



24e Conférence sur le Traitement Automatique des Langues Naturelles (TALN)

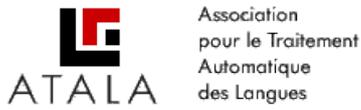
Orléans, France – 26-30 juin 2017

<https://taln2017.cnrs.fr>

Actes de TALN 2017, volume 1 : articles longs

Iris Eshkol, Jean-Yves Antoine (Eds.)

Sponsors :



Préface

Bienvenue à TALN-RECITAL-SITAL 2017

La 24^e édition de la conférence TALN et 19^e édition du RECITAL se déroulent cette année à Orléans, en plein cœur de la région Centre Val de Loire, célèbre pour ses châteaux, inscrite au patrimoine mondial de l'Unesco et située à 125 km au sud de Paris. Nous accueillons des participants sur le campus boisé d'Orléans du lundi 26 au vendredi 30 juin 2017. La conférence est organisée par trois laboratoires de référence dans le domaine du TAL, de l'informatique et de la linguistique de corpus : LLL (Laboratoire Ligérien de Linguistique), LIFO (Laboratoire d'Informatique Fondamentale d'Orléans) et LI (Laboratoire Informatique) de Tours. Après les Sables d'Olonne, Marseille, Paris, la conférence arrive dans une région mondialement réputée pour sa contribution au rayonnement de la langue française. En marge de la conférence, des activités culturelles, sportives et gastronomiques seront proposées pour découvrir la région du Val de Loire tout au long de la semaine, en après-midi ainsi qu'en soirée.

La conférence accueille 150 participants. Le premier jour, lundi, est consacré à quatre ateliers thématiques : « DEFT Fouille de texte » consacré cette année à l'analyse d'opinion et langage figuratif dans des messages postés sur Twitter ; la seconde édition du « HackaTAL » consacrée au résumé automatique de description de produits à partir de leurs commentaires et la prédiction automatique de la brevetabilité de termes liés à des technologies selon leur historique ; l'atelier « Les corpus annotés du français : ressources disponibles et exploitation en TAL » proposé pour la première fois à TALN et l'atelier « DiLiTAL – Diversité linguistique et TAL » dédié au traitement automatique des langues peu dotées. Nous remercions les organisateurs de ces ateliers pour leurs propositions, leur animation et leur organisation de ces événements. Les trois jours qui suivent associent dans de mêmes séances TALN et la conférence jeunes chercheurs RECITAL. Le dernier jour, vendredi 30 juin, est dédié au salon de l'innovation (SITAL) destiné à accueillir les entreprises et les chercheurs dans ce domaine du numérique de plus en plus émergent avec le développement du big data sur les données textuelles, ceci afin d'échanger leurs idées sur les développements actuels et futurs du domaine. Dans le cadre du SITAL, trois tables rondes « TAL et Humanités numériques », « Traitement automatique de la langue biomédicale » et « TAL dans l'expérience utilisateur : analyses et outils » sont proposées. Nous remercions les animateurs des tables rondes pour leur aide dans l'organisation de ces événements. Les tables rondes sont suivies par la session des posters et de démonstrations qui permettent aux participants de discuter plus longuement sur leurs travaux et de faire connaissance avec de nouveaux logiciels développés. La conférence se termine par une grande table ronde intitulée

« TAL aujourd’hui et demain : nouvelles méthodes, nouveaux usages, nouvelles applications ». Tout au long de la conférence, les étudiants ont la possibilité de s’inscrire et de participer aux déjeuners avec des experts d’un domaine.

Nous accueillons deux conférenciers invités. Tout d’abord Shuly Wintner (Université de Haifa, Israël) spécialiste en grammaires formelles, morphologie, traduction automatique et acquisition du langage, et Laura Kallmeyer (Université de Duesseldorf, Allemagne) qui travaille depuis des années sur l’analyse syntaxique, les grammaires d’arbres adjoints et l’interface syntaxe-sémantique. Nous les remercions d’avoir accepté notre invitation.

TALN2017 et RECITAL2017 ont permis aux chercheurs de présenter leurs travaux sous forme de communications orales, de posters et de démonstrations. 122 papiers au total ont été soumis aux deux conférences, 71 papiers (58%) ont été acceptés. Sur 28 articles longs soumis, 14 articles (50%) ont été acceptés. Sur 60 articles courts proposés, 30 articles (50%) ont été acceptés dont 8 pour la présentation orale et 22 en poster. Enfin, 17 papiers ont été soumis pour la session de démonstrations dont 14 ont été acceptés. En ce qui concerne la conférence RECITAL, 17 articles ont été soumis, 13 articles ont été acceptés dont 5 pour la présentation orale et 8 en poster.

La procédure de relecture et de la prise de décision finale est une procédure complexe, c’est pourquoi nous remercions le comité de relecture de TALN et de RECITAL ainsi que le comité de programme pour le temps consacré, pour leur patience et leur bonne volonté. Nous tenons à remercier le comité permanent de la conférence (CPERM) et son président pour l’aide dans l’organisation de l’événement et dans la prise de décisions pour certaines questions. Nous remercions le comité d’organisation constitué de chercheurs de trois laboratoires LLL, LIFO et LI, d’agents CNRS, de doctorants et d’étudiants de master et de licence. C’est une équipe formidable qui a mis tout en œuvre pour pouvoir permettre le meilleur accueil et le meilleur déroulement de la conférence. Nous tenons à remercier nos partenaires et nos sponsors : ATALA, DGLFLF, Orléans Métropole, ORTOLANG, les entreprises Inbenta, Aktan, Acatus, Proxem, sans lesquels cette édition de TALN, de RECITAL et le salon SITAL n’aurait pas pu avoir lieu. Un grand merci à l’Université d’Orléans et plus particulièrement à la Faculté LLSH d’avoir prêté les locaux et d’avoir permis l’organisation de cet événement dans les meilleures conditions.

Iris Eshkol-Taravella et Jean-Yves Antoine
Co-Présidents de TALN2017

Comités

Présidente du comité d'organisation de TALN 2017

Iris Eshkol-Taravella (LLL-UMR 7270, Université d'Orléans)

Vice-président du comité d'organisation de TALN 2017

Jean-Yves Antoine (LI-EA 6300, Université de Tours)

Président du comité d'organisation de RECITAL 2017

Yannick Parmentier (LIFO-EA 4022, Université d'Orléans)

Vice-présidente du comité d'organisation de RECITAL 2017

Hélène Flamein (LLL-UMR 7270, Université d'Orléans)

Présidente du salon d'innovation de TALN 2017

Gabrielle Bosshard (Aktan, Orléans)

Vice-présidente du salon d'innovation de TALN 2017

Sandra Cestic (LLL-UMR 7270, Acatu Informatique, Orléans)

Comité d'Organisation

Ahmed Abid (LI, Université de Tours)

Catherine Aléonard (LLL, Université d'Orléans)

Flora Badin (LLL, Université d'Orléans)

Gabriel Bergounioux (LLL, Université d'Orléans)

Sylvie Billot (LIFO, Université d'Orléans)

Marwa Boulakbech (LI, Université de Tours)

Caroline Cance (LLL, Université d'Orléans)

Guillaume Cleuziou (LIFO, Université d'Orléans)

Hyun Jung Kang (LLL, Université d'Orléans)
Anaïs Lefeuvre-Halftermeyer (LIFO, Université d'Orléans)
Denis Maurel (LI, Université de Tours)
Agata Savary (LI, Université de Tours)
Emmanuel Schang (LLL, Université d'Orléans)

Étudiants du Master Linguistique, spécialité LASTIC de l'Université d'Orléans

Hélène Couderc
Laetitia Delay
Siqi Fan
Lorraine Gaspard
Mélanie Lefeuvre
Sara Masaud
Stéphanie Nogueira
Camille Pertin
Cathy Querineau
Céline Vaschalde
Jidong Xie

Comité de Programme de TALN

Jean-Yves Antoine (Université de Tours, LI)
Delphine Bernhard (LiLPa - Université de Strasbourg)
Laurent Besacier (Université de Grenoble, LIG)
Nathalie Camelin (LIUM - Université du Maine)
Benoît Crabbé (Université Paris Diderot, LLF)
Iris Eshkol-Taravella (Université d'Orléans, LLL)
Olivier Ferret (CEA LIST)
Claire Gardent (CNRS, LORIA)
Thierry Hamon (Université Paris 13, LIMSI)
Philippe Langlais (Université de Montréal, RALI)
Emmanuel Morin (Université de Nantes, LINA)
Philippe Muller (CNRS, CLLE)
Adeline Nazarenko (Université Paris-Nord, LIPN)

Comité de Relecture de TALN 2017

Stergos Afantenos, CNRS/Université Paul Sabatier
Salah Ait-Mokhtar, Xerox Research Centre Europe
Maxime Amblard, Université de Lorraine
Jean-Yves Antoine, LI, Université de Tours
Delphine Batistelli, MoDyCo, Université Paris Ouest Nanterre La Défense
Frédéric Bechet, Aix Marseille Université - LIF

Delphine Bernhard, LiLPa, Université de Strasbourg
Romaric Besançon, CEA LIST
Philippe Blache, CNRS & Université de Provence
Hervé Blanchon, Laboratoire d'Informatique de Grenoble - Equipe GETALP
Florian Boudin, Université de Nantes
Annelies Braffort, LIMSI-CNRS
Nathalie Camelin, LIUM, Université du Maine
Thierry Charnois, LIPN CNRS University of PARIS 13
Guillaume Cleuziou, LIFO, Université d'Orléans
Benoit Crabbé, Paris 7 et INRIA
Béatrice Daille, Laboratoire d'Informatique Nantes Atlantique (LINA)
Marco Dinarelli, Lattice-CNRS
Iris Eshkol-Taravella, University of Orléans
Yannick Estève, LIUM - Université du Maine
Cécile Fabre, Université Toulouse 2
Karèn Fort, Paris 4
Thomas François, Université Catholique de Louvain
Nathalie Friburger, LI, Université de Tours
Eric Gaussier, LIG-UJF
Natalia Grabar, STL CNRS Université Lille 3
Lamia Hadrich, MIRACL Laboratory
Nicolas Hernandez, Université de Nantes - LINA CNRS UMR 6241
Stéphane Huet, LIA - Université d'Avignon
Sylvain Kahane, Université Paris Ouest Nanterre & CNRS
Olivier Kraif, Université Stendhal Grenoble 3
Mathieu Lafourcade, LIRMM
Guy Lapalme, RALI-DIRO, Université de Montréal
Joseph Le Roux, Laboratoire d'Informatique de Paris Nord
Jean-Marc Lecarpentier, GREYC
Anaïs Lefeuvre-Halftermeyer LIFO, Université d'Orléans
Anne-Laure Ligozat, LIMSI-CNRS
Denis Maurel, LI, Université de Tours
Richard Moot, CNRS (LaBRI) & Bordeaux University
Véronique Moriceau, LIMSI-CNRS
Philippe Muller, IRIT, Toulouse University
Aurélié Névéol, CNRS
Jian-Yun Nie, Université de Montréal
Yannick Parmentier, LIFO, Université d'Orléans
Thierry Poibeau, LaTTiCe-CNRS
Andrei Popescu-Belis, IDIAP Research Institute
Jean-Philippe Prost, LIRMM, Université Montpellier 2
Solen Quiniou, LINA - Université de Nantes

Christian Raymond, UEB/INRIA/IRISA/INSA
Christian Retoré, Université de Montpellier ; LIRMM-CNRS
Mathieu Roche, Cirad, TETIS
Didier Schwab, Univ. Grenoble Alpes
Michel Simard, National Research Council Canada (NRC)
Kamel Smaili, LORIA
Xavier Tannier, LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay
Isabelle Tellier, PARIS 3, Lattice
Juan-Manuel Torres-Moreno, Laboratoire Informatique d'Avignon / UAPV
Christel Vrain, LIFO, university of Orléans
Eric Wehrli, University of Geneva
Guillaume Wisniewski, LIMSI-UPS
François Yvon, LIMSI/CNRS et Université Paris-Sud
Pierre Zweigenbaum, LIMSI-CNRS

Table des matières

Session « oral »

Ajout automatique de disfluences pour la synthèse de la parole spontanée : formalisation et preuve de concept <i>Raheel Qader, Gwénolé Lecorvé, Damien Lolive, Pascale Sébillot</i>	1
Normalisation automatique du vocabulaire source pour traduire depuis une langue à morphologie riche <i>Franck Burlot, François Yvon</i>	16
Représentations continues dérivées des caractères pour un modèle de langue neuronal à vocabulaire ouvert <i>Matthieu Labeau, Alexandre Allauzen</i>	32
Utilisation d'indices phraséologiques pour évaluer des textes en langue étrangère : comparaison des bigrammes et des trigrammes <i>Yves Bestgen</i>	47
Traitement des Mots Hors Vocabulaire pour la Traduction Automatique de Document OCRisés en Arabe <i>Kamel Bouzidi, Zied Elloumi, Laurent Besacier, Benjamin Lecouteux, Mohamed-Faouzi Benze-ghiba</i>	63
Représentation et analyse automatique des discontinuités syntaxiques dans les corpus arborés en constituants du français <i>Maximin Coavoux, Benoît Crabbé</i>	77
Construire des représentations denses à partir de thésaurus distributionnels <i>Olivier Ferret</i>	93
Projection Aléatoire Non-Négative pour le Calcul de Word Embedding / Non-Negative Randomized Word Embedding <i>Behrang Qasemizadeh, Laura Kallmeyer, Aurelie Herbelot</i>	109

Création et validation de signatures sémantiques : application à la mesure de similarité sémantique et à la substitution lexicale	
<i>Mokhtar Boumedyen Billami, Núria Gala</i>	123
Vers une solution légère de production de données pour le TAL : création d'un tagger de l'alsacien par crowdsourcing bénévole	
<i>Alice Millour, Karën Fort, Delphine Bernhard, Lucie Steiblé</i>	139
Sciences participatives et TAL : jusqu'où ? comment ? pourquoi ?	
<i>Jean-Yves Antoine, Anaïs Lefeuvre-Halftermeyer</i>	155
Construction automatique d'une base de données étymologique à partir du wiktionary	
<i>Benoît Sagot</i>	169
Apprendre des représentations jointes de mots et d'entités pour la désambiguïsation d'entités	
<i>José Moreno, Romaric Besançon, Romain Beaumont, Eva D'Hondt, Anne-Laure Ligozat, Sophie Rosset, Xavier Tannier, Brigitte Grau</i>	182
Analyse et évolution de la compréhension de termes techniques	
<i>Natalia Grabar, Thierry Hamon</i>	196