

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 years old, French driving license updated: 2010-01-21

Education

- **2007, PhD in computer science**
Laboratoire d'Informatique d'Avignon (LIA), University of Avignon (France)
- **2003, Master Degree in computer science**
Fundamental computer science, University of Avignon
- **2001, Bachelor's Degree in computer science**
mathematics and computer science engineering, networks speciality, University of Avignon

Research

- **Postdoc research**
Coupling semantic interpretation and automatic speech recognition for spoken dialogs (French and Italian). Funded as part of the PORT MEDIA project.
Punctuation detection in speech from various genres (broadcast news, broadcast conversations, meetings) and languages (English, Arabic, Mandarin). Impact of punctuation on subsequent tasks. Long span dependency models. Funded as part of the GALE project.
Automatic summarization of meeting recordings. Relevant keyphrase detection and refinement with user interactions. Global optimization frameworks for summarization. Part of the AMIDA and CALO projects.
- **PhD research work**
We explored the problematics of *Automatic Speech Summarization* applied to *Spoken Document Retrieval* to ease browsing of huge audio information databases. Specific areas studied:
 - spoken information extraction and structuring
 - extractive statistical summarization
 - user interfaces to spoken databasesWork co-funded by the French *Ministère de la Recherche et de l'Industrie* and Thales Land & Joint Systems.
- **Scientific Evaluation Campaigns**
 - ESTER French spoken named entity extraction campaign in 2005
 - DUC and TAC text summarization campaign from 2006 to 2009
 - GALE machine translation evaluation in 2007 to 2009 (in charge of repunctuating speech in source language)
 - GREC text generation evaluation in 2009 (generation of referential expressions)

Work Experience

- **2009 to present - Research Engineer**
Laboratoire d'Informatique de l'Université du Maine, Le Mans (France).
- **2007 to 2009 - Postdoc**
International Computer Science Institute, Berkeley (USA).
- **2004 to 2007 - Research Engineer**
Thales Land & Joint Systems Colombes (France), Multimedia Processing Laboratory.
- **2004 to 2006 - Web master of Association Francophone de la Communication Parlée (AFCP)**
Converted the website to Spip (php based content management portal)
Involved in the French national project Technolanguage

- **2001 to 2003 - Company creation project - *Allegorian***
Automatic speech indexing to allow content access to multimedia documents.
 - studied the speech application market
 - developed a business model and communication strategy
 - created a distributed speech indexing platform and spoken information retrieval engine.
- **2002 - Final year project - LIA**
Adapted a spoken document retrieval engine to French.
- **2001 - 2nd year training, University of Avignon**
Ported to Linux a business management software for PARTS EQUIPMENT, a French reseller.

Technical Skills

- **Programming:** Perl, Java, C++, C, Python
- **Machine Learning:** AdaBoost, CRF, SVM
- **Software Engineering:** UML, Corba
- **System:** Linux/Unix, OSX
- **Networks:** Hardware, Administration, TCP/IP, Nat, Bind, Dhcp
- **Databases:** Postgres, Mysql, Access, Oracle
- **Internet:** Apache, HTML, PHP, Javascript, Ajax, Flash
- **Languages:** written and spoken scientific English, native French

Teaching (156 hours)

- **2005 to 2006 (16 hours) - Data compression, M.Sc., University of Avignon**
Ultra-low bandwidth speech compression using a phoneme decoder and speech synthesis
- **2003 to 2004 (80 hours) - Networks tutorials, BSc level, University of Avignon**
RJ45 cables, network traffic analysis, CISCO routers
- **2000 (60 hours) - Office computing tutorials, University of Avignon**
Taught MS Office to Biology and Literature students.

Publications

- **Journals (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 Graciarena, 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*", to appear in 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

• **Peer reviewed international conferences (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, Lisboa (Portugal) – 2005

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

- **Peer reviewed national conferences (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), Fès (Morocco) – 2004

- **Other (7)**

Daniel Gillick, Benoit Favre, Dilek Hakkani-Tür, Bernd Bohnet, Yang Liu, Shasha Xie, "The ICSI/UTD Summarization System at TAC 2009", to appear in 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", PhD Thesis, University of Avignon (France) – 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”*, Master Thesis, University of Avignon (France) – 2003

Opensource software

- **icsiboost, an opensource implementation of BoosTexter**
Boosting decision stumps – <http://code.google.com/p/icsiboost>
- **icsisumm, integer linear programming based summarization system**
participated to TAC’08-09 – <http://code.google.com/p/icsisumm>
- **icsicrf-grecneg, generating referential expressions**
participated to GREC’09 – <http://code.google.com/p/icsicrf-grecneg>
- **srlevel, semantic role labeling evaluation**
can process automatic speech transcription output – <http://code.google.com/p/srlevel>
- **miralium, sequence classifier with MIRA learning**
Online learning, compatible with CRF++ – <http://code.google.com/p/miralium>