato, Do Humans and speaker verification system use the same information to differentiate voices?, 2009, InterSpeech 2009, Brighton, UK

T1.288 Ibekwe-SanJuan F., S. Fernandez, E. SanJuan, E. Charton, Annotation of Scientific Summaries for Information Retrieval, 2008, ECIR-ESAIR 2008

T1.287 T. Waszak, J-M. Torres, Compression entropique de phrases contrôlée par un perceptron, 2008, JADT 2008

T1.286 J-F. Bonastre, N. Scheffer, D. Matrouf, C. Fredouille, A. Larcher, A. Prete, G. Pouchoulin, N. Evans, B. Fauve, J. Mason, ALIZE/SpkDet: a state-of-the-art open source software for speaker recognition, Speaker Odyssey, South Africa, January 2008

T1.285 A. Moschitti, G. Riccardi, C. Raymond, Spoken Language Understanding With Kernels For Syntactic/Semantic Structures, 2008, Proceedings IEEE Workshop on Automatic Speech Recognition and Understanding

T1.284 S. Oger, G. Linarès, F. Béchet, Local methods for on-demand out-of-vocabulary word retrieval, 2008, International conference on Language Resources and Evaluation, LREC 2008

- T1.283** C. Raymond, G. Riccardi, *Learning with Noisy Supervision for Spoken Language Understanding*, 2008, *International IEEE Conference on Acoustic Speech and Signal Processing*
- T1.282** E. SanJuan, N. Flavier, F. Ibekwe-SanJuan, P. Bellot, *Univ. of Avignon & Lyon III at TREC'08 : Enterprise Track*, 2009, *NIST Special Publication 500-277: The 17th Text REtrieval Conference Proceedings (TREC 2008)* ",
- T1.281** S. Fernandez, P. Velázquez, S. Mandin, E. SanJuan, J-M. Torres, *Les systèmes de résumé automatique sont-ils vraiment des mauvais élèves ?*, *JADT 2008*
- T1.280** E. Charton, J-M. Torres, E. SanJuan, *Réécriture automatique de phrases par modèle de langage*, *JADT 2008*
- T1.279** L. Sitbon, P. Bellot, P. Blache, *Evaluation of lexical resources and semantic networks on a corpus of mental associations*, 2008, 6th edition of the *Language Resources and Evaluation Conference (LREC 2008)*, Marrakech (Maroc), mai 2008.
- T1.278** L. Sitbon, P. Bellot, P. Blache, *A corpus of real-life questions for evaluating robustness of QA systems*, 2008, 6th edition of the *Language Resources and Evaluation Conference (LREC 2008)*, Marrakech (Maroc), mai 2008.
- T1.277** C. Fredouille, N. Evans, *The LIA RT07 Speaker Diarization System*, 2008, *Multimodal Technologies for Perception of Humans. LNCS - Springer - Editor Stiefelhaven, Bowers, Fiscus*
- T1.276** F. Boudin, Juan-Manuel Torres-Moreno, P. Velazquez-Morales. *An efficient Statistical Approach for Automatic Organic Chemistry Summarization. In Proceedings of the International Conference on Natural Language Processing (GoTAL), Gothenburg (Sweden), 2008 25-27 August. The Springer LNCS 5221 - pp 89-99.*
- T1.275** F. Boudin, J-M. Torres, M. El-Bèze, *Mixing Statistical and Symbolic Approaches for Chemical Names Recognition.*, 2008, *Proceedings of the conference CICLing 2008, Haifa (Israel), 2008 17-23 February. The Springer LNCS Proceedings - p 334-343*
- T1.274** S. Hahn, P. Lehnen, C. Raymond, H. Ney, *A Comparison of Various Methods for Concept Tagging for Spoken Language Understanding*, 2008, *Proceedings of the Language Resources and Evaluation Conference*
- T1.273** F. Duvert, M-J. Meurs, C. Servan, F. Béchet, F. Lefèvre, R. De Mori, *Semantic composition process in a speech understanding system*, 2008, *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'08)*
- T1.272** A. Larcher, J-F. Bonastre, J. S.D. Mason, *From GMM to HMM for Embedded Password-Based Speaker Recognition*, 2008, "*European Signal and Image Processing Conference (EUSIPCO)*",
- T1.271** A. Prete, F. Capman, B. Ravera, J-F. Bonastre, *An application constrained front end for speaker verification*, 2008, *EUSIPCO*
- T1.270** L. Sitbon, P. Bellot, *How to cope with questions typed by dyslexic users*, 2008, "*ACM SIGIR Workshop on "Analytics for Noisy Unstructured Text Data"*", *Singapour*",
- T1.269** H-Q NGUYEN, P. Nocera, E. Castelli, TRINH Van Loan, *Tone Recognition of Vietnamese Continuous Speech using Hidden Markov Model*, 2008, "*HUT-ICCE*",
- T1.268** M-J. Meurs, F. Duvert, F. Béchet, F. Lefèvre, R. De Mori, *Semantic Frame Annotation on the French MEDIA corpus*, 2008, *International conference on Language, Resource and Evaluation LREC*
- T1.267** A. Larcher, J-F. Bonastre, J. S. D. Mason, *Reinforced Temporal Structure Information For Embedded Utterance-Based*, 2008, *Interspeech*
- T1.266** M-J. Meurs, E. SanJuan, *Combining Optimal and Atomic Decomposition of Terminology Association graphs*, 2008, *MLG 2008*,
- T1.265** F. Ibekwe-SanJuan, E. SanJuan, M. Vogele, *Graph decomposition of terminology graphs for Domain Knowledge Acquisition (Poster)*, 2008, *ACM 17th Conf. on Information and Knowledge Management (CIKM 2008)*,
- T1.264** M-J. Meurs, F. Duvert, F. Lefèvre, R. De Mori, *Markov Logic Networks for Spoken Language Interpretation*, 2008, *Intelligent Information Systems International Conference, Zakopane*
- T1.263** M. Rouvier, G. Linarès, B. Lecouteux, *On-the-fly term spotting by phonetic filtering and request-driven decoding*, 2008, *IEEE/ISCA Workshop on Speech and Language Technology SLT'08*, Goa, India
- T1.262** H-Q Nguyen, P. Nocera, E. Castelli, Trinh Van Loan, *A Novel Approach in Continuous Speech Recognition for Vietnamese, an isolating tonal language*, 2008, "*Interspeech 2008*",
- T1.261** N. Camelin, G. Damnati, F. Béchet, R. De Mori, *Automatic customer feedback processing :*

alarm detection in open question spoken messages, 2008, "Interspeech'08",

T1.260 M-J. Meurs, F. Lefèvre, R. De Mori, A Bayesian approach to semantic composition for spoken language interpretation, 2008, ISCA Interspeech, Brisbane

T1.259 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Giovanni, Dysphonic Voices and the 0-3000Hz Frequency Band, 2008, Interspeech'08, Brisbane, Australia, September 2008

T1.258 S. Oger, G. Linarès, F. Béchet, P. Nocera, On-demand new word learning using World Wide Web, 2008, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008, Las Vegas, USA

T1.257 F. Boudin, M. El-Bèze, J-M. Torres, A Scalable {MMR} Approach to Sentence Scoring for Multi-Document Update Summarization, 2008, "Coling 2008: Companion volume: Posters and Demonstrations",

T1.256 B. Lecouteux, G. Linarès, Using prompts to produce quality corpus for training automatic speech recognition systems, 2008, IEEE Mediterranean Electrotechnical Conference, MELCON'08, Ajaccio, France

T1.255 B. Lecouteux, G. Linarès, Y. Estève, G. Gravier, Generalized driven decoding for speech recognition system combination, 2008, IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'08)

T1.254 L. Barrault, D. Matrouf, G. Linarès, R. De Mori, Frame-Based Acoustic Feature Integration for Speech Understanding, 2008, IEEE International Conference on Acoustic, Speech and Signal Processing, ICASSP'08, Las Vegas, USA

T1.253 C. Chen, F. Ibekwe-SanJuan, E. SanJuan, M. Vaugeley, Identifying Thematic Variations in SDSS Research : The Impact of Individual Data Releases, 9th JADT, Lyon, France, 10-14 March 2008

T1.252 C. Raymond, Kepa Joseba Rodriguez, G. Riccardi, Active Annotation in the LUNA Italian Corpus of Spontaneous Dialogues, 2008, Proceedings of the Language Resources and Evaluation Conference

T1.251 L. Sitbon, P. Bellot, A readability measure for an information retrieval process adapted to dyslexics, 2008, "Second international workshop on Adaptive Information Retrieval (AIR 2008) (in conjunction with IliX 2008)",

T1.250 R. Kessler, M. El-Bèze, E-Gen : traitement automatique des offres d'emploi, 2008, JADT

T1.249 F. Duvert, M-J. Meurs, C. Servan, F. Béchet, F. Lefèvre, R. De Mori, Semantic Composition Process in a Spoken Understanding System, 2008, Intelligent Information Systems International Conference, Zakopane

T1.248 C. Fredouille, N. Evans, New implementations of the E-HMM based system for speaker diarization in meeting rooms, 2008, ICASSP'08, Las Vegas, USA, April 2008

T1.247 SE Mezaache, J-F. Bonastre, D. Matrouf, Analysis of impostor tests with high scores in NIST-SRE context, 2008, Interspeech 2008, Brisbane (Australia)

T1.246 M. McLaren, D. Matrouf, R. Vogt, J-F. Bonastre, Combining Continuous Progressive Model Adaptation and Factor Analysis for Speaker Verification, 2008, pp 857-860, Interspeech 2008, Brisbane (Australia)

T1.245 G. Damnati, F. Béchet, R. De Mori, Experiments on the France Telecom 3000 Voice Agency corpus: academic research on an industrial spoken dialog system, 2007, HTL-NAACL Workshop - Bridging the Gap: Academic and Industrial Research in Dialog Technologies - Rochester, NY, Avril 2007

T1.244 G. Damnati, F. Béchet, R. De Mori, Spoken Language Understanding strategies on the France Telecom 3000 Voice Agency corpus, 2007, ICASSP 07, Honolulu, HI

T1.243 L. Sitbon, P. Bellot, P. Blache, Phonetic based sentence level rewriting of questions typed by dyslexic spellers in an information retrieval context, 2007, Interspeech 2007, Anvers (Belgique), p. 2543 à 2548

T1.242 B. Favre, J-F. Bonastre, P. Bellot, An Interactive Timeline for Speech Database Browsing, 2007, Interspeech 2007, Anvers (Belgique), p. 2413 à 2416

T1.241 L. Sitbon, P. Bellot, Topic segmentation using weighted lexical links (WLL), 2007, ACM SIGIR 07, ACM Press, Amsterdam (Pays-Bas), p. 737-738

T1.240 G. Linarès, Nocéra Pascal, D. Massonié, D. Matrouf, The LIA speech recognition system : from 10xRT to 1xRT, 2007, International Conférence on Speech, Text and Dialogue (ISCA), LNCS, Springer, ISSN 0302-9743, ISBN 978-3-540-74627-0, volume 4629/2007, pp. 302-308, Pilsen, 2007

- T1.239** A. Nimaan, P. Nocera, F. Béchet, J-F. Bonastre, *Information Retrieval Strategies for Accessing African Audio Corpora*, 2007, *Interspeech 2007*, Antwerp, Belgium
- T1.238** B. Lecouteux, G. Linarès, F. Beaugendre, P. Nocera, *Text Island spotting in large speech database*, 2007, *Conference of the International Speech Communication Association, ISCA, InterSpeech 2007*, Antwerp, Belgium
- T1.237** C. Lévy, G. Linarès, J-F. Bonastre, *Fast adaptation of GMM-based compact models*, 2007, *Conference of the International Speech Communication Association, ISCA, InterSpeech 2007*, Antwerp, Belgium
- T1.236** F. Boudin, B. Favre, F. Béchet, M. El-Bèze, L. Gillard, J-M. Torres, *The LIA-Thales summarization system at DUC-2007*, *DUC-NIST Proceedings Document Understanding Conference (DUC-2007)*, New York, 26-27 april 2007
- T1.235** F. Boudin, J-M. Torres, *A Cosine Maximization-Minimization approach for User-Oriented Multi-Document Update Summarization*, In *Proceedings of the conference RANLP Borovets (Bulgaria)*, 27-29 Sept. 2007, pp 81-87.
- T1.234** A. Preti, J-F. Bonastre, D. Matrouf, F. Capman, B. Ravera, *Confidence measure based unsupervised target model adaptation for speaker verification*, 2007, *Interspeech'07*, Antwerp, Belgium
- T1.233** G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Giovanni, *Frequency Study for the Characterization of the Dysphonic Voices*, 2007, *Interspeech'07*, Antwerp, Belgium, September 2007
- T1.232** J-F. Bonastre, C. Fredouille, A. Ghio, A. Giovanni, G. Pouchoulin, J. Revis, B. Teston, P. Yu, *Complementary approaches for voice disorder assessment*, 2007, *Interspeech'07*, Antwerp, Belgium, September 2007
- T1.231** M. Didi Biha, Bangaly Kaba, M-J. Meurs, E. SanJuan, *Graph decomposition approaches for terminology graphs*, 2007, *MICAI'07, LNCS Volume 4827/2007*, Aguascalientes, Mexico
- T1.230** Iria da Cunha, S. Fernandez, P. Velazquez Morales, Jorge Vivaldi, E. SanJuan, J-M. Torres, *A new hybrid summarizer based on Vector Space model*, *Statistical Physics and Linguistics*, 2007 *MICAI'07, LNCS Volume 4827/2007*, Aguascalientes, Mexico
- T1.229** N. Camelin, F. Béchet, G. Damnati, R. De Mori, *Speech Mining in Noisy Audio Message Corpus*, 2007, *Proc. Interspeech*, August 2007, Antwerp, Belgium
- T1.228** E. SanJuan, F. Ibekwe-SanJuan, J-M. Torres, P. Velazquez-Morales, *Combining Vector Space Model and Multi Word Term Extraction for Semantic Query Expansion*, 2007, *NLDB 07 Paris, LNCS 4592 Springer 2007, ISBN 978-3-540-73350-8*, p. 252-263
- T1.227** S. Fernandez, E. SanJuan, J-M. Torres, *Textual Energy of Associative Memories: performants applications of ENERTEX algorithm in text summarization and topic segmentation*, 2007, *MICAI 2007, LNCS Volume 4827/2007*, Aguascalientes, Mexico
- T1.226** R. Kessler, J-M. Torres, M. El-Bèze, *E-Gen: Automatic Job Offer Processing system for Human Ressources*, 2007, *MICAI 2007, LNCS Volume 4827/2007*, Agusalien
- T1.225** F. Boudin, J-M. Torres, *NEO-CORTEX : a performant user-oriented multi-document summarization system*, 2007, *Proceedings of CICLing-2007, Mexico DF, The Springer LNCS Proceedings 4394, ISSN 0302-9734, ISBN-10: 3-540-70938-X, ISBN-13: 978 3-540-70938-1 pp. 551-562*
- T1.224** G. Linarès, C. Lévy, *Fast discriminative training of semi-continuous HMM*, 2007 I, *International Conférence on Speech, Text and Dialogue (ISCA), Lectures notes on Computer Science, Springer, ISSN 0302-9743, ISBN 978-3-540-74627-0, volume 4629/2007, pp. 398-405, Pilsen, CZ, 2007*
- T1.223** X. Polanco, E. SanJuan, *Hypergraph Modelling and Graph Clustering Process Applied to Co-word Analysis*, 2007, *11th International Conference of the International Society for Scientometrics and Informetrics, Madrid, Spain, June 25-27, 2007. Proceedings, vol. 2, p. 613-618*
- T1.222** E. SanJuan, J-M. Torres, *Textual Energy of Associative Memories: performants applications of ENERTEX algorithm in text summarization and topic segmentation*, 2007, *MICAI 2007, Aguascalientes, Mexique*
- T1.221** B. Fauve, N. Evans, N. Pearson, J-F. Bonastre, J. Mason, *Influence of task duration in text-independent speaker verification*, 2007, *Interspeech 2007*, Antwerp (Belgium)
- T1.220** J-F. Bonastre, D. Matrouf, C. Fredouille, *Artificial impostor voice transformation effects on false acceptance rates*, 2007, *Interspeech'07*, Antwerp, Belgium, September 2007
- T1.219** G. Linarès, D. Massonié, P. Nocera, C. Lévy, *A scalable system for embeded large vocabulary continuous speech recognition*, 2007, *IEEE Workshop on DSP in Mobile and vehicular systems*,

Istanbul, Turkey, 2007

T1.218 D. Matrouf, N. Scheffer, B. Fauve, J-F. Bonastre, A Straightforward and Efficient Implementation of the Factor Analysis Model, *Interspeech 2007*, Antwerp (Belgium)

T1.217 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, J. Revis, Characterization of the pathological voices (dysphonia) in the frequency space, *International Congress on Phonetic Sciences, ICPhS 2007*, Saarbrücken, Germany

T1.216 B. Minescu, G. Damnati, F. Béchet, R. De Mori Conditional use of Word Lattices, Confusion Networks and 1-best string hypotheses in a Sequential Interpretation Strategy, *Interspeech 2007*, Anvers, Belgique, Aout 2007

T1.215 C. Raymond, G. Riccardi, Generative and Discriminative Algorithms for Spoken Language Understanding, 2007, *Interspeech*, Antwerp, Belgium

T1.214 KJ Rodriguez, S. Dipper, M. Götze, M. Poesio, G. Riccardi, C. Raymond, J. Rabiega-Wisniewska, Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus, 2007, *Linguistic Annotation Workshop, LAW2007*

T1.213 C. Raymond, G. Riccardi, KJ Rodriguez, J. Wisniewska, LUNA Corpus: an Annotation Scheme for a Multi-domain Multi-lingual Dialogue Corpus, *Workshop on the Semantics and Pragmatics of Dialogue, DECALOG 2007*,

T1.212 F. Lefèvre, Dynamic Bayesian Networks and Discriminative Classifiers for Multi-stage Semantic Interpretation, 2007, *IEEE International Conference on Speech and Signal Processing (ICASSP'07)*, Hawaiï, USA

T1.211 B. Lecouteux, G. Linarès, Y. Esteve, J. Mauclair, System combination by driven decoding, 2007, *IEEE International Conference on Speech and Signal Processing (ICASSP'07)*, Hawaiï, USA

T1.210 C. Fredouille, N. Evans, The influence of speech activity detection and overlap on speaker diarization for meeting room recordings, 2007, *Interspeech'07*, Antwerp, Belgium, September 2007

T1.209 I. Da Cunha Fanego, Fernández, S., Velázquez, P., Vivaldi, J., E. SanJuan, J-M. Torres, A new hybrid summarizer based on Vector Space Model, *Statistical Physics and Linguistics*, 2007, LNCS 4827. Berlin: Springer. 872-882. ISSN 0302-9743.

T1.208 B. Fauve, N. Evans, N. Pearson, J-F. Bonastre, J. Mason, Influence of task duration in text-independent speaker verification, 2007, *Interspeech 2007*, Antwerp (Belgium)

T1.207 A. Nimaan, P. Nocera, J-F. Bonastre, Towards automatic transcription of Somali language, *LREC 2006*. Genes.

T1.206 A. Nimaan, P. Nocera, J-M. Torres, Boîtes à outils TAL pour les langues peu informatisées : Le cas du Somalï, *JADT 2006*. Besançon.

T1.205 D. Matrouf, J-F. Bonastre, C. Fredouille, Effect of speech transformation on impostor acceptance, 2006, *ICASSP 06*, Toulouse, France, may 2006

T1.204 C. Lévy, G. Linarès, J-F. Bonastre, GMM-based acoustic modeling for embedded speech recognition, 2006, *International Conference on Computer Speech and Language Processing, ICSLP/InterSpeech*, Pittsburgh - USA

T1.203 A. Nimaan, P. Nocera, J-F. Bonastre, Automatic transcription of Somali language, 2006, *Interspeech 2006 - ICSLP*. Pittsburgh PA, USA.

T1.202 B. Lecouteux, G. Linarès, P. Nocera, J-F. Bonastre, Imperfect transcript driven speech recognition, 2006, *International Conference on Computer Speech and Language Processing, ICSLP/InterSpeech*, Pittsburgh - USA

T1.201 A. Nimaan, P. Nocera, J-F. Bonastre, Towards automatic transcription of Somali language, 2006, *LREC 2006*. Genova. Italia.

T1.200 C. Kobus, G. Damnati, L. Delphin Poulat, R. De Mori, Exploiting semantic relations for a Spoken Language Understanding application, 2006, *Proc. International Conference on Spoken Language Processing (Interspeech 2006)*, Pittsburg, PA.

T1.199 R. Gemello, F. Mana, S. Scanzio, P. Laface, R. De Mori, Adaptation Of Hybrid Ann/Hmm Models Using Linear Hidden Transformations And Conservative Training, 2006, *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 06*, Toulouse (France)

T1.198 M. Benzeguiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, Chris, Automatic Speech Recognition and Intrinsic Speech Variation, 2006, *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 06*, Toulouse (France)

- T1.197** L. Barrault, D. Matrouf, R. De Mori, R. Gemello, F. Mana, Dynamic selection of acoustic features in an automatic speech recognition system, *EUropean Signal Processing COncference (EUSIPCO)*, Florence, September 2006
- T1.196** L. Barrault, D. Matrouf, R. De Mori, R. Gemello, F. Mana, Characterizing Feature Variability in Automatic Speech Recognition Systems, 2006, *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 06*, Toulouse (France)
- T1.195** H. Bonneau-Maynard, C. Ayache, F. Béchet, A. Denis, A. Kuhn, F. Lefèvre, D. Mostefa, M. Quignard, S. Rosset, C. Servan, and J. Villaneau, Results of the French Evalda-Media evaluation campaign for literal understanding, 2006, In *Proceedings of the International Conference on Language Resources and Evaluation*
- T1.194** F. Lefèvre, A DBN-based multi-level stochastic spoken language understanding system, 2006, In *Proceedings of IEEE/ACL Workshop on Spoken Language Technology*
- T1.193** C. Servan, C. Raymond, F. Béchet, P. Nocera, Conceptual decoding from word lattices: application to the spoken dialogue corpus MEDIA, 2006, *ISCLP 06*, Pittsburgh, USA
- T1.192** C. Kobus, G. Damnati, L. Delphin-Poulat, R. De Mori, Exploiting semantic relations for a Spoken Language Understanding application, 2006, *ICSLP 2006*, Pittsburgh, Pennsylvania, USA
- T1.191** J-F. Bonastre, D. Matrouf, C. Fredouille, Transfer function-based voice transformation for speaker recognition, 2006, *Speaker Odyssey 06*, June 2006, San Juan, Puerto Rico
- T1.190** A. Preti, J-F. Bonastre, Unsupervised model adaptation for speaker verification, 2006, *ICSLP 2006*, Pittsburgh, PA, USA. p2090-2093
- T1.189** A. Preti, N. Scheffer, J-F. Bonastre, Discriminant approaches for GMM-based speaker detection systems, 2006, Toulouse, France, *MMUA 06*, p50-56
- T1.188** W. Anthony, C. Fredouille, J-F. Bonastre, On the use of linguistics information for Broadcast News, 2006, *ICASSP 06*, Toulouse, France, May 2006
- T1.187** S. Galliano, E. Geoffrois, G. Gravier, J-F. Bonastre, D. Mostefa, K. Choukri, Corpus description of the ESTER Evaluation Campaign for the Rich Transcription of French Broadcast News, *LREC 2006*, 24-26 May 2006, Genoa, ITALY
- T1.186** O. Kraif, M. El-Bèze, R. Meyer, C. Richard, Le corpus Carmel : Un corpus multilingue de récits de voyage, 2006, 7^e *Conférence Internationale Teaching and Language Corpora*
- T1.185** N. Scheffer, J-F. Bonastre, UBM-driven discriminative approach for Speaker Verification, 2006, Puerto Rico, USA, *Odyssey 06* "IBM best paper student award"
- T1.184** N. Scheffer, A multiclass framework for Speaker Verification within an Acoustic Event Sequence system, 2006, *ICSLP 2006*, Pittsburgh, USA. pp 501-504
- T1.183** L. Gillard, P. Bellot, M. El-Bèze, Question Answering Evaluation Survey, 2006, actes de la 5^{ième} conférence Language Resources and Evaluation Conference (LREC), Gênes (Italie), 24-26 mai 2006.
- T1.182** F. Lefèvre, J-L. Gauvain, Discriminant Initialization for Factor Analyzed HMM Training, 2006, In *Proceedings of ICASSP*, Toulouse, May 2006.
- T1.181** F. Béchet, G. Damnati, N. Camelin, R. De Mori, Spoken opinion extraction for detecting variations in user satisfaction, 2006, *IEEE/ACL Workshop on Spoken Language Technology (SLT 06)*, Décembre 2006, Aruba
- T1.180** A. Preti, J-F. Bonastre, F. Capman, A continuous unsupervised adaptation method for speaker verification, 2006, *International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CIS2E 06)*
- T1.179** L. Sitbon, P. Bellot, Tools and methods for topic segmentation of texts and contextual evaluation, 2006, *Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Italie.
- T1.178** B. Favre, F. Béchet, P. Bellot, F. Boudin, M. El-Bèze, L. Gillard, G. Lapalme, J-M. Torres, The LIA-Thales summarization system at DUC-2006, 2006, Actes de la conférence Document Understanding Conference (DUC-2006), New York (USA), 8-9 juin 2006.
- T1.177** N. Camelin, G. Damnati, F. Béchet, R. De Mori, Opinion mining in a telephone survey corpus, 2006, *Proc. International Conference on Spoken Language Processing (ICSLP 06)*, Pittsburg, PA.
- T1.176** A. Nimaan, P. Nocera, J-F. Bonastre, F. Béchet, Speech mining to make African oral patrimony accessible, 2006, *LREC 2006*. Genova. Italia.
- T1.175** C. Fredouille, G. Senay, Technical improvements of the E-HMM based speaker diarization

system for meeting records, 2006, in *Machine Learning for Multimodal Interaction: 3rd International Workshop, MLMI 2006*. Springer LNCS Series

T1.174 G. Cloarec, D. Jouvét, Denis J. Monné, *Analysis of the modeling of pitch and voicing parameters for speaker-independent speech recognition systems*, 65-70. SRIV, ISCA 2006

C-ACTN : communications avec actes dans un congrès national

T1.156 Raphaël Rubino, Iria da Cunha, Georges Linarès, *Extracción automática de léxico bilingüe: experimentos en español y catalán*, ICL 2011, Huelva, Spain

T1.155 Iria da Cunha, Juan-Manuel Torres-Moreno y Gerardo Sierra, *Metodología y desarrollo del primer corpus en español anotado con relaciones retóricas*, XXVII Congreso de la SEPLN, Huelva 5-7 septembre 2011

T1.154 Thierry Bazillon, Benjamin Maza, Mickael Rouvier, Frederic Bechet, Alexis Nasr, *Qui êtes vous ? Catégoriser les questions pour déterminer le rôle des locuteurs dans des conversations orales* TALN 2011

T1.153 R. Rubino, G. Linarès, *Une approche multi-vue pour l'extraction terminologique bilingue*, 2011 Conference en Recherche d'Information et Applications, CORIA, 2011, à paraître, Avignon, France

T1.152 Bonnefoy Ludovic, P. Bellot, Benoit Michel, *Mesure non-supervisée du degré d'appartenance d'une entité à un type*, 2011, TALN 2011, Montpellier (France)

T1.151 R. Deveaud, E. San-Juan, P. Bellot, *Ajout d'informations contextuelles issues de Wikipédia pour la recherche de passages*, 2011, TALN 2011, Montpellier (France)

T1.150 Bonnefoy Ludovic, P. Bellot, Benoit Michel, *Une approche non supervisée pour le typage et la validation d'une réponse à une question en langage naturel : application à la tâche Entity de TREC 2010*, 2011, 8e Conférence en Recherche d'Informations et Applications (CORIA), Avignon (France), mars 2011

T1.149 R. Deveaud, Boudin Florian, P. Bellot, E. San-Juan, *Correction de césures et enrichissement de requêtes par Wikipédia appliqués à la recherche de livres*, 2011, 8e Conférence en Recherche d'Informations et Applications (CORIA), Avignon (France), mars 2011

T1.148 S. Huet, F. Boudin, J-M. Torres, *Utilisation d'un score de qualité de traduction pour le résumé multi -document cross-lingue*, 2011, TALN 2011, Montpellier (France)

T1.147 S. Huet, F. Lefèvre, *Alignement automatique pour la compréhension littérale de l'oral par approche segmentale*, 2011, TALN 2011, Montpellier (France), article court

T1.146 B. Jabaian, L. Besacier, F. Lefèvre, *Comparaison et combinaison d'approches pour la portabilité vers une nouvelle langue d'un système de compréhension de l'oral*, TALN 2011, Montpellier (France)

T1.145 S. Oger, V. Popescu, G. Linarès, *Modèles de langage probabilistes et possibilistes basés sur le Web*, 2010, Journées d'Etude sur la Parole, JEP 2010, Mons (Belgique)

T1.144 Stanislas Oger, Mickael Rouvier, Nathalie Camelin, Rémy Kessler, Fabrice Lefèvre et Juan-Manuel Torres-Moreno, *Système du LIA pour la campagne DEFT'10 : datation et localisation d'articles de presse francophones* DEFT 2010

T1.143 G. Senay, G. Linarès, B. Lecouteux, S. Oger, Thierry Michel, *Décodage interactif de la parole*, 2010 Journées d'Etude sur la Parole, JEP 2010, Mons (Belgique)

T1.142 Mickael Rouvier, Georges Linarès et Driss Matrouf, *Identification du genre vidéo à la volée par combinaison de paramètres acoustiques* JEP 2010

T1.140 F. Pinault, F. Lefèvre, R. De Mori, *Structures de frames sémantiques pour le dialogue Homme-Machine par processus de décision markoviens*, 2010 JEP (Journées d'étude de la parole)

T1.139 J. Kahn, Rossato Solange, J-F. Bonastre, *Modéliser un locuteur : Influence des signaux d'apprentissage sur les performances d'un système de RAL*, 2010, JEP 2010, Mons

T1.138 P. Gotab, F. Béchet, Damnati G., *Stratégies d'apprentissage actif pour l'adaptation de modèles de compréhension dans un système de dialogue oral déployé*, 2010, Journée d'étude de la parole 2010, JEP, Mons, Belgique

T1.137 S. Oger, M. Rouvier, G. Linarès, *Classification du genre vidéo reposant sur des transcriptions automatiques*, TALN, 2010, Montréal

T1.136 A. Larcher, C. Lévy, D. Matrouf, J-F. Bonastre, *Reconnaissance Automatique du Locuteur*

embarquée dans un téléphone portable, JEP 2010

T1.135 G. Sierra, J-M. Torres, Regroupement sémantique de définitions en espagnol, 2010, EGC/EVALECD

T1.134 J. Kahn, N. Audibert, S. Rossato, J-F. Bonastre, Modéliser un locuteur : Influence des signaux d'apprentissage sur les performances d'un système de RAL, 2010, JEP, Mons

T1.133 I. Da Cunha Fanego, Los diccionarios de términos inmobiliarios en español: carencias y deficiencias, 2009, Ahumada Lara, Ignacio (coord.). *Metalexicografía variacional: diccionarios de regionalismos y diccionarios de especialidad*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.

T1.132 I. Da Cunha Fanego, Ferraro, G., Cabré, M.T., *Propuesta de etiquetaje discursivo y sintáctico-comunicativo orientado a la evaluación de un modelo lingüístico de resumen automático*, 2009, Bretones, M. C. et al. (eds). *Applied Linguistics Now: Understanding Language and Mind*. Almería: Universidad de Almería. 1333-1347. ISBN: 978-84-692-1479-4.

T1.131 B. Lecouteux, P. Nocéra, G. Linarès, Décodage guidé par un modèle cache sémantique, 2010, JEP Mons, Belgique

T1.130 G. Senay, G. Linarès, B. Lecouteux, S. Oger, T. Michel, Décodage interactif de la parole, 2010, JEP, Mons, Belgique

T1.129 S. Oger, V. Popescu, G. Linarès, Modèles de langage probabilistes et possibilistes basés sur le Web, 2010, JEP, Mons, Belgique

T1.128 F. Pinault, F. Lefèvre, R. De Mori, Structures de frames sémantiques pour le dialogue Homme-Machine par processus de décision markoviens, 2010, JEP

T1.127 T. Waszak, C. de Loupy, P. Bellot, Identification et structuration hiérarchique des titres dans les documents HTML, 2009, 6è CORIA, Toulon (France)

T1.127 R. Lavalley, P. Bellot, M. El-Bèze, Interactions entre le calcul de collocations et la catégorisation automatique de textes, 2009, 6è CORIA, Toulon (France)

T1.126 I. Da Cunha Fanego, Iruskieta, M., La influencia del anotador y las técnicas de traducción en el desarrollo de árboles retóricos. Un estudio en español y euskera, 2009, *Proceedings of 7th Brazilian Symposium in Information and Human Language Technology (STIL 2009)*. São Carlos, Brasil: Universidade de São Paulo. ISSN 2175-6201.

T1.125 E. Charton, Combinaison de contenus encyclopédiques multilingues pour une reconnaissance d'entités nommées en contexte, 2009, "RECITAL 2009"

T1.122 E. Charton, J-M. Torres, Classification d'un contenu encyclopédique en vue d'un étiquetage par entités nommées, 2009, TALN 2009, Senlis, France

T1.121 J-M. Torres, M. El-Bèze, F. Béchet, N. Camelin, Fusion probabiliste appliquée à la détection et classification d'opinions, 2009, DEFT'09

T1.120 R. Kessler, J-M. Torres, M. El-Bèze, Profilage de candidatures assisté par Relevance Feedback, 2009, TALN 2009, Senlis, France

T1.120 C. Vidrequin, J-M. Torres, J-J. Schneider, M. El-Bèze, Génération et enrichissement automatique de listes de patrons de phrases pour les moteurs de questions-réponses, 2007, 7ème Conférence Extraction et Gestion des Connaissances (EGC 2007), Namur, Belgique 23-26 janvier 2007

T1.119 M. Rouvier, G. Linarès, D. Matrouf, Identification robuste du genre vidéo par l'audio, 2009, MajecSTIC, Avignon, France

T1.118 B. Lecouteux, G. Linarès, B. Favre, Détection des mots hors vocabulaires par combinaison de mesures de confiances haut et bas niveaux, 2009, MajecSTIC, Avignon, France

T1.117 H. Mauchretien, G. Linarès, T. Jimenez, C. Fredouille, Identification Musicale par des technologies de traitement automatique de la parole, 2009, "MajecSTIC",

T1.116 L. Bonnefoy, R. Deveaud, E. Charton, Interrogations de moteurs de recherche par des requêtes formulées en langage naturel, 2009, MajecSTIC,

T1.115 P. Gotab, Apprentissage automatique et Co-training, 2009, TALN/RECITAL 2009, Senlis, France

T1.114 S. Fernandez, E. SanJuan, J-M. Torres, Résumés de textes par extraction de phrases, algorithmes de graphe et énergie textuelle, 2009, XVI èmes rencontres de la Société Francophone de Classification

T1.113 B. Lecouteux, G. Linarès, B. Favre, Détection de mots hors-vocabulaire par combinaison de

mesures de confiance de haut et bas niveaux, 2009, MajecSTIC'09, (Avignon, France)

T1.112 L. Sitbon, P. Bellot, P. Blache, Lisibilité et recherche d'information : vers une meilleure accessibilité, 2008, 5è CORIA, Trégastel (France), mars 2008

T1.111 F. Duvert, M-J. Meurs, C. Servan, F. Béchet, F. Lefèvre, R. De Mori, Composition sémantique pour la compréhension de la parole dans le cadre de dialogue, 2008, JEP

T1.110 Gillard, P. Bellot, M. El-Bèze, Quelles combinaisons de scores et de critères numériques pour un système de Questions/Réponses ?, Actes de TALN 2008, Avignon (France)

T1.109 F. Poulard, T. Waszak, N. Hernandez, P. Bellot, Repérage de citations, classification des styles de discours et identification des constituants, citationnels en écrits journalistiques, actes de TALN 2008, Avignon (France)

T1.108 A. Prete, F. Capman, B. Ravera, J-F. Bonastre, Surveillance vocale de réseaux de communications professionnels par la reconnaissance du locuteur, 2008 JEP

T1.107 M-J. Meurs, F. Lefèvre, R. De Mori, Interprétation du dialogue oral : pour une approche bayésienne de la composition sémantique., 2008, MajecSTIC, (Best Paper Award)

T1.106 M-J. Meurs, F. Duvert, F. Béchet, F. Lefèvre, R. De Mori, Annotation en Frames Sémantiques du corpus de dialogue MEDIA, 2008, TALN

T1.104 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Giovanni, Analyse Phonétique dans le Domaine Fréquentiel pour la Classification des Voix Dysphoniques, 2008, XXVIIème JEP, JEP'08, Avignon, France

T1.103 S. Oger, G. Linarès, F. Béchet, P. Nocera, Enrichissement dynamique du vocabulaire à partir du Web, 2008, "Actes des JEP, JEP 2008 ",

T1.102 R. Kessler, J-M. Torres, M. El-Bèze, E-Gen : classification automatique des CV, TALN 2008, Avignon, France

T1.101 C. Vidrequin, J-M. Torres, J-J. Schneider, M. El-Bèze, Extraction automatique d'information à partir de microtextes non structurés, TALN 2008, Avignon 2008

T1.100 B. Lecouteux, G. Linarès, Y. Estève, G. Gravier, Combinaison de systèmes par décodage guidé, 2008, Journées d'études de la Parole (JEP)

T1.99 L. Barrault, D. Matrouf, G. Linarès, Combinaison de différents jeux de paramètres acoustiques pour la reconnaissance de la parole, 2008, "Journées d'études de la Parole (JEP)",

T1.98 C. Lévy, G. Linarès, J-F. Bonastre, Adpatation rapide de modèles acoutisques compacts, 2008, "Journées d'études de la Parole (JEP)"

T1.97 E. Charton, N. Camelin, R. Acuna-Agost, P. Gotab, R. Lavalley, R. Kessler, S. Fernandez, Pré-traitements classiques ou par analyse distributionnelle: application aux méthodes de classification automatique déployées pour DEFT08, 2008, TALN08 - DEFT08

T1.96 S. Fernandez, E. SanJuan, J-M. Torres, Enertex: un système basé sur l'énergie textuelle, 2008, Traitement Automatique de la Langue Naturelle (TALN 2008)

T1.95 H-Q. Nguyen, P. Nocera, E. Castelli, Trinh Van Loan, Reconnaissance de la parole continue à grand vocabulaire en vietnamien, une langue syllabique tonale, 2008, JEP 2008

T1.94 N. Camelin, F. Béchet, R. De Mori, G. Damnati, 2008, Analyse des erreurs d'une stratégie automatique de sondage d'opinions, Dans les actes des Journées d'Étude sur la Parole, Juin, Avignon

T1.93 Vincent Jousse, Y. Estève, F. Béchet, T. Bazillon, G. Linarès, Caractérisation et détection de parole spontanée dans de larges collections de documents audio, 2008, Journées d'études de la Parole,

T1.92 E. Charton, Teva Merlin, C. Lévy, A. Larcher, S. Meigner, J-F. Bonastre, L. Besacier, J. Farinas, B. Ravera, Mistral : plateforme d'authentification biométrique, 2008, JEP

T1.91 A. Larcher, J-F. Bonastre, J. S.D. Mason, Utilisation de la structure de mots de passe personnalisés pour la reconnaissance de locuteurs embarquée, JEP, 2008

T1.90 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Giovanni, Analyse Phonétique dans le Domaine Fréquentiel pour la Classification des Voix Dysphoniques, 2008, XXVIIe JEP, JEP'08, Avignon, France

T1.89 I. Da Cunha Fanego, Importancia del marcaje de las relaciones discursivas para la generación automática de resúmenes, 2007, Cano, P.; Fernández, I.; González, M.; Prego, G.; Souto, M. (eds.) Proceedings of the 5th General Linguistics Conference. Santiago de Compostela: Arco Libros. 835-846. ISBN 978-84-7635-669.

- T1.88** S. Fernandez, E. SanJuan, J-M. Torres, *Energie textuelle de mémoires associatives*, TALN 2007 Vol 1, pp 25-34
- T1.87** I. Da Cunha Fanego, Llopis, J., *Constatación de la validez de los resúmenes adjuntos a artículos médicos de investigación de cara a la evaluación de resúmenes automáticos*, 2007, Mairal, R. et al. (eds.) *Language Learning, Language Use, and Cognitive Modelling: Applied Perspectives Across Disciplines*. Madrid: UNED. ISBN 978-84-611-6897-2.
- T1.86** L. Sitbon, L. Gillard, *Evaluation modulaire d'un système de questions-réponses sur un corpus de questions semi-spontanées*, 2007, Actes de la 4^e CORIA 07
- T1.85** L. Gillard, P. Bellot, M. El-Bèze, *D'une compacité positionnelle à une compacité probabiliste pour un système de Questions/Réponses*, 2007, actes de la 4^e CORIA, Saint-Etienne (France), mars 2007
- T1.84** L. Sitbon, P. Bellot, P. Blache, *Traitements phrastiques phonétiques pour la réécriture de phrases dysorthographiées*, 2007, Actes de TALN 2007, Toulouse (France)
- T1.83** Gillard, P. Bellot, M. El-Bèze, *Analyse des échecs d'une méthode pour traiter les questions définitoires soumises à un système de Questions/Réponses*, 2007, actes de TALN 2007, Toulouse (France)
- T1.82** A. Denis, F. Béchet, M. Quignard, *Résolution de la référence dans des dialogues homme-machine : évaluation sur corpus de deux approches symbolique et probabiliste*, TALN 07, Juin 2007
- T1.81** L. Sitbon, *Combinaison de ressources linguistiques pour l'aide à l'accès lexical : étude de faisabilité*, 2007 Actes de la conférence RECITAL
- T1.80** R. Lavalley, R. Rubino, *Mise en oeuvre de méthodes de TAL afin d'aider les enseignants dans l'élaboration d'exercices d'apprentissage de la lecture pour enfants dyslexiques*, 2008, actes de MajecSTIC 2008, Marseille, France, p. 45 à 53
- T1.79** J-M. Torres, M. El-Bèze, F. Béchet, N. Camelin, *Comment faire pour que l'opinion forgée à la sortie des urnes soit la bonne ? Application au défi DEFT 2007*, DEFT07, pp 119-133, Plate-forme AFIA 2007, Grenoble.
- T1.78** A. Preți, *Utilisation de mesures de confiance pour l'adaptation non supervisée des modèles de locuteurs en vérification du locuteur*, 2007, Rencontre Jeunes Chercheurs en Parole 2007
- T1.77** N. Camelin, F. Béchet, G. Damnati, R. De Mori, *Analyse automatique de sondages téléphoniques d'opinion*, 2007, TALN 07
- T1.76** N. Flavier, P. Bellot, *Vers un appariement automatique de questions extraites de courriers électroniques*, 2007, Conférence Francophone sur l'Apprentissage Automatique (CAP 2007), Grenoble (France)
- T1.75** G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, M. Azzarello, A. Giovanni, *Modélisation Statistique et Informations Pertinentes pour la Caractérisation des Voix Pathologiques (Dysphonies)*, 2006 JEP, 12-15 Mai 2006, Dinard, France
- T1.74** J-F. Bonastre, D. Matrouf, C. Fredouille, *Augmentation du taux de fausse acceptation par transformation inaudible de la voix des imposteurs*, 2006, JEP, 12-15 Mai 2006, Dinard, France
- T1.73** B. Lecouteux, G. Linarès, P. Nocera, J-F. Bonastre, *Reconnaissance de la parole guidée par des transcriptions approchées*, 2006, JEP 06, Dinard (France)
- T1.72** A. Nimaan, P. Nocera, J-F. Bonastre, *Reconnaissance automatique de la parole en langue somalienne*, 2006, Journées d'Etudes sur la Parole 2006. Dinard. France.
- T1.71** C. Servan, C. Raymond, F. Béchet, P. Nocera, *Décodage conceptuel à partir de graphes de mots sur le corpus de dialogue Homme-Machine MEDIA*, 2006, JEP, Dinard, Juin 2006
- T1.70** H. Bonneau-Maynard, F. Lefèvre, *Un modèle stochastique de compréhension de la parole à 2+1 niveaux*, 2006, In Actes des XXVies JEP
- T1.69** *Utilisation des transducteurs dans le décodage conceptuel application au corpus MEDIA*, 2006, MajecSTIC, Lorient, Novembre 2006
- T1.68** G. Linarès, C. Lévy, J.C. Plagniol, *Estimation rapide de modèles semi-continus discriminants*, 2006 JEP 06, Dinard - France
- T1.67** C. Lévy, G. Linarès, J-F. Bonastre, *Représentation acoustique compacte pour un système de reconnaissance de la parole embarquée*, 2006, JEP 06, Dinard - France
- T1.66** L. Sitbon, J. Grivolla, L. Gillard, P. Bellot, P. Blache, *Vers une prédiction automatique de la difficulté d'une question en langue naturelle*, 2006, actes de la 13^{ième} conférence TALN, Louvain (Belgique), 10-13 avril 2006, pp. 337 à 346.

- T1.65** C. Kobus, G. Damnati, L. Delphin-Poulat, *Mesure de confiance de relation sémantique dans le cadre d'un modèle de langage sémantique*, 2006, JEP 2006, Dinard, France
- T1.64** L. Gillard, P. Bellot, M. El-Bèze, *Influence de mesures de densité pour la recherche de passages et l'extraction de réponses dans un système de questions-réponses*, actes de la 3^e CORIA, Lyon (France), 15-17 mars 2006, pp. 193-204.
- T1.63** L. Gillard, P. Bellot, M. El-Bèze, *Questions Booléennes : Oui ou Non, des Questions et des Réponses*, 2006, actes de la 13^e conférence TALN, Louvain (Belgique), 10-13 avril 2006, pp. 159 à 166.
- T1.62** F. Lefèvre, J-L. Gauvain, *Transformation linéaire discriminante pour l'apprentissage des HMM à analyse factorielle*, actes des XXV^{es} JEP, 2006
- T1.61** N. Camelin, G. Damnati, F. Béchet, R. De Mori, *Détection automatique d'opinions dans des corpus de messages oraux*, 2006, JEP, Dinard
- T1.60** B. Favre, J-F. Bonastre, P. Bellot, F. Capman, *Accès aux connaissances orales par le résumé automatique*, 2006, 6^e journées francophones "Extraction et Gestion des Connaissances" EGC 2006, Lille (France), janvier 2006
- T1.59** G. Cloarec, D. Juvet, J. Monné, *Influence de la corrélation entre le pitch et les paramètres acoustiques en reconnaissance de la parole*, JEP 2006, Nancy
- T1.58** I. Da Cunha Fanego, Wanner, L., *Resumen automático de artículos médicos en castellano: integración de técnicas de análisis textual, léxico, discursivo y sintáctico-comunicativo*, 2006 Proceedings of the 7th General Linguistics Conference. Barcelona: Publicacions i Edicions de la Universitat de Barcelona. ISBN 84-475-2086-8.

C-INV (Conférences invités)

- T1.43** G. Linarès, *Speech Processing*, 2010, Second bootcamp of the European Network of Excellence PASCAL (Pattern Analysis, Statistical modelling and Computational Learning), Marseille
- T1.42** E. Charton, *Application de la Génération Automatique de textes au traitement des langues Africaines*, Workshop AfricanHLT, Djibouti, Janvier 2010
- T1.41** E. Charton, *NLGBase, a statistical ontology exploiting the Wikipedia content*, Workshop Wikimania09, Buenos Aires, août 2009,
- T1.40** J.-F. Bonastre, *Vocal Identification*, YESS 2009 (Colloque annuel de la Mission pour la Science et la Technologie, Ambassade de France), Washington, Juillet 2009
- T1.39** G. Linarès, *Automatic speech recognition : is this the end ?*, 2009, Keynote, Young researchers Workshop in Speech Technology, Dublin, Irlande
- T1.38** M. El-Bèze, J-M. Torres, P. Bellot, *Variations autour du Résumé Automatique de Textes*, 2008, Inria Grenoble, Séminaires InTech, 2008
- T1.37** R. De Mori, *Probabilistic frame instantiations for spoken language understanding*, 2008, "LangTech 2008, Rome, Italy",
- T1.36** R. De Mori, *Spoken language understanding: a survey*, 2007, "Keynote speaker, IEEE - ASRU Kyoto, Japan,
- T1.35** J-F. Bonastre, *Le traitement automatique de la parole : une technique d'investigation ?*, Journées Scientifiques IUF 2007, Avignon, Mai 2007
- T1.34** J-F. Bonastre, *Recherche d'Information dans des documents audio radiodiffusés*, Ecole d'été ERMITES, Septembre 2007
- T1.33** J-F. Bonastre, *Evaluation paradigm: Advantages and limits*, 2006, MMUA Workshop, Toulouse, May 2006
- T1.32** J-F. Bonastre, *Reconnaissance du Locuteur Indexation Reconnaissance du Locuteur et Indexation de Documents Audio*, 2006, ERMITES 2006 (Ecole été Recherche Multimodale Information), Giens, Septembre 2006
- T1.31** P. Bellot, *Traitement automatique des langues et classification automatique : méthodes et applications pour la recherche d'informations*, 2006, RIAS 2006, Lyon, mars 2006 (<http://eric.univ-lyon2.fr/~rias2006/>)
- T1.30** J.F. Bonastre, *La Voix pour l'identification des personnes, Sécurité dans les transports, inrets*, Paris, 2006

C-COM (Communications orales sans actes dans un congrès national ou international)

T1.30 Pouchoulin Gilles, A. Ghio, L. Crevier-Buchman, C. Fougeron, C. Fredouille, S. Pinto, A. Giovanni, D. Robert, A. Simon, B. Teston, F. Viallet, *Préconisations pour la conservation et l'exploitation de données collectées pour l'étude des troubles de la voix et de la parole*, **2011**, 4e Journées de Phonétique Clinique (JPC4), 19-21 mai, Strasbourg, France.

T1.29 Pouchoulin Gilles, C. Fredouille, C. Fougeron, A. Ghio, N. Lévêque, C. Meunier, O. Panseri, *Analyse par bandes de fréquences pour la détection des zones de «déviance» dans la parole dysarthrique*, **2011**

4e Journées de Phonétique Clinique (JPC4), 19-21 mai, Strasbourg, France

T1.28 C. Fredouille, C. Fougeron, O. Panseri, N. Lévêque, *Automatic speech processing for the detection of abnormal speech zones in dysarthric speech*, **2010**, 13th Meeting of the International Clinical Phonetics and Linguistics Association (ICPLA), Oslo, 23-26 jun, 2010 (sur résumé)

T1.27 E. Charton, *Mistral Plateforme d'authentification biométrique*, 2007, "Grand Colloque Stic", ANR, Paris 2007

T1.26 L. Crevier Buchman, C. Fougeron, C. Fredouille, A. Ghio, C. Meunier, et al., *A new approach for an acoustic-phonetic description of dysarthria*, 2010, 2nd International Symposium Basal Ganglia Speech Disorders Deep & Brain Stimulation, Aix en Provence

T1.25 P. Clément, C. Fredouille, N. Lévêque, *Méthodes objectives appliquées à la dysarthrie*, 2009, Troisièmes Journées Phonétique Clinique (JPC3) - Aix-en-Provence - 4 et 5 décembre 2009

T1.24 C. Fredouille, G. Pouchoulin, J-F. Bonastre, A. Ghio, A. Giovanni, *Pertinence des consonnes sourdes pour l'observation des phénomènes liés à la dysphonie ?*, 2009, Troisièmes Journées Phonétique Clinique (JPC3) - Aix-en-Provence - 4 et 5 décembre 2009

T1.23 P. Aventurier, H. Leiser, H. Richard, P. Bellot, *OntoFruit : Ecophysiologie végétale de l'arboriculture fruitière - un référentiel documentaire indexé par une ontologie du domaine*, 2008

T1.22 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Giovanni, *Frequency-Based Analysis for the Characterization of the Dysphonic Voices*, 2007, PEVOC 07: Pan European Voice Conference, Groningen, The Netherlands

T1.21 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Giovanni, *Analyse fréquentielle pour la caractérisation des voix dysphoniques*, 2007, 2^e Journées Phonétique Clinique (JPC2) - Grenoble - 13 et 14 Déc. 2007

T1.20 J. Révis, O. Champavère, C. Fredouille, A. Giovanni, *Evaluation de la dysphonie à partir de techniques issues de la reconnaissance du locuteur*, 2007, 2^e Journées Phonétique Clinique (JPC2) - Grenoble - 13 et 14 Décembre 2007

T1.17 Erick Galani Maziero, Thiago Alexandre Salgueiro Pardo, Iria da Cunha, Juan-Manuel Torres-Moreno, Eric SanJuan, *DiZer 2.0 – An Adaptable On-line Discourse Parser*, III Workshop "A RST e os Estudos do Texto" STIL'11, 24-26 octobre 2011, Cuiabá, Mato Grosso (Brésil).

T1.16 Brenda Gabriela Castro Roln, Gerardo Sierra Martinez, Juan-Manuel Torres-Moreno, Iria da Cunha Fanego, *El discurso y la semantica como recursos para la deteccion de similitud textual*, III Workshop "A RST e os Estudos do Texto" STIL'11, 24-26 octobre 2011, Cuiaba, Mato Grosso (Brésil).

T1.15 Iria da Cunha, Juan-Manuel Torres-Moreno, Gerardo Sierra, *On the Development of the RST Spanish Tree-bank*, 5th Linguistic Annotation Workshop, ACL, June 23-24, 2011 : LAW Workshop, Portland, Oregon.

T1.13 I. Da Cunha Fanego, J-M. Torres, *Automatic Discourse Segmentation : Review and Perspectives*, **2010**, African HLT, Djibouti

T1.12 J-M. Torres, Javier Ramirez, I. Da Cunha Fanego, *Un resumeur a base de graphes, indépendant de la langue*, 2010, African HLT, Djibouti

T1.11 F. Verdet, D. Matrouf, J-F. Bonastre, *NIST LRE 2009 - LIA System Description*, NIST Language Recognition Evaluation 2009 Workshop

T1.10 I. Da Cunha Fanego, J-M. Torres, *Automatic Discourse Segmentation : Review and Perspectives*, 2010, African HLT, Djibouti

T1.9 J-M. Torres, J. Ramirez, I. Da Cunha Fanego, *Un résuméur à base de graphes, indépendant de la langue*, 2010, African HLT, Djibouti

T1.8 F. Béchet, E. Charton, *Ester2 entités nommées*, Participation à la campagne ESTER 2, 2009

T1.7 G. Pouchoulin, C. Fredouille, J-F. Bonastre, A. Ghio, A. Marques, A. Giovanni, *Are the unvoiced consonants relevant for dysphonia phenomenon observation ?*, 2009, 3rd Advanced Voice Function

Assessment International workshop - AVFA'09

T1.6 Co. Fredouille, N. Evans, *The LIA-EURECOM RT'09 Speaker Diarization System*, 2009, NIST Rich Transcription (RT) Workshop, Maryland, mai 2009

T1.5 H-Q. Nguyen, P. Nocera, E. Castelli, T. Van Loan, *Large vocabulary continuous speech recognition for Vietnamese, an under-resourced language*, 2008, SLTU, Spoken Languages Technologies for Under-resourced languages, 2008

T1.4 E. Charton, R. Acuna-Agost, *Quel modèle pour détecter une opinion ? Trois propositions pour généraliser l'extraction d'une idée dans un corpus*, 2007, Actes de Deft'07, Défi fouille de texte

T1.3 C. Richard, R. Meyer, M. El-Bèze, *Projet CARMEL : récits de voyages*, 2006, Atelier Technolanguage, TALN 2006, Louvain

T1.2 D. Matrouf, L. Barrault, R. De Mori, *A General Method for Combining Acoustic Features in an Automatic Speech Recognition System*, 2006, Speech Recognition and Intrinsic Variation (SRIV) Workshop, Sept. 2006, Toulouse, France

T1.1 Gillard, L. Sitbon, E. Blaudez, P. Bellot, M. El-Bèze, *The LIA at QA@CLEF2006*, 2006, dans les Working Notes of the Cross Language Evaluation Forum (CLEF) 2006, Alicante (Espagne), 20-22 septembre 2006.

Publications effectuées par Stéphanne Huet dans son établissement d'origine

OS

T1.476 Stéphanne Huet, Gwénolé Lecorvé, Guillaume Gravier et Pascale Sébillot. *Multimodal Processing and Interaction: Audio, Video, Text*. Petros Maragos, Alexandros Potamianos, Patrick Gros (eds.). Chapitre *Toward the Integration of Natural Language Processing and Automatic Speech Recognition: Using Morpho-syntax and Pragmatics for Transcription*. pages 201-218, Springer, 2008.

ACTN

T1.452 Stéphanne Huet, Guillaume Gravier et Pascale Sébillot. *Morpho-Syntactic Post-Processing of N-Best Lists for Improved French Automatic Speech Recognition*. *Computer Speech and Language*, Volume 12(4), pages 663-684, 2010.

T1.448 J. Bourdaillet, S. Huet, P. Langlais, G. Lapalme, *TransSearch: from a Bilingual Concordancer to a Translation Finder*, 2010, *Machine Translation Journal*, 24(3-4), pp. 241-271

C-ACTI

T1.373 Stéphanne Huet, Guillaume Gravier et Pascale Sébillot. *Morphosyntactic Processing of N-Best Lists for Improved Recognition and Confidence Measure Computation*. *Proceedings of the 8th Annual Conference of the International Speech Communication Association (Interspeech)*, pages 1741-1744, Anvers, Belgique, Août 2007.

T1.375 Stéphanne Huet, Julien Bourdaillet, Philippe Langlais et Guy Lapalme. *Harnessing the Redundant Results of Translation Spotting*. *Proceedings of the 12th Machine Translation Summit (MT Summit)*, 8 pages, Ottawa, Canada, Août 2009.

T1.374 Stéphanne Huet, Julien Bourdaillet et Philippe Langlais. *TS3: an Improved Version of the Bilingual Concordancer TransSearch*. *Proceedings of the 13th Annual Conference of the European Association for Machine Translation (EAMT)*, pages 20-27, Barcelone, Espagne, Mai 2009.

T1.14 Stéphanne Huet, Julien Bourdaillet, Alexandre Patry et Philippe Langlais. *The RALI Machine Translation System for WMT 2010*. *Proceedings of the ACL Joint 5th Workshop on Statistical Machine Translation and Metrics (WMT)*, 7 pages, Uppsala, Suède, Juillet 2010.

C-ACTN

T1.141 Julien Bourdaillet, Stéphanne Huet et Philippe Langlais. *Alignement de traductions rares à l'aide de paires de phrases non alignées*. Actes de la 17ème conférence sur le Traitement Automatique des Langues Naturelles (TALN), 10 pages, Montréal, Canada, Juillet 2010.

T1.124 Stéphane Huet, Julien Bourdaillet et Philippe Langlais. *Intégration de l'alignement de mots dans le concordancier bilingue TransSearch*. Actes de la 16ème conférence sur le Traitement Automatique des Langues Naturelles (TALN), 10 pages, Senlis, France, Juin 2009.

T1.123 Julien Bourdaillet, Stéphane Huet, Fabrizio Gotti, Guy Lapalme et Philippe Langlais. *Enhancing the Bilingual Concordancer TransSearch with Word-level Alignment*. Proceedings of the 22nd Conference of the Canadian Society for Computational Studies of Intelligence (Canadian AI), Lecture Notes in Computer Science, Volume 5549/2009, pages 27-38, Kelowna, Canada, Mai 2009.

T1.105 Stéphane Huet, Guillaume Gravier et Pascale Sébillot. *Un modèle multi-sources pour la segmentation en sujets de journaux radiophoniques*. Actes de la 15ème conférence sur le Traitement Automatique des Langues Naturelles (TALN), pages 49-58, Avignon, France, Juin 2008.