



6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition) (JEP-TALN-RÉCITAL) ¹

Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition).

Volume 3 : Rencontre des Étudiants Chercheurs en Informatique pour le TAL

Christophe Benzitoun, Chloé Braud, Laurine Huber, David Langlois, Slim Ouni, Sylvain Pogodalla, Stéphane Schneider (Éds.)

Nancy, France, 08-19 juin 2020

1. <https://jep-taln2020.loria.fr/>

Crédits : L'image utilisée en bannière est une photographie du vitrail « Roses et Mouettes », visible dans la maison Bergeret à Nancy. La [photographie](#) a été prise par Alexandre Prevot, diffusée sur flickr sous la licence [CC-BY-SA 2.0](#).

Le logo de la conférence a été créé par Annabelle Arena.

©2020 ATALA et AFCP

Avec le soutien de



Message des présidents de l’AFCP et de l’ATALA

En ce printemps 2020, et les circonstances exceptionnelles qui l’accompagnent, c’est avec une émotion toute particulière que nous vous convions à la 6e édition conjointe des Journées d’Études sur la Parole (JEP), de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) et des Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL). Après une première édition commune en 2002 (à Nancy, déjà!), et une expérience renouvelée avec succès en 2004, c’est désormais tous les quatre ans (Avignon 2008, Grenoble 2012) que se répète cet événement commun, attendu de pied ferme par les membres des deux communautés scientifiques voisines.

Cette édition 2020 est exceptionnelle, puisque dans le cadre des mesures sanitaires liées à la pandémie mondiale de COVID-19 (confinement strict, puis déconfinement progressif), la conférence ne peut avoir lieu à Nancy comme initialement prévu, mais se déroule à distance, sous forme virtuelle, soutenue par les technologies de l’information et de la communication. Nous remercions ici chaleureusement les organisateurs, Christophe Benzitoun, Chloé Braud, Laurine Huber, David Langlois, Slim Ouni et Sylvain Pogodalla, qui ont dû faire preuve de souplesse, d’inventivité, de détermination, de puissance de travail, et de tant d’autres qualités encore, afin de maintenir la conférence dans ces circonstances, en proposant un format inédit. Grâce aux différentes solutions mises en œuvre dans un délai court, la publication des communications scientifiques est assurée, structurée, et les échanges scientifiques sont favorisés, même à distance.

Bien entendu, nous regrettons tous que cette réunion JEP-TALN-RECITAL ne permette pas, comme ses prédécesseurs, de nouer ou renforcer les liens sociaux entre les différents membres de nos communautés respectives – chercheurs, jeunes et moins jeunes, académiques et industriels, professionnels et étudiants – autour d’une passionnante discussion scientifique ou d’un mémorable événement social. . . Notre conviction est qu’il est indispensable de maintenir à l’avenir de tels lieux d’échanges dans le domaine francophone, afin bien sûr de permettre aux jeunes diplômés de venir présenter leurs travaux et poser leurs questions sans la barrière de la langue, mais aussi de dynamiser nos communautés, de renforcer les échanges et les collaborations, et d’ouvrir la discussion autour des enjeux d’avenir, qui questionnent plus que jamais la place de la science et des scientifiques dans notre société.

Lors de la précédente édition, nous nous interrogeons sur les phénomènes et tendances liés à l’apprentissage profond et sur leurs impacts sur les domaines de la Parole et du TAL. Force est de constater que l’engouement pour ces approches dans nos domaines a permis un retour sur le devant de la scène des domaines liés à l’Intelligence Artificielle, animant parfois un débat tant philosophique que technique sur la place de la machine dans la société, notamment à travers le questionnement sur la vie privée de l’utilisateur. Ces questionnements impactent tant la Parole que le TAL, d’une part sur la place de la gestion des données, d’autre part sur les modèles eux-mêmes. Malgré ces questionnements, nous constatons que les acquis et les expertises perdurent, et les nouvelles approches liées à l’apprentissage profond ont permis un rapprochement des domaines de la Parole et du TAL, sans les dénaturer, à la manière des conférences JEP-TALN-RECITAL qui créent un espace plus grand d’échange et d’enrichissement réciproques.

Nous terminons ces quelques mots d’ouverture en remerciant l’ensemble des personnes qui ont rendu possible cet événement qui restera, nous l’espérons, riche et passionnant, malgré les circonstances. L’ATALA et l’AFCP tiennent tout d’abord à réitérer leurs remerciements aux organisateurs des JEP, de TALN et de RECITAL, qui sont parvenus à maintenir le cap à travers vents et marées. Nos remerciements vont également à l’ensemble des membres des comités de programme, dont le travail et l’implication ont permis de garantir la qualité et la cohérence du programme finalement retenu. Un grand merci aux relecteurs pour le temps et le soin qu’ils ont dédiés à ce travail anonyme et indispensable. Ils se reflètent dans la qualité des soumissions que chacun pourra découvrir sur le site de la conférence.

En conclusion, cette 6e édition conjointe JEP-TALN-RECITAL est exceptionnelle parce qu’elle se tient dans un contexte de crise généralisée — crise sanitaire, économique, voire sociale et politique. Mais nous

formons le vœu qu'elle reste également dans les annales pour la qualité des échanges scientifiques qu'elle aura suscités, et pour le message envoyé à nos communautés scientifiques et à la société dans son ensemble, un message de détermination et de confiance en l'avenir, où la science et les nouvelles technologies restent au service de l'humain.

Véronique Delvaux, présidente de l'Association Francophone de la Communication Parlée
Christophe Servan, président de l'Association pour le Traitement Automatique des Langues

Préface

En 2002, l'AFCP (Association Francophone pour la Communication Parlée) et l'ATALA (Association pour le Traitement Automatique des Langues) organisèrent conjointement leurs principales conférences afin de réunir en un seul lieu, à Nancy, les communautés du traitement automatique et de la description des langues écrites, parlées et signées.

En 2020, la sixième conférence commune revient à Nancy, après Fès (2004), Avignon (2008), Grenoble (2012) et Paris (2016). Elle est organisée par le LORIA (Laboratoire lorrain de recherche en informatique et ses applications, UMR 7503), l'ATILF (Analyse et traitement informatique de la langue française, UMR 7118) et l'INIST (Institut de l'information scientifique et technique) et regroupe :

- les 33^{es} Journées d'Études sur la Parole (JEP),
- la 27^e conférence sur le Traitement Automatique des Langues Naturelles (TALN),
- la 22^e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL).

Les circonstances particulières liées à l'épidémie de Covid-19 en France et dans le monde ont conduit à une virtualisation de la conférence. Ainsi, malgré un rassemblement physique qui n'a pu avoir lieu, diffusions, présentations (au gré des auteurs) et discussions des articles acceptés ont lieu sur le site internet de la conférence. Les tutoriels, certains ateliers, et le salon de l'innovation qui accompagnent la conférence ont cependant dû être annulés, mais les ateliers suivants sont maintenus :

- Défi Fouille de Textes (DEFT 2020),
- Éthique et TRaitement Automatique des Langues (ÉTeRNAL).

La conférence accueille également des conférencières et conférenciers invités dont les exposés sont diffusés sur le site : Dirk Hovy (université de Bocconi, Milan, Italie, invité ÉTeRNAL) ainsi que Marie-Jean Meurs (Université du Québec à Montréal, UQAM, Canada) et Hugo Cyr (Faculté de science politique et droit à l'Université du Québec à Montréal, UQAM, Canada). En raison des circonstances particulières, un exposé conjoint de Christine Meunier (Laboratoire Parole et Langage LPL, CNRS, Aix-en-Provence, France) et Christophe Stécoli (police technique et scientifique française) a dû être annulé et reporté à une journée spéciale en septembre 2020.

Ces actes regroupent les articles des conférences JEP (volume 1), TALN (volume 2), RÉCITAL (volume 3), les articles décrivant les démonstrations (volume 4), et les articles des ateliers DEFT (volume 5) et ÉTeRNAL (volume 6). Pour la première fois, un appel spécifique à résumés en français d'articles parus dans une sélection de conférences internationales en 2019 était également proposé (volume 4). Un appel spécifique apprenti·e·s chercheur·euse·s destiné aux étudiants de licence, de master, ou en première année de thèse a également été proposé, pour leur proposer des présentations courtes ou sous forme de poster de leurs projets.

Pour les JEP, 87 articles ont été soumis, parmi lesquels 74 ont été sélectionnés, soit un taux de sélection de 85%.

Pour TALN, 58 articles ont été soumis, parmi lesquels 37 ont été sélectionnés, soit un taux de sélection de 63%, dont 10 comme article longs (17% des soumissions) et 27 comme article courts dont 20 en présentation orale (34% des soumissions) et 7 en présentation poster (12% des soumissions).

Pour RÉCITAL, 22 articles ont été soumis, parmi lesquels 16 ont été sélectionnés, soit un taux de sélection de 73%.

Nous souhaitons vivement remercier toutes les personnes qui ont participé à ce travail de relecture et de sélection :

- l'ensemble des relecteurs (voir page xi),
- le comité de programme des JEP (voir page viii),
- le comité de programme de TALN (voir page ix),
- le comité de programme de RÉCITAL (voir page x).

Nous souhaitons également remercier nos sociétés savantes : l'AFCP, assurant la continuité des éditions successives des JEP, et l'ATALA, dont le CPerm (comité permanent) assure la continuité des éditions

successives de TALN.

Nous remercions le comité d'organisation et les nombreuses personnes qui ont assuré le soutien administratif et technique pour que cette conférence se déroule dans les meilleures conditions, et en particulier Yannick Parmentier pour son travail pour la diffusion de ces actes sur HAL et les différents sites d'archives ouvertes ([anthologie ACL](#) et [talnarchives.atala.org/](#)).

Nous remercions enfin tous les partenaires institutionnels et industriels qui nous ont fait confiance, en particulier l'université de Lorraine, le CNRS, l'Inria, le LORIA, l'ATILF, l'INIST, le master TAL de l'Institut des Sciences du Digital Management & Cognition (IDMC), le projet OLKI de l'initiative Lorraine Université d'Excellence (LUE), la Région Grand Est, *The Evaluations and Language resources Distribution Agency* (ELDA), le projet ANR PARSEME-FR, la délégation générale à la langue française et aux langues de France (DGLFLF), l'Association des Professionnels des Industries de la Langue (APIL) et les entreprises Synapse, Yseop et Orange.

Bonne conférence à toutes et à tous !

Les présidentes et présidents JEP :	David Langlois et Slim Ouni
TALN :	Chloé Braud et Sylvain Pogodalla
RÉCITAL :	Christophe Benzitoun et Laurine Huber

Comités

Comité de programme des JEP

Martine Adda-Decker (Laboratoire de Phonétique et Phonologie, CNRS)
Jean-Francois Bonastre (LIA, Université d'Avignon)
Fethi Bougares (LIUM, Le Mans Université) Philippe Boula De Mareüil (LIMSI, CNRS)
Hervé Bredin (LIMSI, CNRS)
Olivier Crouzet (LLING, Université de Nantes)
Elisabeth Delais-Roussarie (LLING, Université de Nantes)
Véronique Delvaux (Laboratoire de Phonétique, IRSTL, Université de Mons)
Camille Fauth (LiLPa, Université de Strasbourg)
Emmanuel Ferragne (CLILLAC-ARP, Université de Paris)
Cecile Fougeron (Laboratoire de Phonétique et Phonologie, CNRS)
Corinne Fredouille (LIA, Université d'Avignon)
Alain Ghio (LPL, CNRS)
Camille Guinaudeau (LIMSI, Université Paris Sud)
Anne Guyot Talbot (CLILLAC-ARP, Université de Paris 7)
Bernard Harmegnies (Laboratoire de Phonétique, IRSTL, Université de Mons)
Nathalie Henrich Bernardoni (Gipsa-lab, CNRS)
Bassam Jabaian (LIA, Université d'Avignon)
David Langlois (LORIA, Université de Lorraine)
Yves Laprie (LORIA, CNRS)
Anthony Larcher (LIUM, Université du Maine)
Gwénoél Lecorvé (IRISA, Université de Rennes)
Benjamin Lecouteux (LIG, Université Grenoble Alpes)
Georges Linarès (LIA, Université d'Avignon)
Damien Lolive (IRISA, Université Rennes)
Julie Mauclair (IRIT)
Yohann Meynadier (LPL, Aix-Marseille Université)
Slim Ouni (LORIA, Université de Lorraine)
Thomas Pellegrini (IRIT, Université de Toulouse)
François Portet (LIG, Grenoble INP)
Fabian Santiago (Structures Formelles du Langage, Université de Paris 8)
Christophe Savariaux (Gipsa-lab, CNRS)
Nathalie Vallee (Gipsa-lab, Université Grenoble Alpes)
Ioana Vasilescu (LIMSI, CNRS)

Comités de programme TALN

Maxime Amblard (LORIA, Université de Lorraine)
Chloé Braud (IRIT, CNRS)
Caroline Brun (Naver Labs Europe)
Nathalie Camelin (LIUM, Université du Maine)
Marie Candito (Université Paris 7)
Vincent Claveau (IRISA, CNRS)
Chloé Clavel (Telecom-ParisTech)
Mathieu Constant (ATILF, CNRS, Université de Lorraine)
Pascal Denis (Inria)
Cécile Fabre (Université Toulouse 2)
Thomas François (Université catholique de Louvain)
Núria Gala (LPL, CNRS, Aix-Marseille Université)
Natalia Grabar (STL, CNRS, Université Lille 3)
Anne-Laure Ligozat (LIMSI, CNRS, ENSIE, Université Paris-Saclay)
Emmanuel Morin (LINA, Université de Nantes)
Sylvain Pogodalla (LORIA, Inria)
Solen Quiniou (LINA, Université de Nantes)
Corentin Ribeyre (Etermind)
Tim van de Cruys (IRIT, CNRS)
Pierre Zweigenbaum (LIMSI, CNRS, Université Paris-Saclay)

Comité de programme RÉCITAL

Jean-Yves Antoine (Université François Rabelais de Tours)
Sonia Badene (Linagora, IRIT)
Frédéric Béchet (LIF, Aix Marseille Université)
Christophe Benzitoun (ATILF, Université de Lorraine)
Maria Boritchev (LORIA, Inria)
Léo Bouscarrat (EURA NOVA, Aix-Marseille Université)
Manon Cassier (INALCO, Paris)
Kevin Deturck (Viseo Technologies)
Emmanuelle Esperança-Rodier (GETALP, Université Grenoble Alpes)
Kim Gerdes (sorbonne nouvelle)
Nicolas Hernandez (LINA, UMR 6241, CNRS, Université de Nantes)
Lydia-Mai Ho-Dac (CLLE-ERSS, Université Toulouse Jean Jaurès)
Laurine Huber (LORIA, Université de Lorraine)
Sylvain Kahane (Modyco, Université Paris Ouest Nanterre)
Gwénolé Lecorvé (IRISA, Université de Rennes, CNRS)
Joël Legrand (LORIA, Inria, CNRS)
Anne-Laure Ligozat (LIMSI, CNRS, ENSIE, Université Paris-Saclay)
Pierre Ludmann (LORIA, Université de Lorraine)
Yann Mathet (Université de Caen)
Anne-Lyse Minard (IRISA, CNRS)
Sandrine Ollinger (ATILF, UMR 7118, CNRS)
Yannick Parmentier (LORIA, Université de Lorraine)
Justine Reynaud (LORIA, Université de Lorraine)
Stella Zevio (LIPN, Université de Paris 13)

Relectrices et relecteurs

- Gilles Adda (LIMSI, CNRS) Salah Ait-Mokhtar (Naver Labs Europe)
- Charlotte Alazard (Université Toulouse 2 Jean Jaurès)
- Alexandre Allauzen (LIMSI-CNRS, Université Paris-Sud)
- Pascal Amsili (Université Paris Diderot)
- Pierre André Hallé (Laboratoire de Phonétique et Phonologie, CNRS–Université Paris 3)
- Régine André-Obrecht (Université Paul Sabatier Toulouse III)
- Jean-Yves Antoine (Université François Rabelais de Tours)
- Nicolas Audibert (Laboratoire de Phonétique et Phonologie, CNRS–Université Paris 3)
- Nelly Barbot (IRISA, Université de Rennes 1)
- Claude Barras (LIMSI, CNRS)
- Loïc Barrault (University of Sheffield)
- Katarina Bartkova (ATILF, Université de Lorraine)
- Frédéric Béchet (LIF, Aix Marseille Université)
- Nathalie Bedoin (DDL, Université Lyon 2)
- Patrice Bellot (LSIS, CNRS, Aix-Marseille Université)
- Asma Ben Abacha (National Library of Medicine, National Institutes of Health)
- Delphine Bernhard (LiLPa, Université de Strasbourg)
- Roxane Bertrand (LPL, CNRS, Aix-Marseille Université)
- Laurent Besacier (Laboratoire d'Informatique de Grenoble)
- Yves Bestgen (F.R.S-FNRS et Université Catholique de Louvain)
- Frédéric Bimbot (IRISA, CNRS)
- Caroline Bogliotti (MODYCO, UMR 7114, CNRS, Université Paris Nanterre)
- Anne Bonneau (LORIA, CNRS)
- Stéphanie Borel (Université de Tours)
- Féthi Bougarès (LIUM, Le Mans Université)
- Leila Boutora (Laboratoire Parole et Langage, Aix Marseille Université)
- Paul Caillon (LORIA, Université de Lorraine)
- Mélanie Canault (DDL, Université Lyon 2)
- Thierry Charnois (LIPN, CNRS, Université de Paris 13)
- Chloé Clavel (Telecom-ParisTech)
- Maximin Coavoux (Université Grenoble Alpes, CNRS)
- Vincent Colotte (LORIA, Université de Lorraine)
- Juan Manuel Coria (LIMSI, Université Paris-Saclay Paris 13)
- Benoît Crabbé (Université Paris 7)
- Lise Crevier Buchman (Laboratoire de Phonétique et Phonologie, CNRS, Hôpital Foch)
- Béatrice Daille (LINA, Université de Nantes)
- Géraldine Damnati (Orange Labs)
- Dan Dediu (Dynamique du Langage, UMR5596, Université Lumière Lyon 2)
- Joseph Di Martino (LORIA, Université de Lorraine)
- Gaël Dias (Université Caen Normandie)
- Amazouz Djegdjiga (LPP, Université Sorbonne Nouvelle – Paris 3)
- Benjamin Elie (IMSIA, ENSTA ParisTech)
- Iris Eshkol-Taravella (Université d'Orléans)
- Emmanuelle Esperança-Rodier (GETALP, Université Grenoble Alpes)
- Yannick Estève (LIA, Université d'Avignon)
- Dominique Estival (Western Sydney University)
- Olivier Ferret (CEA LIST)
- Lionel Fontan (Archean Labs)
- Karën Fort (Sorbonne Université)
- Claire Gardent (LORIA, CNRS)
- Eric Gaussier (LIG, Université Grenoble Alpes)
- Cédric Gendrot (LPP, Université Sorbonne Nouvelle – Paris 3)
- James German (Laboratoire Parole et Langage, Aix Marseille Université)
- Cyril Goutte (National Research Council Canada)
- Cyril Grouin (LIMSI, CNRS, Université Paris-Saclay)
- Pierre André Hallé (LPP, Université Sorbonne Nouvelle – Paris 3)
- Olivier Hamon (Syllabs)
- Thierry Hamon (LIMSI, Université Paris-Saclay, CNRS, Université Sorbonne Paris Nord)
- Bernard Harmegnies (Institut de Recherche en Sciences et Technologies du Langage, Université de Mons)
- Nabil Hathout (CLLE, CNRS)
- Amir Hazem (LS2N, Université de Nantes)
- Nicolas Hernandez (LS2N, Université de Nantes)
- Fabrice Hirsch (Praxiling, Université Paul Valéry Montpellier 3)
- Thomas Hueber (GIPSA-lab, CNRS)
- Kathy Huet (Institut de Recherche en Sciences et Technologies du Langage, Université de Mons)

Stéphane Huet (LIA, Université d'Avignon)
 Mathilde Hutin (LIMSI, Université Paris-Saclay, CNRS, Université Sorbonne Paris Nord)
 Irina Illina (LORIA, Université de Lorraine)
 Christine Jacquin (LS2N Université de Nantes)
 Adèle Jatteau (STL, UMR 8163, Université de Lille, CNRS)
 Denis Jouvét (LORIA, Inria)
 Sylvain Kahane (Modyco, Université Paris Ouest Nanterre)
 Takeki Kamiyama (LPP, Université Paris 8 Vincennes-Saint-Denis)
 Hannah King (CLILLAC-ARP, Université Paris Diderot)
 Olivier Kraif (Université Grenoble Alpes)
 Matthieu Labeau (Telecom Paris)
 Mathieu Lafourcade (LIRMM, Université de Montpellier)
 Mohamed Lahrouchi (SFL, UMR 7023, CNRS Université Paris 8)
 Muriel Lalain (LPL, CNRS, Aix-Marseille Université)
 Joseph Lark (Dictanovia)
 Thomas Lavergne (LIMSI, CNRS, Univ. Paris Sud, Université Paris Saclay)
 Guillaume Le Berre (LORIA, Université de Lorraine)
 Gwénolé Lecorvé (IRISA, Université de Rennes, CNRS)
 Benjamin Lecouteux (Laboratoire Informatique de Grenoble)
 Claire Lemaire (Université Grenoble Alpes)
 Yves Lepage (Waseda University)
 Joseph Le Roux (LIPN, Université de Paris 13)
 Veronika Lux (ATILF, CNRS)
 Paolo Mairano (STL, UMR 8163, Université de Lille)
 Anna Marczyk (LPL, CNRS, Aix-Marseille Université)
 Denis Maurel (Université François Rabelais de Tours)
 Christine Meunier (LPL, CNRS, Aix-Marseille Université)
 Alexis Michaud (LACITO, CNRS)
 Richard Moot (LIRMM, CNRS)
 Véronique Moriceau (LIMSI, CNRS)
 Philippe Muller (IRIT, Université de Toulouse)
 Alexis Nasr (LIF, Université de la Méditerranée)
 Sylvain Navarro (CLLE-ERSS, CNRS)
 Luka Nerima (Université de Genève)
 Aurélie Névéol (LIMSI, CNRS, Université Paris-Saclay)

Jian-Yun Nie (Université de Montreal)
 Damien Nouvel (INaLCO)
 Nicolas Obin (IRCAM)
 Yannick Parmentier (LORIA, Université de Lorraine)
 Sebastian Peña Saldarriaga (Dictanovia)
 Marie Philippart de Foy (Université de Mons)
 Myriam Piccaluga (Institut de Recherche en Sciences et Technologies du Langage, Université de Mons)
 Claire Pillot-Loiseau (LPP, UMR 7018, CNRS, Université Sorbonne Nouvelle – Paris 3)
 Serge Pinto (LPL, CNRS, Aix-Marseille Université)
 Agnès Piquard (LORIA, CNRS, Université de Lorraine)
 Thierry Poibeau (LaTTiCe, CNRS)
 Alain Polguère (ATILF Université de Lorraine)
 Laurent Prévot (LPL, CNRS, Aix-Marseille Université)
 Jean-Philippe Prost (LIRMM, Université de Montpellier)
 Christian Raymond (IRISA, INSA de Rennes)
 Christian Retoré (LIRMM, Université de Montpellier)
 Albert Rilliard (LIMSI, CNRS, Université Paris-Saclay)
 Virginie Roland (Institut de Recherche en Sciences et Technologies du Langage, Université de Mons)
 Sophie Rosset (LIMSI, CNRS, Université Paris-Saclay)
 Véronique Sabadell (LPC, Aix Marseille Université)
 Stéphane Schneider (INIST, CNRS)
 Didier Schwab (Université Grenoble Alpes)
 Pascale Sébillot (IRISA, INSA de Rennes)
 Djamé Seddah (Almanach, Université Paris la Sorbonne)
 Gilles Serasset (LIG, Université Grenoble Alpes)
 Romain Serizel (LORIA, Université de Lorraine)
 Kamel Smaïli (LORIA, Université de Lorraine)
 Rudolph Sock (LiLPa, Université de Strasbourg)
 Ludovic Tanguy (CLLE, CNRS)
 Xavier Tannier (LIMICS, Sorbonne Université, INSERM)
 Andon Tchechmedjiev (IMR, Mines Alès)
 Juan-Manuel Torres-Moreno (LIA, Université d'Avignon)
 Nicolas Turenne (LISIS, INRA)
 Béatrice Vaxelaire (LiLPa, Université de Strasbourg)

Anne Vilain (GIPSA-lab, Université de Grenoble Alpes)

Coriandre Vilain (GIPSA-lab, Université de Grenoble Alpes)

Guillaume Wisniewski (LLF, Université de Paris)

Jane Wottawa (LIUM, Le Mans Université)

Yaru Wu (LPP, MoDyCo, Université Paris Nanterre)

Kossi Seto Yibokou (LiLPa, Université de Strasbourg)

François Yvon (LIMSI, CNRS, Université Paris-Sud)

Table des matières

Segmentation de texte non-supervisée pour la détection de thématiques à l'aide de plongements lexicaux	1
<i>Alexandra Benamar</i>	
Spécificités des erreurs d'orthographe des personnes dyslexiques : analyse d'un corpus de productions écrites	15
<i>Johana Bodard</i>	
Ré-entraîner ou entraîner soi-même ? Stratégies de pré-entraînement de BERT en domaine médical	29
<i>Hicham El Boukkouri</i>	
Évaluation systématique d'une méthode commune de génération	43
<i>Hugo Boulanger</i>	
Analyse de la régulation de la longueur dans un système neuronal de compression de phrase : une étude du modèle LenInit	57
<i>François Buet</i>	
Exploitation de modèles distributionnels pour l'étude de la nomination dans un corpus d'interviews politiques	71
<i>Manon Cassier</i>	
L'adaptabilité comme compétence pour les systèmes de dialogue orientés tâche	85
<i>Oralie Cattan</i>	
Simplification de textes : un état de l'art	96
<i>Sofiane Elguendouze</i>	
Évolution phonologique des langues et réseaux de neurones : travaux préliminaires	110
<i>Clémentine Fourier</i>	
Comparing PTB and UD information for PDTB discourseconnective identification	123
<i>Kelvin Han, Phyllicia Leavitt, Srilakshmi Balard</i>	
Transformations syntaxiques entre niveaux de simplification dans le corpus Newsela	137
<i>Rita Hijazi</i>	
La désambiguïsation des abréviations du domaine médical	151
<i>Anaïs Koptient</i>	
Apprentissage de plongements de mots sur des corpus en langue de spécialité : une étude d'impact	164
<i>Valentin Pelloin, Thibault Prouteau</i>	
Représentation vectorielle de paires de verbes pour la prédiction de relations lexicales	179
<i>Etienne Rigaud</i>	
TTS voice corpus reduction for audio-book generation	193
<i>Meysam Shamsi</i>	

Exploiter des modèles de langue pour évaluer des sorties de logiciels d'OCR pour des documents français du XVIIe siècle

205

Jean-Baptiste Tanguy