



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

# AVIS DE SOUTENANCE DE THESE DE DOCTORAT

**Monsieur Eric CHARTON soutiendra une thèse  
le 12 novembre 2010 à 14h**

---

**LIA**

SPÉCIALITÉ : INFORMATIQUE

Titre de la thèse : Génération de phrases multilingues par apprentissage automatique de modèles de phrases

Membres du jury :

DANLOS Laurence PR Linguistique informatique Université Paris 7,  
LAPALME Guy, PR Linguistique informatique, Université de Montréal, Canada,  
GAGNON Michel, PR Linguistique informatique, École Polytechnique de Montréal,  
Canada,  
BONASTRE Jean François, PR TAL, Université d'Avignon et des Pays de Vaucluse,  
BESACIER Laurent, PR TAL, Université de Grenoble,  
TORRES-MORENO Juan-Manuel, MCF HDR TAL, Université d'Avignon et des Pays de  
Vaucluse.

Résumé de la thèse :

La Génération Automatique de Texte (GAT) est le champ de recherche de la linguistique informatique qui étudie la possibilité d'attribuer à une machine la faculté de produire du texte intelligible.

Nous présentons une proposition de système de GAT reposant exclusivement sur des méthodes statistiques. Son originalité est d'exploiter un corpus en tant que ressource de formation de phrases. Cette méthode offre plusieurs avantages : elle simplifie l'implémentation d'un système de GAT en plusieurs langues et améliore les capacités d'adaptations d'un système de génération à un domaine sémantique particulier.

La production, d'après un corpus d'apprentissage, des modèles de phrases finement étiquetées requises par notre générateur de texte nous a conduit à mener des recherches approfondies dans le domaine de l'extraction d'information et de la classification. Nous décrivons le système d'étiquetage et de classification de contenus encyclopédique mis au point à cette fin.

Dans les étapes finales du processus de génération, les modèles de phrases sont exploités par un module de génération de texte multilingue. Ce module exploite des algorithmes de recherche d'information pour extraire du modèle une phrase préexistante, utilisable en tant que support sémantique et syntaxique de l'intention à communiquer.

Plusieurs méthodes sont proposées pour générer une phrase, choisies en fonction de la complexité du contenu sémantique à exprimer. Nous présentons notamment parmi ces méthodes une proposition originale de génération de phrases complexes par agrégation de proto-phrases de type Sujet, Verbe, Objet.

Nous envisageons dans nos conclusions que cette méthode particulière de génération puisse ouvrir des voies d'investigations prometteuses sur la nature du processus de formation de phrases.

UNIVERSITÉ D'AVIGNON  
ET DES PAYS DE VAUCLUSE

Collège des Etudes Doctorales  
case 3

74 rue Louis Pasteur  
84029 Avignon cedex 1

+33 (0) 4 90 16 25 29 (tél)  
+33 (0) 4 90 16 25 31 (fax)

bureau 1W69

etudes-doctorales@univ-avignon.fr