



18e Conférence en Recherche d'Information et Applications
16e Rencontres Jeunes Chercheurs en RI
30e Conférence sur le Traitement Automatique des Langues Naturelles
25e Rencontre des Étudiants Chercheurs en Informatique pour le
Traitement Automatique des Langues
(CORIA-TALN) ¹

Actes de CORIA-TALN 2023.

Actes des 16e Rencontres Jeunes Chercheurs en RI (RJCRI) et 25e Rencontre des
Étudiants Chercheurs
en Informatique pour le Traitement Automatique des Langues (RÉCITAL)

Marie Candito, Thomas Gerald, José G Moreno (Éds.)

Paris, France, 5 au 9 juin 2023

1. <https://coria-taln-2023.sciencesconf.org/>

Avec le soutien de



Préface

Organisée conjointement par les laboratoires franciliens sous l'égide de l'Association francophone de Recherche d'Information et Applications (ARIA) et l'Association pour le Traitement Automatique des Langues (ATALA), la conférence CORIA-TALN-RJCRI-RECITAL 2023 regroupe :

- la 18e Conférence en Recherche d'Information et Applications (CORIA)
- la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN) ;

ainsi que les deux conférences associées, destinées aux jeunes chercheuses et chercheurs :

- Les 16e Rencontres Jeunes Chercheurs en RI (RJCRI)
- la 25e Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL)

Ces deux derniers événements ont été nommés cette année les "Rencontres Jeunes Chercheurs", et ont fait l'objet d'un processus de relecture et de sélection unique. Nous avons reçu 13 soumissions, dont 11 ont été acceptées (à comparer pour RECITAL en 2022, 15 soumissions et 9 articles acceptés).

Lors de la conférence, qui se tient à Paris, du 6 au 9 juin 2023, les articles RJC sont présentés dans des sessions communes aux conférences CORIA et/ou TALN. Quatre des onze articles RJC sont présentés oralement, et les sept autres sous forme de poster.

Comme c'est le propre des Rencontres Jeunes Chercheurs, cette édition conjointe RJCRI/RECITAL a de nouveau été l'occasion pour de jeunes chercheuses et chercheurs, de se former à l'écriture d'articles scientifiques, de diffuser leurs travaux en cours et d'obtenir des avis de la communauté scientifique, tant lors du processus de relecture que, pour les articles acceptés, lors de leur présentation pendant la conférence.

Nous tenons à remercier chaleureusement les membres du comité de programme des RJC, pour leurs relectures exigeantes mais constructives. Nous espérons que celles-ci ont permis aux autrices et auteurs ayant soumis aux RJC d'avancer dans leurs travaux.

Un grand merci également à toute l'équipe d'organisation de l'édition 2023 ainsi qu'aux sponsors, pour tout le travail réalisé et le soutien financier afin que ces rencontres soient une réussite.

Marie Candito, Thomas Gérald et José G. Moreno, co-présidents des RJC 2023

Comités

Comité de programme

Présidence

- Marie Candito
- Thomas Gerald
- José Moreno

Membres

- Catherine Berrut
- Sylvie Calabretto
- Thierry Charnois
- Maximin Coavoux
- Kata Gabor
- Aina Garí Soler
- Gaël Guibon
- Paul Lerner
- Jesus Lovon
- Aurélie Névéol
- Yannick Parmentier
- Karen Pinel-Sauvagnat
- Carlos Ramisch
- Arnaud Soulet
- Julien Velcin

Table des matières

Les jeux de données en compréhension du langage naturel et parlé : paradigmes d'annotation et représentations sémantiques	1
<i>Rim Abrougui</i>	
Étude de la fidélité des entités dans les résumés par abstraction	21
<i>Eunice Akani</i>	
Mise en place d'un modèle compact à architecture Transformer pour la détection jointe des intentions et des concepts dans le cadre d'un système interactif de questions-réponses	37
<i>Nadège Alavoine, Arthur Babin</i>	
Utiliser les syntagmes nominaux complexes anglais pour évaluer la robustesse des systèmes de traduction anglais-français en langue de spécialité	57
<i>Maud Bénard</i>	
Vers une implémentation de la théorie sens-texte avec les grammaires catégorielles abstraites	72
<i>Marie Cousin</i>	
Analyse de la légitimité des start-ups	87
<i>Asmaa Lagrid</i>	
Approches neuronales pour la détection des chaînes de coréférences : un état de l'art	101
<i>Fabien Lopez</i>	
Etudes sur la géolocalisation de Tweets	114
<i>Thibaud Martin</i>	
IR-SenTransBio : Modèles Neuronaux Siamois pour la Recherche d'Information Biomédicale	131
<i>Safaa Menad</i>	
L'évaluation de la traduction automatique du caractère au document : un état de l'art	143
<i>Mariam Nakhlé</i>	
Normalisation lexicale de contenus générés par les utilisateurs sur les réseaux sociaux	160
<i>Lydia Nishimwe</i>	