



MINISTÈRE DE L'ÉDUCATION  
NATIONALE, DE L'ENSEIGNEMENT  
SUPÉRIEUR ET DE LA RECHERCHE

MAISON DE LA  
RECHERCHE

# AVIS DE SOUTENANCE DE THESE DE DOCTORAT

**Mademoiselle Marie-Jean MEURS soutiendra une thèse  
le 10 décembre 2009 à 11h**

**CERI**

SPÉCIALITÉ : INFORMATIQUE ED 166

Titre de la thèse : Approche stochastique bayésienne de la composition sémantique pour les modules de compréhension automatique de la parole dans les systèmes de dialogue homme-machine

Membres du jury :

LAMEL Lori, DR Informatique, CNRS

BELLEGARDA Jerome, DR Informatique, Apple Inc. , Cupertino, Californie, Etats-Unis d'Amérique,

BESACIER Laurent, PR Informatique, Université Joseph Fourier, Grenoble,

BRETIER Philippe, Dr Informatique, France Télécom, Lannion,

NEY Hermann, Prof. Dr.-Ing. Informatique, Rheinische Westfälische Technische Hochschule Aachen, Allemagne,

DE MORI Renato, Pr Informatique, Université d'Avignon et des Pays de Vaucluse,

LEFEVRE Fabrice, MCF-HDR Informatique, Université d'Avignon et des Pays de Vaucluse,

Résumé de la thèse :

Les systèmes de dialogue homme-machine ont pour objectif de permettre un échange oral efficace et convivial entre un utilisateur humain et un ordinateur. Leurs domaines d'applications sont variés, depuis la gestion d'échanges commerciaux jusqu'au tutorat ou l'aide à la personne.

Cependant, les capacités de communication de ces systèmes sont actuellement limités par leur aptitude à comprendre la parole spontanée.

Nos travaux s'intéressent au module de compréhension de la parole et présentent une proposition entièrement basée sur des approches stochastiques, permettant l'élaboration d'une hypothèse sémantique complète. Notre démarche s'appuie sur une représentation hiérarchisée du sens d'une phrase à base de frames sémantiques.

La première partie du travail a consisté en l'élaboration d'une base de connaissances sémantiques adaptée au domaine du corpus d'expérimentation MEDIA (information touristique et réservation d'hôtel). Nous avons eu recours au formalisme FrameNet pour assurer une généricité maximale à notre représentation sémantique. Le développement d'un système à base de règles et d'inférences logiques nous a ensuite permis d'annoter automatiquement le corpus.

La seconde partie concerne l'étude du module de composition sémantique lui-même. En nous appuyant sur une première étape d'interprétation littérale produisant des unités conceptuelles de base (non reliées), nous proposons de générer des fragments sémantiques (sous-arbres) à l'aide de réseaux bayésiens dynamiques. Les fragments sémantiques générés fournissent une représentation sémantique partielle du message de l'utilisateur. Pour parvenir à la représentation sémantique globale complète, nous proposons et évaluons un algorithme de composition d'arbres décliné selon deux variantes.

La première est basée sur une heuristique visant à construire un arbre de taille et de poids minimum. La seconde s'appuie sur une méthode de classification à base de séparateurs à vaste marge pour décider des opérations de composition à réaliser.

Le module de compréhension construit au cours de ce travail peut être adapté au traitement de tout type de dialogue.

Il repose sur une représentation sémantique riche et les modèles utilisés permettent de fournir des listes d'hypothèses sémantiques scorées. Les résultats obtenus sur les données expérimentales confirment la robustesse de l'approche proposée aux données incertaines et son aptitude à produire une représentation sémantique consistante.

UNIVERSITÉ D'AVIGNON  
ET DES PAYS DE VAUCLUSE

COLLEGE DES ETUDES DOCTORALES  
CASE 20

74 rue Louis Pasteur  
84029 AVIGNON CEDEX 1  
<http://www.univ-avignon.fr>

tél : +33(0)4 90 16 25 29

fax : +33(0)4 90 16 27 44

joelle.derbaisse@univ-avignon.fr