

## Formation universitaire

2005	Doctorat	<b>Informatique</b> , mention très honorable Laboratoire d'Informatique d'Avignon, Université d'Avignon
2001	D.E.A.	<b>Informatique</b> , mention assez bien Département Informatique de Luminy, Marseille
2000	Ingénieur-maître	<b>Génie Mathématique et Informatique</b> (GMI), spécialité RÉSEAU, mention assez bien Institut Universitaire Professionnalisé (IUP), Université d'Avignon
2000	Maîtrise-Licence-DEUG	<b>Génie Mathématique et Informatique</b> Institut Universitaire Professionnalisé (IUP), Université d'Avignon
1997	D.U.T.	<b>Génie des Télécommunications &amp; Réseaux</b> (GTR) Université d'Aix-Marseille II

## Parcours et travaux de recherche

### Résumé court

#### Parcours :

- 2009 Maître de Conférence** à l'INSA de Rennes
- 2008 Chercheur** au Laboratoire Informatique d'Avignon
- 2007 ATER** au Laboratoire Informatique d'Avignon
- 2006 Post-Doc** à l'université de Trento, Italy
- 2005 ATER** demi poste au Laboratoire Informatique d'Avignon
- 2002 Doctorant** au Laboratoire Informatique d'Avignon

#### Thèmes de recherches en quelques mots :

- ✓ **Reconnaissance de la parole** : modélisation du langage, mesures de confiance
- ✓ **Compréhension de la parole spontanée** : Méthodes du TALN appliquées à la parole
- ✓ **Apprentissage automatique** : Modèles de Markov, Machines à vecteurs de support (SVM), Champs conditionnels Aléatoires (CRF), Arbres de décision (SCT), [algorithmes de Boosting](#), Apprentissage actif (Active Learning).

## Résumé détaillé

### D.E.A. EN INFORMATIQUE AU LABORATOIRE INFORMATIQUE D'AVIGNON, L.I.A.

J'ai effectué mon mémoire de D.E.A. [35] dans la thématique « recherche documentaire » au Laboratoire Informatique d'Avignon (LIA) sous la direction de Patrice Bellot, Maître de conférences à l'UAPV et de Marc El-Bèze Professeur.

Ces travaux ont fait l'objet d'une publication dans la conférence nationale RFIA en 2002 [33]. Le mémoire complet ainsi que cette publication sont disponibles sur ma page WEB, en voici le résumé : « Une difficulté majeure dans l'utilisation d'un système de recherche documentaire est le choix du vocabulaire à employer pour exprimer une requête. L'enrichissement de la requête peut prendre plusieurs formes : ajout de mots extraits automatiquement des documents rapportés, réestimation des poids attribués à chacun des mots de la requête initiale, etc. Le système de recherche documentaire SIAC est utilisé pour extraire un premier jeu de documents à partir d'une requête. Une méthode de classification non supervisée, à base d'arbres de décision, est ensuite exploitée pour classer les phrases des documents trouvés et les documents eux-mêmes, selon qu'elles/ils contiennent ou non certains mots extraits automatiquement de l'ensemble des documents rapportés. à chaque nœud de l'arbre, peut être associée une expression booléenne mettant en jeu les mots sélectionnés lors de la classification. Nous montrons, à l'aide des données de la seconde campagne d'évaluation Amaryllis, que la réécriture de la requête suivant les expressions booléennes correspondant aux meilleures feuilles permet d'améliorer la précision de la recherche documentaire. La réécriture de la requête, au vu de la structure des arbres de décision, amène à se pencher sur le traitement de la négation dans les systèmes de recherche ainsi qu'à une reformulation des critères de pondérations habituellement utilisés. »

### THÈSE DE DOCTORAT EN INFORMATIQUE AU LABORATOIRE INFORMATIQUE D'AVIGNON, L.I.A., EN COLLABORATION AVEC FRANCE TÉLÉCOM RECHERCHE ET DÉVELOPPEMENT

#### Contexte

Mes travaux de thèse [34] ont été menés dans l'équipe « Dialogue » du Laboratoire Informatique d'Avignon de l'Université d'Avignon et des pays de Vaucluse. Ces travaux concernent la reconnaissance de la parole et la compréhension dans les systèmes de dialogue oral. Ils ont été effectués en collaboration avec France Télécom Recherche et Développement (FTRD), Lannion, qui a financé la thèse à travers le contrat n° 021B178.

Thèse soutenue devant un jury composé de :

Jean-Paul HATON	Professeur	LORIA, Nancy	Rapporteur
Giuseppe RICCARDI	Professeur	DIT, Trento (Italie)	Rapporteur
Marc EL-BÈZE	Professeur	LIA, Avignon	Examineur
Géraldine DAMNATI	Ingénieur	France Télécom R&D, Lannion	Examineur
Renato DE MORI	Professeur	LIA, Avignon	Directeur de Thèse
Frédéric BÉCHET	Maître de Conf.	LIA, Avignon	Co-directeur de Thèse

#### Travaux de recherche

Les systèmes de dialogue oral homme-machine fournissent des services (consultation d'horaires d'avions ou de trains, consultation de la météo, recherche de restaurants, ...) à des utilisateurs, tout en leur offrant la possibilité de s'exprimer en langage naturel. Ces systèmes sont couplés avec une base de données en relation avec le service fourni. La difficulté principale est de comprendre le sens des paroles de l'utilisateur. Pour effectuer cette tâche, la plupart des systèmes de ce type font appel à un module de reconnaissance de la parole permettant de transformer le signal vocal en version textuelle. Cette transcription est ensuite analysée de manière à extraire les informations sémantiques indispensables au système pour répondre aux attentes de l'utilisateur. Dans cette architecture séquentielle de traitement des tâches, la qualité de l'interprétation sémantique est très dépendante de la qualité du processus de reconnaissance automatique de la parole (RAP). Ce module de RAP utilise généralement des informations acoustiques pour convertir le signal en unités linguistiques de base (phonèmes, syllabes ou mots) et des informations linguistiques à portée réduite (modèle de langage  $N$ -grammes). Le module de compréhension s'appuie sur des unités sémantiques élémentaires, que l'on appelle concepts souvent représentés sous la forme d'une paire « attribut/valeur ». Ils sont ensuite composés dans le but de construire une représentation sémantique. L'objectif de ces travaux de thèse a été l'amélioration du processus d'extraction des concepts élémentaires en faisant intervenir des informations de plus haut niveau dans la recherche de la transcription. Les différents points suivants résument les travaux effectués durant la thèse :

**MODÈLE DE LANGAGE CONCEPTUEL :** J'ai présenté l'élaboration d'un modèle de langage basé sur le formalisme des transducteurs à états finis. Ce transducteur encode des grammaires régulières permettant de faire l'association entre les mots et les concepts utilisés par le module de compréhension. Il permet d'analyser toute phrase construite à partir du lexique de l'application visée et d'émettre les séquences conceptuelles qu'il est possible d'y trouver. Ce transducteur est ensuite optimisé afin de retirer les ambiguïtés qu'il est possible de lever à ce niveau du processus. Ces ambiguïtés peuvent être, par exemple, qu'à partir d'une même chaîne de mot le transducteur émette plusieurs séquences conceptuelles identiques, ou alors que plusieurs concepts peuvent être émis alors que le contexte immédiat dans la phrase permet de connaître le bon. Ce modèle de langage permet d'en-

richir l'espace de recherche généré par le module de reconnaissance automatique de la parole (graphe de mots représenté sous la forme d'un automate à états fini) d'informations conceptuelles par une simple opération de composition [3,23].

**DÉCODAGE CONCEPTUEL :** Dans cet espace enrichi, représenté par un transducteur, les entrées correspondent aux mots tandis que les sorties correspondent aux concepts associés. Il est alors aisé de passer du graphe de mots au graphe de concepts en ne considérant que les sorties du transducteur. Dans ce nouvel espace de recherche, il est maintenant possible de faire intervenir d'autres connaissances, sur l'ordonnement des concepts où les prédictions du gestionnaire de dialogue dans la recherche de la transcription. Sans toutefois utiliser d'autres scores, je propose un processus de décodage qui cherche d'abord pour les meilleures interprétations (séquences de concepts) possibles et fournit une liste structurée des  $N$ -meilleures hypothèses conceptuelles accompagnées de leur meilleure chaîne de mots la supportant. Cette liste permet d'obtenir en quelques hypothèses un résumé exhaustif et non-redondant du point de vue de la compréhension du graphe de mots [3].

**MESURES DE CONFIANCE :** Il est primordial dans ce genre de système de pouvoir estimer la qualité de la reconnaissance afin de palier aux inévitables erreurs de reconnaissance. J'ai proposé différentes mesures de confiance faisant intervenir des critères linguistiques, acoustiques ou sémantiques pour évaluer la qualité de la transcription à différents niveaux : mot, phrase et concept. Les mesures linguistiques chargées de diagnostiquer la qualité de la transcription au niveau de la phrase sont basées sur le fait que l'utilisation de méthodes de repli dans la construction d'un modèle de langage  $N$ -grammes est souvent générateur d'erreurs. Nous avons proposé de raisonner sur des situations de confiance obtenues par des décodages en parallèle effectués avec des modèles de langage augmentés permettant de faire appel aux méthodes de repli le moins souvent possible. Une mesure calculant la fréquence d'utilisation de ces méthodes de repli est également proposée et montre des résultats intéressants [4,32]. Ont été également développées des mesures de confiance au niveau conceptuel qui utilisent des informations acoustiques ou plus originalement, des classifieurs textuels appris automatiquement.

**STRATÉGIE D'AIDE À LA DÉCISION :** La dernière partie de ces travaux s'est attachée à fournir au gestionnaire de dialogue une information pertinente sur la qualité du processus de décodage. Je propose une stratégie basée sur un arbre de décision où chaque nœud est une unité de décision qui fait de la validation en fonction de la liste structurée et des mesures de confiance. Ces unités exploitent des consensus de réponses données par différents classifieurs, LIA-SCT (arbre de décision sémantique), BoostTexter (classifieur à large marge, implémentation de l'algorithme de boosting) et SVM-Torch (machine à support vectoriel). J'utilise dans mes expériences une stratégie avec deux unités de décisions. J'aboutis alors à 4 situations de fiabilité différentes avec dans certaines une probabilité très forte que l'hypothèse conceptuelle décodée soit correcte et d'autres non. De plus la stratégie employée permet de prédire le type d'erreur commis sur les concepts (substitution, insertion, ou suppression). Dans les situations où la probabilité que les hypothèses soient correctes est faible je présente des méthodes de correction d'erreurs [2], voir aussi [20,21,19]. La correction d'erreurs revient à chercher une alternative crédible à la meilleure hypothèse au sein de la liste structurée des  $N$ -meilleures hypothèses. Cette stratégie permet de donner les moyens au gestionnaire de dialogue de prendre une décision plus optimale sur ses choix à effectuer dans la gestion du dialogue [2].

L'avancement des travaux a fait l'objet de rapports remis à France Télécom de périodicité bi-annuelle.

## **TRAVAUX EN TANT QU'ATER AU LABORATOIRE INFORMATIQUE D'AVIGNON, L.I.A.**

Durant l'année d'ATER au Laboratoire Informatique d'Avignon, j'ai poursuivi mes travaux sur les applications de dialogue homme-machine, plus particulièrement au niveau de la Reconnaissance Automatique de la Parole (RAP) et la compréhension. Les travaux réalisés durant ma thèse sont exploités sur les données MEDIA (Méthodologie d'évaluation automatique de la compréhension hors et en contexte du DIALogue). L'opération de décodage conceptuel, au moyen d'un modèle de langage conceptuel stochastique est effectuée directement sur le graphe de mots (approche intégrée) produit par un module de RAP plutôt que sur la ou les  $N$ -meilleures hypothèses de reconnaissance (approche séquentielle). Cette opération permet de gagner significativement au niveau du taux d'erreur sur l'extraction des concepts [18,30].

## **TRAVAUX DE POST-DOC AU DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE, D.I.S.I., À L'UNIVERSITÉ DE TRENTE, ITALIE**

J'ai réalisé une année de Post-Doc à l'université de Trento en Italie dans l'équipe de Giuseppe Riccardi pour travailler sur le projet européen LUNA (spoken Language UNderstanding in multilinguAl communication systems) <http://www.ist-luna.eu/>. C'est un projet ambitieux et très motivant dont l'objectif est d'améliorer la compréhension de la parole spontanée dans les applications téléphoniques. L'originalité et l'ambition de ce projet est de s'attaquer à la compréhension dans des systèmes complexes (e.g. centre d'appels pour l'assistance technique) et dans différentes langues (i.e. français, italien, polonais). Dans ce projet, j'ai travaillé sur les sujets suivants : dans la continuité de mes précédents travaux j'ai étudié les moyens d'exploiter les sources de connaissances diverses (connaissances *a priori*, historique du dialogue ou prédictions du système) dans le processus de compréhension de la parole. Cela m'a conduit à étudier les modèles discriminants qui permettent d'intégrer de manière souple ce genre d'information. Des travaux préliminaires comparant différents modèles (i.e. les automates à états fini (FSM), les Machines à Vecteur de Support (SVM) et les Champs de Markov Aléatoires (CRF) ont été publiés dans [15]. J'ai également collaboré à un travail sur l'intégration d'informations liées à la structure syntaxique/sémantique au sein d'un classifieur [14]. J'ai participé étroitement à la définition du protocole d'annotation sémantique des corpus de dialogue qui seront produits au cours de ce projet,

travaux publiés dans [17,16]. L'annotation de corpus, m'a amené à travailler sur le principe d'« Active Learning », dont le but est de tirer au mieux parti des données d'apprentissage pour les modèles statistiques. Un modèle est appris sur un nombre restreint d'exemples annotés, ce modèle sélectionne ensuite automatiquement les exemples les plus pertinents (les plus ambigus de son point de vue) qui devront être annotés. L'Active Learning permet d'obtenir des modèles aussi (ou plus) robustes que ceux qui auraient été appris avec l'intégralité des données, tout en économisant le temps et l'argent nécessaire pour faire ces annotations. J'ai travaillé sur des stratégies permettant de prendre en compte les erreurs d'annotation, qui ont un impact très négatif sur les modèles d'apprentissage statistiques, dans le processus d'Active Learning, des premiers résultats sont publiés dans [13].

### **TRAVAUX EN TANT QU'ATER AU LABORATOIRE INFORMATIQUE D'AVIGNON, L.I.A.**

J'ai ré-intégré le Laboratoire Informatique d'Avignon en tant qu'ATER en continuant ma participation sur le projet européen LUNA. En collaboration avec l'équipe dirigée par Herman Ney à l'université d'Aachen (Aix-la-Chapelle) j'ai poursuivi les évaluations des performances des différents modèles statistiques sur des tâches de compréhension de la parole, les résultats ont été publiés à [11]. J'ai développé un système automatique de pré-annotation et de vérification visant à accélérer et améliorer la qualité du processus d'annotation sémantique des corpus italiens que nous produisons au sein du projet LUNA [12,29].

### **TRAVAUX EN TANT QUE CHERCHEUR CONTRACTUEL AU LABORATOIRE INFORMATIQUE D'AVIGNON, L.I.A.**

J'ai pris cette année la direction du WorkPackage 4 « *on-line, context sensitive validation* » du projet européen LUNA, et je travaille actuellement sur la représentation sémantique en frame pour le dialogue Homme-Machine. L'objectif est de proposer une représentation sémantique structurée allant au delà de la représentation conceptuelle « à plat ».

## **Activités dans la communauté**

Relecteur pour la conférence ICMLA (2011,2012), IEEE International Conference on Machine Learning and Applications  
Relecteur pour la conférence Interspeech (2011,2012), Annual Conference of the International Speech Communication Association  
Relecteur pour la conférence IEEE ICASSP (2008 et 2009), International Conference on Acoustics, Speech, and Signal Processing  
Relecteur pour la conférence LREC 2008,2010 Language Resources and Evaluation Conference  
Relecteur pour la conférence EACL 2009, The 12th Conference of the European Chapter of the Association for Computational Linguistics

## Publications

Les publications suivantes sont disponibles en version électronique à l'adresse suivante :

<http://www.irisa.fr/texmex/people/raymond/publi.html>

### Reuves internationales (4)

- [1] STEFAN HAHN, MARCO DINARELLI, CHRISTIAN RAYMOND, FABRICE LEFÈVRE, PATRICK LEHNEN, RENATO DE MORI, ALESSANDRO MOSCHITTI, HERMANN NEY ET GIUSEPPE RICCARDI (2010)  
« Comparing Stochastic Approaches to Spoken Language Understanding in Multiple Languages ».  
*IEEE Transactions on Audio, Speech and Language Processing*, PP :1.
- [2] CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET, NATHALIE CAMELIN, RENATO DE MORI, GÉRALDINE DAMNATI (2007).  
« Sequential decision strategies for machine interpretation of speech ».  
*IEEE Transactions on Audio, Speech and Language Processing*, 15(1) :162-171.
- [3] CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET, RENATO DE MORI ET GÉRALDINE DAMNATI (2006).  
« On the use of finite state transducers for semantic interpretation ».  
*Speech Communication*, 48 :288-304.
- [4] YANNICK ESTÈVE, CHRISTIAN RAYMOND, RENATO DE MORI ET DAVID JANISZEK (2003).  
« On the use of linguistic consistency in systems for human-computer dialogs ».  
*IEEE Transactions on Speech and Audio Processing*, 11(6) :746-756.

### Conférences et workshops internationaux (22)

- [5] JULIEN FAYOLLE, MURAT SARACLAR, FABIENNE MOREAU, CHRISTIAN RAYMOND ET GUILLAUME GRAVIER (2012).  
« Lexical-phonetic automata for spoken utterance indexing and retrieval »  
*Spoken Language Technologies Workshop*, Berkeley, California, U.S.A.
- [6] FRÉDÉRIC BÉCHET, FRÉDÉRIC DUVERT, CHRISTIAN RAYMOND ET RENATO DE MORI (2010).  
« Frame Based Interpretation Of Conversational Speech »  
*Spoken Language Technologies Workshop*, Berkeley, California, U.S.A.
- [7] JULIEN FAYOLLE, FABIENNE MOREAU, CHRISTIAN RAYMOND ET GUILLAUME GRAVIER (2010).  
« Reshaping Automatic Speech Transcripts for Robust High-level Spoken Document Analysis »  
*Analytics for Noisy Unstructured Text Data*, Toronto, Canada.
- [8] JULIEN FAYOLLE, FABIENNE MOREAU, CHRISTIAN RAYMOND, GUILLAUME GRAVIER ET PATRICK GROS (2010).  
« CRF-based Combination of Contextual Features to Improve A Posteriori Word-level Confidence Measures »  
*InterSpeech*, Makuari, Japan.
- [9] CHRISTIAN RAYMOND ET JULIEN FAYOLLE (2010).  
« Reconnaissance robuste d'entités nommées sur de la parole transcrite automatiquement »  
*Traitement Automatique des Langues Naturelles*, TALN'10, Montréal, Canada.
- [10] CHRISTOPHE SERVAN, NATHALIE CAMELIN, CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET ET RENATO DE MORI (2010).  
« On The Use Of Machine Translation For Spoken Language Understanding Portability »  
*In Proceedings of the International Conference on Acoustic Speech and Signal Processing ICASSP'10*, Dallas, USA.
- [11] STEFAN HAHN, PATRICK LEHNEN, CHRISTIAN RAYMOND ET HERMANN NEY (2008).  
« A Comparison of Various Methods for Concept Tagging for Spoken Language Understanding »  
*Language Resources and Evaluation Conference*, LREC08, Marrakech, Maroc.
- [12] CHRISTIAN RAYMOND, KEPA JOSEBA RODRIGUEZ ET GIUSEPPE RICCARDI (2008).  
« Active Annotation in the LUNA Italian Corpus of Spontaneous Dialogues »  
*Language Resources and Evaluation Conference*, LREC08, Marrakech, Maroc.
- [13] CHRISTIAN RAYMOND ET GIUSEPPE RICCARDI (2008).  
« Learning with Noisy Supervision for Spoken Language Understanding ».

*In Proceedings of the International Conference on Acoustic Speech and Signal Processing, Las Vegas, USA.*

[14] ALESSANDRO MOSCHITTI, GIUSEPPE RICCARDI ET CHRISTIAN RAYMOND (2007).

« Spoken Language Understanding With Kernels For Syntactic/Semantic Structures ».

*Proceedings IEEE Workshop on Automatic Speech Recognition and Understanding, Kyoto, Japon.*

[15] CHRISTIAN RAYMOND ET GIUSEPPE RICCARDI (2007).

« Generative and Discriminative Algorithms for Spoken Language Understanding ».

*Interspeech, Anvers, Belgique.*

[16] CHRISTIAN RAYMOND, GIUSEPPE RICCARDI, KEPA JOSEBA RODRIGUEZ ET JOANNA WIŚNIEWSKA (2007).

« The LUNA Corpus : an Annotation Scheme for a Multi-domain Multi-lingual Dialogue Corpus ».

*Workshop on the Semantics and Pragmatics of Dialogue, DECALOG'2007, Rovereto, Italie.*

[17] KEPA JOSEBA RODRIGUEZ, STEFANIE DIPPER, MICHAEL GÖTZE, MASSIMO POESIO, GIUSEPPE RICCARDI, CHRISTIAN RAYMOND ET JOANNA RABIEGA-WIŚNIEWSKA (2007).

« Standoff Coordination for Multi-Tool Annotation in a Dialogue Corpus ».

*Linguistic Annotation Workshop, LAW07, Prague, République tchèque.*

[18] CHRISTOPHE SERVAN, CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET ET PASCAL NOCÉRA (2006).

« Conceptual decoding from word lattices : application to the spoken dialogue corpus MEDIA »

*In Proceedings of the International Conference on Spoken Language Processing, ICSLP'06. Pittsburgh, USA*

[19] CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET, RENATO DE MORI ET GÉRALDINE DAMNATI (2005).

« Semantic interpretation with error correction ».

*In Proceedings of the International Conference on Acoustic Speech and Signal Processing, ICASSP'05, Philadelphie, USA.*

[20] CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET, RENATO DE MORI ET GÉRALDINE DAMNATI (2004).

« On the use of confidence for statistical decision in dialogue strategies ».

*In 5<sup>th</sup> SIGdial Workshop on Discourse and Dialogue, Boston, USA.*

[21] CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET, RENATO DE MORI, GÉRALDINE DAMNATI ET YANNICK ESTÈVE (2004).

« Automatic learning of interpretation strategies for spoken dialogue systems ».

*In Proceedings of the International Conference on Acoustic Speech and Signal Processing, ICASSP'04, Montréal, Canada.*

[22] CHRISTIAN RAYMOND, YANNICK ESTÈVE, FRÉDÉRIC BÉCHET, RENATO DE MORI ET GÉRALDINE DAMNATI (2003).

« Belief confirmation in spoken dialogue systems using confidence measures ».

*In Proceedings IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU'03, St. Thomas, US-Virgin Islands.*

[23] YANNICK ESTÈVE, CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET ET RENATO DE MORI (2003).

« Conceptual decoding for spoken dialog systems ».

*In Proceedings of European Conference on Speech Communication and Technology, Eurospeech'03, Genève, Suisse.*

[24] RENATO DE MORI, YANNICK ESTÈVE ET CHRISTIAN RAYMOND (2002).

« On the use of structures in language models for dialogue ».

*In Proceedings of the International Conference on Spoken Language Processing, ICSLP'02, Denver, Colorado, USA.*

[25] YANNICK ESTÈVE, CHRISTIAN RAYMOND ET RENATO DE MORI (2002).

« On the use of structure in language models for dialogue : specific solutions for specific problems ».

*In ISCA Tutorial and Research Workshop on Multi-Modal Dialogue in Mobile Environments, Kloster Irsee, Allemagne.*

## **Conférences et ateliers nationaux (8)**

[26] JULIEN FAYOLLE, FABIENNE MOREAU, CHRISTIAN RAYMOND ET GUILLAUME GRAVIER (2012).

« Automates lexico-phonétiques pour l'indexation et la recherche de segments de parole »

*Journée d'Études sur la Parole, JEP'12, Juin, Grenoble, France*

[27] CHRISTIAN RAYMOND ET VINCENT CLAVEAU (2012).

« Participation de l'IRISA à DEFT 2012 : recherche d'information et apprentissage pour la génération de mots-clés »

*Défi Fouille de Texte, Grenoble, Juin, France*

[28] CHRISTIAN RAYMOND ET VINCENT CLAVEAU (2011).

« Participation de l'IRISA à DEFT 2011 : expériences avec des approches d'apprentissage supervisé et non-supervisé »

*Défi Fouille de Texte*, Montpellier, France.

[29] CHRISTIAN RAYMOND, KEPA JOSEBA RODRIGUEZ (2008).

« Annotation dynamique dans le corpus italien de dialogues spontanés LUNA ».

*XXVII<sup>ième</sup> Journée d'Études sur la Parole, JEP'08*, Avignon, France.

[30] CHRISTOPHE SERVAN, CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET AND PASCAL NOCÉRA (2006).

« Décodage conceptuel à partir de graphes de mots sur le corpus de dialogue Homme-Machine MEDIA ».

*XXVI<sup>ième</sup> Journée d'Études sur la Parole, JEP'06*, Rennes, France.

[31] CHRISTIAN RAYMOND, FRÉDÉRIC BÉCHET, RENATO DE MORI ET GÉRALDINE DAMNATI (2004).

« Stratégie de décodage conceptuel pour les applications de dialogue oral ».

*XXV<sup>ième</sup> Journée d'Études sur la Parole, JEP'04*, Fès, Maroc.

[32] CHRISTIAN RAYMOND (2003).

« Mesures de confiance pour la reconnaissance de la parole dans des applications de dialogue homme-machine ».

*Majestic*, Marseille, France.

[33] CHRISTIAN RAYMOND, PATRICE BELLOT ET MARC EL-BÈZE (2002).

« Enrichissement de requêtes pour la recherche documentaire selon une classification non-supervisée ».

*13<sup>ième</sup> Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et d'Intelligence Artificielle (RFIA'2002)*, pages 625 à 632, Angers, France.

## **Mémoire et Thèse**

[34] CHRISTIAN RAYMOND (2005).

« Décodage conceptuel : co-articulation des processus de transcription et compréhension dans les systèmes de dialogue ».

*Thèse de doctorat*, Université d'Avignon et des Pays de Vaucluse.

[35] CHRISTIAN RAYMOND (2001).

« Réécriture de requêtes pour la recherche documentaire selon une méthode de classification à base d'arbres de décision non-supervisés ».

*Mémoire de DEA*, Marseille (Luminy).

## Enseignement

Les enseignements que j'ai dispensés en tant que vacataire et ATER à l'université d'Avignon entre 2001 et 2006 ainsi que 2007/2008 sont présentés ci-dessous sous la forme d'un tableau. Celui-ci contient une brève description du contenu de ces enseignements ainsi que leur volume horaire (totalisés sur toutes les années). Un second tableaux résume mon parcours enseignement. De part ma formation initiale, mes compétences en enseignements informatiques ne se limitent pas aux enseignements que j'ai effectués.

<b>Réseaux et Télécommunications</b> (Licence) 189 heures
Objectifs du module : comprendre et de maîtriser les couches basses des réseaux informatiques : la transmission des données, la communication entre machines et l'interconnexion dans les réseaux. Sont abordés le codage de l'information, l'accès au média, le routage et le contrôle des flux et des erreurs dans le modèle OSI et le protocole TCP/IP avec l'adressage et les masques de sous-réseaux.
<b>Base de données</b> (Licence) 46 heures
Objectifs du module : concevoir et modéliser une base de données relationnelle. Savoir décomposer une relation jusqu'à la quatrième forme normale pour éviter les redondances. Mise en pratique en utilisant l'outil Oracle.
<b>Langages objets Java et C# &amp; API Avancées</b> (Master 1) 67 heures
Objectifs du module : Concepts et APIs de référence en Java et C# : interfaces (swing, AWT, winform), entrées/sorties, gestion des exceptions, accès concurrentiels, accès distants, applets ...
<b>Préparation et Certification C2I</b> (Licence) 108 heures
Objectifs du module : Le Certificat informatique et internet (C2i) niveau 1 est un certificat national (BO du 7 avril 2005) destiné à développer, renforcer et valider la maîtrise des technologies de l'information et de la communication (TIC). Il développe les compétences suivantes : S'approprier son environnement de travail et ses données (Entreprendre) ; Communiquer à distance (Échanger) ; Rechercher l'information (Enquêter) ; Réaliser, élaborer des documents complexes et structurés (Élaborer) ; Communiquer les résultats de ses travaux à l'aide d'un outil adapté (Éditer).
<b>Théorie des langages</b> (Licence) 69 heures
Objectifs du module : Acquérir une aisance dans la mise en relation des principes de compétence et performance, représentations conceptuelles et réalisations concrètes au travers du lien qui existe entre la théorie des langages et plus tard la compilation. Contenu : Après quelques rappels sur les ensembles, des notions de base (monoïde, algorithme, etc.) sont passées en revue avant d'aborder les 2 grands chapitres de la théorie des langages qui couvrent l'analyse lexicale et l'analyse syntaxique. Sont abordées les Grammaires, les règles de production, la typologie des grammaires, les opérations de dérivation et Mise en évidence de problèmes découlant du non déterminisme, de l'ambiguïté, ou de la récursivité. On voit comment transformer une expression régulière en une grammaire régulière ou en un automate régulier. On apprend à rendre les automates réguliers déterministes et à les minimiser. Parmi les stratégies d'analyses syntaxiques déterministes qui sont évoquées, l'accent est mis sur 3 d'entre elles : algorithme d'Earley, méthodes LL(k), et LR(k).
<b>Bureautique</b> (Licence) 18 heures
Objectifs du module : former les étudiants à la maîtrise des outils de bureautique, tels que les traitements de textes, les tableurs, les navigateurs WEB, etc.
<b>Analyse et compréhension</b> (Master 2) 6 heures
Objectifs du module : L'objectif principal de ce cours est de permettre, comprendre et appliquer les principes de méthodes robustes au traitement et la compréhension de la langue naturelle (ACLN) et leurs applications pour manipuler des grandes quantités de texte. Le cours offrira une introduction aux méthodes statistiques utilisées dans TALNE, expliquera les modèles théoriques de base et montrera leurs applications pratiques aux grandes collections de textes.
<b>Administration et gestion des réseaux</b> (Master 1) 12 heures
Objectifs du module : Approche de l'administration réseau par la pratique : protocole SNMP, le serveur d'adressage DHCP, le serveur de nom Internet DNS, le serveur de fichier NFS, le service de messagerie, NFS et Samba.



STATUT	Année	Nature	Public	Enseignement	Heure équivalent TD
ATER	2007-2008	TP	Master 1	Langages objets Java et C#	50
		TD	Master 1	APIs avancées	27
		TD	Licence GMI	Réseau et Télécom	30
		TD	Licence	Préparation c2i	72
		Suivi en ligne	Licence	Certification c2i	36
ATER	2005-2006	TP	Master 2	Analyse et Compréhension	6
		TD	Licence GMI	Théorie des langages	20
		TD	Licence GMI	Réseau et Télécom	45
Vacataire	2004-2005	TD	Licence GMI	Théorie des langages	9
		TD	Licence GMI	Réseau et Télécom	30
	2003-2004	TP	Licence GMI	Théorie des langages	20
		TD	Licence GMI	Réseau et Télécom	32
		Cours/TP	Licence LEA	Bureautique	18
	2002-2003	TD	Licence GMI	Théorie des langages	20
		TD	Licence GMI	Réseau et Télécom	32
		TD	Licence GMI	Base de données	10
		TP	Licence GMI	Base de données	10
	2001-2002	TD	Licence GMI	Réseau et Télécom	32
		TD	DU BAC+3	Réseau et Télécom	18
		TD	Licence GMI	Base de données	14
		TP	Licence GMI	Base de données	12
Volume horaire total					543