

---

## Résumés de thèses

### Rubrique préparée par Sylvain Pogodalla

*Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France  
sylvain.pogodalla@inria.fr*

---

**Mokhtar Boumedyen BILLAMI** : mokhtar.billami@lis-lab.fr

**Titre** : Désambiguïsation sémantique dans le cadre de la simplification lexicale : contributions à un système d'aide à la lecture pour des enfants dyslexiques et faibles lecteurs

**Mots-clés** : Désambiguïsation sémantique, simplification lexicale, traitement automatique des langues, enfants dyslexiques, faibles lecteurs.

**Title**: *Word Sense Disambiguation within Lexical Simplification: Contributions to a Reading Support System for Children with Dyslexia and Poor Readers*

**Keywords**: *Word sense disambiguation, lexical simplification, natural language processing, dyslexic children, poor readers.*

**Thèse de doctorat** en Sciences du Langage, Laboratoire d'Informatique et Systèmes (LIS), Département Informatique et Interactions (DII), Faculté des Sciences, Aix-Marseille Université, sous la direction de Nuria Gala Pavia (MC HDR, Aix-Marseille Université) et Johannes Ziegler (DR, CNRS). Thèse soutenue le 15/11/2018.

**Jury** : Mme Nuria Gala Pavia (MC HDR, Aix-Marseille Université, codirectrice), M. Johannes Ziegler (DR, CNRS, codirecteur), M. Olivier Ferret (IC HDR, CEA LIST, rapporteur), M. Mathieu Lafourcade (MC HDR, Université Montpellier 2, rapporteur), Mme Cécile Fabre (Pr, Université Toulouse 2, examinatrice), M. Laurent Prévot (Pr, Aix-Marseille Université, examinateur).

**Résumé** : *La lecture est fondamentale pour tout ce qu'un enfant doit apprendre pendant son parcours scolaire. D'après des rapports nationaux (MJENR 2003) ou internationaux (PISA 2009), 20% à 30% des élèves français sont de faibles lecteurs et ont des difficultés pour comprendre les textes écrits, 5% à 10% sont des enfants dys-*

*lexiques. Ces lecteurs sont en grande difficulté face à des textes complexes ou avec un vocabulaire peu courant.*

*Ces dernières années, un nombre important de technologies ont été créées pour venir en aide aux personnes ayant des difficultés pour lire des textes écrits. Les systèmes proposés intègrent des technologies de la parole (lecture à « voix haute ») ou des aides visuelles (paramétrage ou mise en couleur des polices, ou augmentation de l'espace entre lettres et lignes). Cependant, il est essentiel de proposer aussi des transformations sur le contenu afin d'avoir des substituts de mots plus simples et plus fréquents. Cela permettra de rendre les textes plus accessibles et plus faciles à lire et à comprendre. Le but de cette thèse est de contribuer à un système d'aide à la lecture permettant de proposer automatiquement une version simplifiée d'un texte donné tout en gardant le même sens des mots.*

*Le travail présenté traite du problème de l'ambiguïté sémantique (très courant en traitement automatique des langues) et vise à proposer des solutions pour la désambiguïsation sémantique à l'aide de méthodes non supervisées et à base de connaissances provenant de ressources lexico-sémantiques. Dans un premier temps, nous proposons un état de l'art sur les méthodes de désambiguïsation sémantique et les mesures de similarité sémantique (essentiels pour la désambiguïsation sémantique). Par la suite, nous comparons divers algorithmes de désambiguïsation sémantique afin d'identifier le meilleur. Enfin, nous présentons nos contributions pour la création d'une ressource lexicale pour le français proposant des synonymes désambiguïsés et gradués en fonction de leur niveau de difficulté de lecture et compréhension. Nous montrons que cette ressource est utile et peut être intégrée dans un module de simplification lexicale de textes.*

**URL où le mémoire peut être téléchargé :**

<https://hal.archives-ouvertes.fr/tel-01969248>

---

**Chloé CLAVEL** : [chloe.clavel@telecom-paristech.fr](mailto:chloe.clavel@telecom-paristech.fr)

**Titre** : Analyse des opinions dans les interactions : de la fouille de données à l'interaction humain-agent

**Mots-clés** : Informatique affective, traitement automatique des langues, analyse d'opinions, interaction humain-agent, agents conversationnels animés.

**Title**: *Opinion Analysis in Interactions: from Data Mining to Human-Agent Interaction*

**Keywords**: *Affective computing, natural language processing, opinion analysis, human-agent interaction, embodied conversational agents.*

**Habilitation à diriger des recherches** en Informatique, Laboratoire Traitement et Communication de l'Information, UFR 919, Université Pierre et Marie Curie – Paris 6. Habilitation soutenue le 29/05/2017.

**Jury :** M. Dirk Heylen (Pr, Université de Twente, Pays-Bas, rapporteur), M. Nicolas Sabouret (Pr, Université Paris-Sud, rapporteur), M. Björn Schuller (Pr, Imperial College London, rapporteur), M. Frédéric Béchet (Pr, Aix-Marseille Université, examinateur), M. Patrick Paroubek (Ingénieur de recherche, Université Paris-Sud, examinateur), M. Mohamed Chetouani (Pr, Université Pierre et Marie Curie – Paris 6, président).

**Résumé :** *Le premier axe de mon activité de recherche concerne l'analyse des opinions dans les interactions humain-humain (conversations téléphoniques des centres d'appels, enquêtes de satisfaction, forums de relation client, données de micro-blogging données conversationnelles avec le conseiller virtuel d'EDF et interactions face à face avec un agent conversationnel animé). J'ai pu dans ce cadre constituer, à partir des données d'entreprise, des corpus In-the-wild riches en expressions spontanées. C'est sur ces corpus et d'autres corpus académiques disponibles que j'ai développé des modèles d'opinions et de sentiments. Les modèles d'opinions construits reposent sur des grammaires constituées de lexiques et de règles linguistiques et sur des méthodes d'apprentissage hybrides permettant d'intégrer des connaissances linguistiques au sein d'algorithmes d'apprentissage automatique. Dans le cadre des corpus issus d'interactions orales, j'ai travaillé sur la caractérisation des phénomènes de parole spontanée (disfluence, parole superposée) et sur la modélisation jointe des paramètres acoustiques et des marqueurs linguistiques. Dans le cadre des corpus issus d'interactions écrites, les structures énonciatives propres à la communication virtuelle et ses spécificités rédactionnelles ont également été étudiées. Dans le contexte de l'interaction humain-agent, mes contributions portent sur la mise en place d'une méthode de détection des opinions et des sentiments qui ancre son fonctionnement dans le contexte de l'interaction, avec la prise en compte, d'une part, du contexte dialogique (paires adjacentes et énoncés précédents de l'utilisateur) et des modalités communicatives de l'agent et, d'autre part, de la structure de Topic donnée par le scénario d'interaction. Par ailleurs, j'ai choisi de travailler sur des modélisations fines du phénomène des opinions/sentiments, en le circonscrivant aux goûts de l'utilisateur (likes et dislikes) et aux opinions des utilisateurs envers l'interaction (pour la détection d'interactions problématiques).*

*Le deuxième axe de mon activité de recherche porte sur les stratégies d'interactions socioémotionnelles de l'agent face à un utilisateur humain et sur la génération chez l'agent d'énoncés porteurs d'attitudes. La complémentarité de ces deux axes repose sur l'intégration du système d'analyse de sentiments décrit dans le paragraphe précédent dans les stratégies d'interactions socioémotionnelles de l'agent. Nous nous sommes focalisés sur le contenu verbal avec la mise en place de stratégies permettant de décider de l'alignement ou non de l'agent sur l'attitude de l'utilisateur et d'instancier les paramètres de réalisation d'un tel alignement. Enfin, concernant la génération des énoncés de l'agent, nous avons travaillé au niveau prosodique, notamment. Deux approches méthodologiques ont été choisies : une approche basée sur l'annotation manuelle des signaux prosodiques suivie d'une analyse statistique des corrélats perceptifs et une approche basée sur l'apprentissage automatique de séquences de si-*

*gnaux à partir d'un corpus avec l'extraction automatique de descripteurs prosodiques et l'extraction automatique de règles temporelles d'association.*

---

**Mnasri MAALI** : maali.mnasri@gmail.com

**Titre** : Résumé automatique multi-document dynamique

**Mots-clés** : Similarité sémantique, regroupement, ILP, analyse discursive.

**Title**: *Multi-Document Update-Summarization*

**Keywords**: *Semantic similarity, clustering, ILP, discourse analysis.*

**Thèse de doctorat** en Informatique, CEA LIST, LVIC, Université Paris-saclay, sous la direction de Gaël de Chalendar (IC, CEA LIST). Thèse soutenue le 20/09/2018.

**Jury** : M. Gaël de Chalendar (IC, CEA LIST, directeur), Mme Sophie Rosset (DR, CNRS, LIMSI, présidente), M. Jean-Luc Minel (Pr émérite, Université Paris Nanterre, rapporteur), M. Juan-Manuel Torres-Moreno (MC HDR, Université d'Avignon, rapporteur), M. Antoine Doucet (Pr, Université de La Rochelle, examinateur), M. Olivier Ferret (IC HDR, CEA LIST, encadrant scientifique).

**Résumé** : *Cette thèse s'intéresse au résumé automatique de texte et plus particulièrement au résumé mis à jour. Cette problématique de recherche vise à produire un résumé différentiel d'un ensemble de nouveaux documents par rapport à un ensemble de documents supposés connus. Elle intègre ainsi dans la problématique du résumé à la fois la question de la dimension temporelle de l'information et celle de l'historique de l'utilisateur. Dans ce contexte, le travail présenté s'inscrit dans les approches par extraction fondées sur une optimisation linéaire en nombres entiers (ILP) et s'articule autour de deux axes principaux : la détection de la redondance des informations sélectionnées et la maximisation de leur saillance. Pour le premier axe, nous nous sommes plus particulièrement intéressés à l'exploitation des similarités interphrastiques pour détecter, par la définition d'une méthode de regroupement sémantique de phrases, les redondances entre les informations des nouveaux documents et celles présentes dans les documents déjà connus. Concernant notre second axe, nous avons étudié l'impact de la prise en compte de la structure discursive des documents, dans le cadre de la Théorie de la Structure Rhétorique (RST), pour favoriser la sélection des informations considérées comme les plus importantes. L'intérêt des méthodes ainsi définies a été démontré dans le cadre d'évaluations menées sur les données des campagnes TAC et DUC. Enfin, l'intégration de ces critères sémantiques et discursifs au travers d'un mécanisme de fusion tardive a permis de montrer dans le même cadre la complémentarité de ces deux axes et le bénéfice de leur combinaison.*

**URL où le mémoire peut être téléchargé :**

<https://tel.archives-ouvertes.fr/tel-01902781>

---