



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).

Atelier DÉfi Fouille de Textes

Rémi Cardon, Natalia Grabar, Cyril Grouin, Thierry Hamon (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

Table des matières

Présentation de la campagne d'évaluation DEFT 2020 : similarité textuelle en domaine ouvert et extraction d'information précise dans des cas cliniques	1
<i>Rémi Cardon, Natalia Grabar, Cyril Grouin, Thierry Hamon</i>	
Calcul de similarité entre phrases : quelles mesures et quels descripteurs ?	14
<i>Davide Buscaldi, Ghazi Felhi, Dhaou Ghoul, Joseph Le Roux, Gaël Lejeune, Xudong Zhang</i>	
Participation d'EDF R&D à DEFT 2020	26
<i>Danrun Cao, Alexandra Benamar, Manel Boumghar, Meryl Bothua, Lydia Ould Ouali, Philippe Suignard</i>	
Contextualized French Language Models for Biomedical Named Entity Recognition	36
<i>Jenny Copara, Julien Knafou, Nona Naderi, Claudia Moro, Patrick Ruch, Douglas Teodoro</i>	
Approche supervisée de calcul de similarité sémantique entre paires de phrases	49
<i>Khadim Dramé, Gorgoumack Sambe, Ibrahima Diop, Lamine Faty</i>	
DEFT 2020 - Extraction d'information fine dans les données cliniques : terminologies spécialisées et graphes de connaissance	55
<i>Thomas Lemaitre, Camille Gosset, Mathieu Lafourcade, Namrata Patel, Guilhem Mayoral</i>	
DOING@DEFT : cascade de CRF pour l'annotation d'entités cliniques imbriquées	66
<i>Anne-Lyse Minard, Andréane Roques, Nicolas Hiot, Mirian Halfeld Ferrari Alves, Agata Savary</i>	
Extraction d'information de spécialité avec un système commercial générique	79
<i>Clothilde Royan, Jean-Marc Langé, Zied Abidi</i>	
DEFT 2020 : détection de similarité entre phrases et extraction d'information	91
<i>Mike Tapi Nzali</i>	
Similarité sémantique entre phrases : apprentissage par transfert interlingue	97
<i>Charles Teissèdre, Thiziri Belkacem, Maxime Arens</i>	
Participation de l'équipe du LIMICS à DEFT 2020	108
<i>Perceval Wajsbürt, Yoann Taillé, Guillaume Lainé, Xavier Tannier</i>	