



---

*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)*<sup>1</sup>

Actes de CORIA-TALN 2023.

Actes de la 18e Conférence en Recherche d'Information et Applications (CORIA)

---

Haïfa Zargayouna (Éds.)

Paris, France, 5 au 9 juin 2023

---

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



Avec le soutien de



## Préface

Après Rennes en 2018, l'ARIA (Association francophone de Recherche d'Information et Applications) et l'ATALA (Association pour le Traitement Automatique des Langues) ont décidé de se retrouver à nouveau pour organiser conjointement leur principale conférence.

Au cours de cette manifestation, se tiendront les conférences CORIA (CONFérence en Recherche d'Information et Applications) – pour sa dix-huitième édition et TALN (conférence sur le Traitement Automatique des Langues Naturelles) – pour sa trentième édition. À ces deux conférences s'ajoutent les journées jeunes chercheurs, à savoir, les seizième Rencontres Jeunes Chercheurs en RI (RJCRI) et les vingt-cinquième Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL). Ces conférences sont les points de rassemblement des communautés francophones respectivement en recherche d'information et en traitement automatique des langues.

Cette année, le comité de programme de CORIA a choisi 18 papiers pour une présentation orale (10 papiers longs, 2 papiers courts et 6 papiers déjà publiés dans des conférences ou revues internationales de renom). Les papiers couvrent des thèmes de recherche d'information assez divers avec une forte coloration autour des architectures neuronales et des modèles de langue.

Pour favoriser les interactions entre les deux communautés, nous avons intégré dans l'organisation des présentations quatre sessions communes à la recherche d'information et au traitement automatique des langues. Pour compléter le contenu scientifique des conférences, le programme de la manifestation a été enrichi par la venue de trois conférenciers invités dont deux de la communauté de la Recherche d'Information :

- Jaap Kamps est enseignant chercheur à l'Université d'Amsterdam, ses recherches autour de la Recherche d'Information se font à l'Institut pour la logique, le langage et le calcul (Institute for Logic, Language, and Computation).
- Jian-Yun Nie est professeur à Université de Montréal et titulaire d'une chaire de recherche du Canada en traitement de langue naturelle. Il travaille dans le laboratoire RALI (Recherche Appliquée en Linguistique Informatique). Ses recherches couvrent divers sujets en recherche d'information (RI) et traitement de langue naturelle.

Enfin, je tiens à remercier l'ensemble des membres du comité d'orientation et le bureau de l'ARIA, les membres du comité de programme pour l'évaluation des soumissions, les organisateurs Franciliens pour leur grande disponibilité, ainsi que toutes les personnes qui se sont investies pour la réussite de cette manifestation. Je remercie, également, les auteurs grâce à qui CORIA continue d'exister.

Je vous souhaite à toutes et à tous de belles journées scientifiques.

**Haïfa Zargayouna – Présidente du Comité de Programme CORIA 2023**

## Comités

### Comité de programme

#### Présidente

Haïfa ZARGAYOUNA

#### Membres

- Ismail BADACHE
- Patrice BELLOT
- Catherine BERRUT
- Romaric BESANCON
- Robert BOSSY
- Mohand BOUGHANEM
- Davide BUSCALDI
- Sylvie CALABRETTO
- Max CHEVALIER
- Jean-Pierre CHEVALLET
- Adrian CHIFU
- Vincent CLAVEAU
- Antoine DOUCET
- Sebastien FOURNIER
- Eric GAUSSIER
- Mathias GÉRY
- Lorraine GOEURIOT
- Gilles HUBERT
- Jose MORENO
- Véronique MORICEAU
- Philippe MULHEM
- Diana NURBAKOVA
- Jian-Yun NIE
- Karen PINEL-SAUVAGNAT
- Benjamin PIWOWARSKI
- Mathieu ROCHE
- Eric SAN JUAN
- Jacques SAVOY
- Christophe SERVAN
- Laure SOULIER
- Lynda TAMINE
- Nicolas TURENNE
- Haïfa ZARGAYOUNA

## Table des matières

<b>I</b>	<b>Articles longs</b>	<b>1</b>
	<b>Impact de l'apprentissage multi-labels actif appliqué aux transformers</b>	<b>2</b>
	<i>Maxime Arens, Charles Teissède, Lucile Callebert, Jose G Moreno, Mohand Boughanem</i>	
	<b>Quelles évolutions sur cette loi ? Entre abstraction et hallucination dans le domaine du résumé de textes juridiques</b>	<b>18</b>
	<i>Nihed Bendahman, Karen Pinel-Sauvagnat, Gilles Hubert, Mokhtar Boumedyen Billami</i>	
	<b>Augmentation de jeux de données RI pour la recherche conversationnelle à initiative mixte</b>	<b>37</b>
	<i>Pierre Erbacher, Philippe Preux, Jian-Yun Nie, Laure Soulier</i>	
	<b>Apprentissage de sous-espaces de préfixes</b>	<b>59</b>
	<i>Louis Falissard, Vincent Guigue, Laure Soulier</i>	
	<b>Recherche cross-modale pour répondre à des questions visuelles</b>	<b>74</b>
	<i>Paul Lerner, Ferret Olivier, Camille Guinaudeau</i>	
	<b>Adaptation de domaine pour la recherche dense par annotation automatique</b>	<b>93</b>
	<i>Minghan Li, Eric Gaussier</i>	
	<b>Extraction d'entités nommées à partir de descriptions d'espèces</b>	<b>111</b>
	<i>Maya Sahraoui, Vincent Guigue, Régine Vignes-Lebbe, Marc Pignal</i>	
	<b>Le théâtre français du XVIIIe siècle : une expérience en catégorisation de textes</b>	<b>127</b>
	<i>Jacques Savoy</i>	
	<b>Enrichissement des modèles de langue pré-entraînés par la distillation mutuelle des connaissances</b>	<b>139</b>
	<i>Raphaël Sourty, Jose G Moreno, François-Paul Servant, Lynda Tamine</i>	
	<b>Constitution de sous-fils de conversations d'emails</b>	<b>157</b>
	<i>Lionel Tadonfouet Tadjou, Eric De La Clergerie, Fabrice Bourge, Tiphaine Marie</i>	
<b>II</b>	<b>Articles courts</b>	<b>172</b>
	<b>Intégration du raisonnement numérique dans les modèles de langue : État de l'art et direction de recherche</b>	<b>173</b>
	<i>Sarah Abchiche, Lynda Said Lhadj, Vincent Guigue, Laure Soulier</i>	
	<b>Reconnaissance d'Entités Nommées fondée sur des Modèles de Langue Enrichis avec des Définitions des Types d'Entités</b>	<b>185</b>
	<i>Jesús Lovón Melgarejo, Jose Moreno, Romaric Besançon, Olivier Ferret, Lynda Tamine</i>	
<b>III</b>	<b>Articles déjà publiés</b>	<b>195</b>
	<b>Entity Enhanced Attention Graph-Based Passages Retrieval</b>	<b>196</b>
	<i>Lucas Albarede, Lorraine Goeuriot, Philippe Mulhem, Claude Le Pape-Gardeux, Sylvain Marie, Trinidad</i>	

<b>Highlighting exact matching via marking strategies for ad hoc document ranking with pretrained contextualized language models</b>	<b>201</b>
<i>Lila Boualili, Jose Moreno, Mohand Boughanem</i>	
<b>Vers l'évaluation continue des systèmes de recherche d'information.</b>	<b>202</b>
<i>Petra Galuscakova, Romain Deveaud, Gabriela Gonzalez-Saez, Philippe Mulhem, Lorraine Goeuriot, Florina Piroi, Martin Popel</i>	
<b>CoSPLADE : Adaptation d'un Modèle Neuronal Basé sur des Représentations Partielles pour la Recherche d'Information Conversationnelle</b>	<b>207</b>
<i>Nam Le Hai, Thomas Gerald, Thibault Formal, Jian-Yun Nie, Benjamin Piwowarski, Laure Soulier</i>	
<b>The Power of Selecting Key Blocks with Local Pre-ranking for Long Document Information Retrieval</b>	<b>213</b>
<i>Minghan Li, Diana Nicoleta Popa, Johan Chagnon, Yagmur Gizem Cinar, Eric Gaussier</i>	
<b>iQPP : Une Référence pour la Prédiction de Performances des Requêtes d'Images</b>	<b>214</b>
<i>Eduard Poesina, Radu Tudor Ionescu, Josiane Mothe</i>	
<b>IV Démonstration</b>	<b>221</b>
<b>XPMIR : Une bibliothèque modulaire pour l'apprentissage d'ordonnancement et les expériences de RI neuronale</b>	<b>222</b>
<i>Yuxuan Zong, Benjamin Piwowarski</i>	

Première partie  
Articles longs