

Benoit Favre

LIUM, Université du Maine, Avenue Laënnec, 72085 Le Mans Cedex 9, France +33 2 43 83 38 52; benoit.favre@gmail.com
28 ans, permis de conduire Français mise à jour : 2010-01-21

Formation

- **2007, Doctorat en informatique**
Laboratoire d'Informatique d'Avignon (LIA), Université d'Avignon (France)
- **2003, DEA en informatique (équivalent Master Recherche)**
Informatique fondamentale, Université d'Avignon
- **2002, Maîtrise en informatique**
Génie Mathématiques et Informatique, spécialité réseaux, titre d'Ingénieur Maître, Université d'Avignon

Recherche

- **Recherche post-doctorale**
Couplage d'interprétation sémantique et reconnaissance automatique de la parole pour l'analyse de dialogues parlés (en Français et Italien). Projet PORT MEDIA.
Reponctuation de discours parlé dans différents styles (news, conversations, réunions) et multiples langues (Anglais, Arabe, Mandarin). Impact de la ponctuation sur les tâches aval. Modèles à dépendances étendues. Projet GALE.
Résumé automatique de réunions. Détection d'expressions salientes et raffinement par interactions utilisateur. Optimisation globale pour le résumé automatique. Projets AMIDA et CALO.
- **Travaux de recherche doctorale**
Nous avons exploré les problématiques du *Résumé automatique de parole* appliqué à la *Recherche Documentaire Audio* pour faciliter la navigation dans les bases de données audio. Les travaux sont spécifiquement menés sur :
 - la structuration automatique de documents audio
 - le résumé automatique par extraction utilisant des méthodes statistiques
 - les interfaces utilisateurs à des bases de données audioTravail co-financé par le Ministère de la Recherche et de l'Industrie et Thales Land & Joint Systems.
- **Campagnes d'Évaluation Scientifique**
 - extraction d'entités nommées dans des émissions de radio (ESTER), en 2005
 - résumé automatique de texte (DUC, TAC), de 2006 à 2009
 - reponctuation de parole pour la traduction automatique (GALE) de 2007 à 2009
 - génération d'expressions textuelles (GREC) en 2009

Expérience professionnelle

- **2009 au présent - Ingénieur de Recherche**
Laboratoire d'Informatique de l'Université du Maine, Le Mans (France).
- **2007 à 2009 - Post-doc**
International Computer Science Institute, Berkeley (USA).
- **2004 à 2007 - Ingénieur de recherche**
Thales Land & Joint Systems, Colombes (France), laboratoire Multimedia Processing.
- **2004 à 2006 - Web master de l'Association Francophone de la Communication Parlée (AFCP)**
Conversion du site à Spip (portail de gestion de contenu en php)
Implication dans le projet national Technolangue

- **2001 à 2003 - Projet de création d'entreprise - Allegorian**
Indexation automatique de la parole pour l'accès aux documents multimédia par le contenu.
- positionnement et étude de marché
- stratégies commerciale et de communication
- développement d'une plate-forme d'indexation distribuée et d'un moteur de recherche sur la parole
- **2002 - Projet de fin d'études - LIA**
Adaptation d'un moteur de recherche documentaire en langue naturelle à l'indexation automatisée de la parole
- **2001 - Stage de 2ème année, Université d'Avignon**
Portage sous linux d'un logiciel de gestion commerciale pour PARTS EQUIPMENT, un revendeur Français.

Compétences Techniques

- **Programmation** : Perl, Java, C++, C, Python
- **Apprentissage Artificiel** : AdaBoost, CRF, SVM
- **Ingénierie logicielle** : UML, Corba
- **Système** : Linux/Unix, OSX
- **Réseau** : Matériels actifs/passifs, Administration, TCP/IP, Nat, Bind, Dhcp
- **Bases de données** : Postgres, Mysql, Access, Oracle
- **Internet** : Apache, HTML, PHP, Javascript, Ajax, Flash
- **Langues** : Anglais scientifique parlé et écrit, Français natif

Enseignements (156 heures)

- **2005 et 2006 (16 heures) - TP Voix-Données-Images, niveau Master Informatique, IUP GMI**
Compression à ultra bas débit par décodage acoustico-phonétique et synthèse vocale
- **2003 et 2004 (80 heures) - TP Réseaux, niveau Licence Informatique, IUP GMI**
câblage RJ45, étude des flux réseau avec *Etheral*, mise en oeuvre de routeurs CISCO
- **2000 (60 hours) - Tutorats de bureautique à l'université d'Avignon**
élèves de 2ème Année de Sciences de la Vie (SV) et Langues Étrangères Appliquées (LEA)

Publications

- **Revues (5)**

Gokhan Tur, Andreas Stolcke, Lynn Voss, Stanley Peters, Dilek Hakkani-Tür, u John Dowding, Benoit Favre, Raquel Fernández, Matthew Frampton, Mike Frandsen, Clint Frederickson, Martin Graciarina, Donald Kintzing, Kyle Leveque, Shane Mason, John Niekrasz, Matthew Purver, Korbinian Riedhammer, Elizabeth Shriberg, Jing Tien, Dimitra Vergyri, Fan Yang, " *The CALO Meeting Assistant System*", à paraître dans IEEE Transactions on Audio, Speech and Language Processing – 2010

Elizabeth Shriberg, Benoit Favre, James Fung, Dilek Hakkani-Tür, Sébastien Cuendet, " *Prosodic Similarities of Dialog Act Boundaries Across Speaking Styles*", Linguistic Patterns in Spontaneous Speech, Language and Linguistics Monograph Series Number A25, Edited by Shu-Chuan Tseng, Institute of Linguistics, Academia Sinica, Taipei (Taiwan) – 2009

Umit Guz, Benoit Favre, Dilek Hakkani-Tür and Gokhan Tur, " *Generative and Discriminative Methods using Morphological Information for Sentence Segmentation of Turkish*", IEEE Transactions on Audio, Speech and Language Processing, Special Section on Processing Morphologically Rich Languages, Volume 17, Number 5, pp. 295-903 – 2009

Mari Ostendorf, Benoit Favre, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Dustin Hillard, Julia B. Hirschberg, Heng Ji, Jeremy G Kahn, Yang Liu, Evgeny Matusov, Hermann Ney, Elizabeth Shriberg, Wen Wang, Chuck Wooters, "Speech Segmentation and Its Impact on Spoken Language Technology", IEEE Signal Processing Magazine – 2008

Sebastien Cuendet, Dilek Hakkani-Tür, Elizabeth Shriberg, James Fung, Benoit Favre, "Cross-Genre Feature Comparisons for Spoken Sentence Segmentation", International Journal of Semantic Computing (IJSC), Volume 1, Issue 3; also published in IEEE International Conference on Semantic Computing (ICSC 2007), pp. 265-271, Irvine (USA) – 2007

- **Conférences internationales avec comité de lecture (20)**

Benoit Favre, Bernd Bohnet, Dilek Hakkani-Tür, "Evaluation of Semantic Role Labeling and Dependency Parsing of Automatic Speech Recognition Output", IEEE International Conference in Acoustics, Speech and Signal Processing (ICASSP), Dallas (USA) – 2010

Kofi Boakye, Benoit Favre, Dilek Hakkani-Tür, "Any Questions? Automatic Question Detection in Meetings", ASRU, Merano (Italy) – 2009

Shasha Xie, Dilek Hakkani-Tür, Benoit Favre, Yang Liu, "Integrating Prosodic Features in Extractive Meeting Summarization", ASRU, Merano (Italy) – 2009

Daniel Gillick, Benoit Favre, "A Scalable Global Model for Summarization", NAACL/HLT 2009 Workshop on Integer Linear Programming for Natural Language Processing – 2009

Shasha Xie, Benoit Favre, Dilek Hakkani-Tür, Yang Liu, "Leveraging Sentence Weights in Concept-based Optimization Framework for Extractive Meeting Summarization", Interspeech, Brighton (UK) – 2009

Nikhil Garg, Benoit Favre, Korbinian Reidhammer, Dilek Hakkani-Tür, "ClusterRank: A Graph Based Method for Meeting Summarization", Interspeech, Brighton (UK) – 2009

Benoit Favre, Dilek Hakkani-Tür, "Phrase-level and Word-level Strategies for Detecting Appositions in Speech", Interspeech, Brighton (UK) – 2009

Benjamin Lecouteux, Georges Linarès, Benoit Favre, "Combined low level and high level features for Out-Of-Vocabulary Word detection", Interspeech, Brighton (UK) – 2009

Benoit Favre, Dilek Hakkani-Tür, Elizabeth Shriberg, "Syntactically-informed Models for Comma Prediction", IEEE International Conference in Acoustics, Speech and Signal Processing (ICASSP), Taipei (Taiwan) – 2009

Daniel Gillick, Korbinian Riedhammer, Benoit Favre, Dilek Hakkani-Tür, "A Global Optimization Framework for Meeting Summarization", IEEE International Conference in Acoustics, Speech and Signal Processing (ICASSP), Taipei (Taiwan) – 2009

Benoit Favre, Dilek Hakkani-Tür, Slav Petrov, Dan Klein, "Efficient Sentence Segmentation Using Syntactic Features", Spoken Language Technologies (SLT), Goa (India) – 2008

Gokhan Tur, Andreas Stolcke, Lynn Voss, John Dowding, Benoit Favre, Raquel Fernandez, Matthew Frampton, Michael Frandsen, Clint Frederickson, Martin Graciarena, Dilek Hakkani-Tür, Donald Kintzing, Kyle Leveque, Shane Mason, John Niekrasz, Stanley Peters, Matthew Purver, Korbinian Riedhammer, Elizabeth Shriberg, Jing Tien, Dimitra Vergyri, Fan Yang, "The CALO Meeting Speech Recognition and Understanding System", Spoken Language Technologies (SLT), Goa (India) – 2008

Korbinian Riedhammer, Benoit Favre, Dilek Hakkani-Tür, "A Keyphrase Based Approach to Interactive Meeting Summarization", Spoken Language Technologies (SLT), Goa (India) – 2008

Korbinian Riedhammer, Daniel Gillick, Benoit Favre, Dilek Hakkani-Tür, "Packing the Meeting Summarization Knapsack", Interspeech, Brisbane (Australia) – 2008

Benoit Favre, Ralph Grishman, Dustin Hillard, Heng Ji, Dilek Hakkani-Tür, Mari Ostendorf, "Punctuating Speech for Information Extraction", IEEE International Conference in Acoustics, Speech and Signal Processing (ICASSP), Las Vegas, Nevada (USA) – 2008

Benoit Favre, Jean-François Bonastre, Patrice Bellot, "An Interactive Timeline for Speech Database Browsing", Interspeech 2007, Antwerp (Belgium) – 2007

Sebastien Cuendet, Elizabeth Shriberg, Benoit Favre, James Fung, Dilek Hakkani-Tür, "An Analysis of Sentence Segmentation Features for Broadcast News, Broadcast Conversations, and Meetings", SIGIR 2007, Searching Spontaneous Conversational Speech (SSCS) workshop, pp. 37-43, Amsterdam (The Netherlands) – 2007

Benoit Favre, Frédéric Béchet, Pascal Nocéra, "Robust Named Entity Extraction from Spoken Archives", Proceedings of HLT-EMNLP'05, Vancouver (Canada) – 2005

Benoit Favre, Frédéric Béchet, Pascal Nocéra, "Mining Broadcast News data: Robust Information Extraction from Word Lattices", Proceeding of Eurospeech'05, Lisbonne (Portugal) – 2005

Benoit Favre, Patrice Bellot, Jean-François Bonastre, "Information retrieval on mixed written and spoken documents", actes de la 7ème conférence RIAO, Avignon (France), pp. 826-835 – 2004

- **Conférences nationales avec comité de lecture (2)**

Benoit Favre, Jean-François Bonastre, Patrice Bellot, François Capman, "Accès aux connaissances orales par le résumé automatique", EGC'06, Lille (France) – 2006

Benoit Favre, Jean-François Bonastre, Patrice Bellot, "Recherche d'information dans un mélange de documents écrits et parlés", Journées d'Etude de la Parole (JEP), Fèz (Maroc) – 2004

- **Autres (7)**

Daniel Gillick, Benoit Favre, Dilek Hakkani-Tür, Bernd Bohnet, Yang Liu, Shasha Xie, "The ICSI/UTD Summarization System at TAC 2009", à paraître dans Text Analysis Conference, Gaithersburg, MD (USA) – 2009

Benoit Favre, Bernd Bohnet, "ICSI-CRF: The Generation of References to the Main Subject and Named Entities Using Conditional Random Fields", ACL-IJCNLP workshop on Language Generation and Summarisation – 2009

Daniel Gillick, Benoit Favre, Dilek Hakkani-Tür, "The ICSI Summarization System at TAC 2008", Text Analysis Conference, Gaithersburg, MD (USA) – 2008

Florian Boudin, Benoit Favre, Frédéric Béchet, Marc El-Bèze, Laurent Gillard, Juan-Manuel Torres-Moreno, "The LIA-Thales summarization system at DUC-2007", DUC-NIST Proceedings Document Understanding Conference (DUC-2007), New York (USA) – 2007

Benoit Favre, "Résumé automatique de parole pour un accès efficace aux bases de données audio", Thèse de doctorat, Université d'Avignon – 2007

Benoit Favre, Frédéric Béchet, Patrice Bellot, Florian Boudin, Marc El-Bèze, Laurent Gillard, Guy Lapalme, Juan-Manuel Torres-Moreno, *“The LIA-Thales summarization system at DUC-2006”*, Document Understanding Conference Workshop, HLT-NAACL’06, New York (USA) – 2006

Benoit Favre, *“Indexation multimédia : caractérisation du déséquilibre entre les modalités texte et parole”*, Mémoire de DEA Informatique, Université d’Avignon – 2003

Logiciels libres

- **icsiboost, une implémentation opensource de BoosTexter**
Boosting decision stumps – <http://code.google.com/p/icsiboost>
- **icsisumm, résumé automatique par programmation linéaire en nombres entiers**
a participé à TAC’08-09 – <http://code.google.com/p/icsisumm>
- **icsicrf-grecneg, génération d’expressions référentielles**
a participé à GREC’09 – <http://code.google.com/p/icsicrf-grecneg>
- **srleval, évaluation de l’étiquetage en rôles sémantiques**
capable de traiter des transcriptions automatiques de parole – <http://code.google.com/p/srleval>
- **miralium, classifieur de séquences avec apprentissage MIRA**
Apprentissage en ligne, compatible avec CRF++ – <http://code.google.com/p/miralium>