



# CORIA-TALN 2025

---

*20e Conférence en Recherche d'Information et Applications (CORIA)  
32ème Conférence sur le Traitement Automatique des Langues  
Naturelles (TALN)  
27ème Rencontre des Étudiants Chercheurs en Informatique pour le  
Traitement Automatique des Langues (RECITAL)  
Les 18e Rencontres Jeunes Chercheurs en RI (RJCRI)  
(CORIA-TALN)<sup>1</sup>*

Actes de CORIA-TALN-RJCRI-RECITAL 2025.

Actes des 18e Rencontres Jeunes Chercheurs en RI (RJCRI) et 27ème Rencontre des  
Étudiants Chercheurs  
en Informatique pour le Traitement Automatique des Langues (RECITAL)

---

Frédéric BECHET, Adrian-Gabriel CHIFU, Karen PINEL-SAUVAGNAT, Benoit FAVRE, Eliot MAES,  
Diana NURBAKOVA (Éds.)

Marseille, France, 30 juin au 4 juillet 2025

---

1. <https://coria-taln-2025.lis-lab.fr>



Avec le soutien de

Organisateurs



Soutiens académiques



Sponsors privés



## Préface

Organisée par l'Université d'Aix-Marseille et les UMR CNRS LIS et LPL, sous l'égide de l'Association francophone de Recherche d'Information et Applications (ARIA) et l'Association pour le Traitement Automatique des Langues (ATALA), l'édition 2025 de CORIA-TALN regroupe :

- la 20e Conférence en Recherche d'Information et Applications (CORIA) ;
- la 32e Conférence sur le Traitement Automatique des Langues Naturelles (TALN) ;
- les 27e Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL) ;
- les 18e Rencontres Jeunes Chercheurs en RI (RJCRI).

Les conférences CORIA et TALN offrent le plus important forum d'échange francophone aux acteurs universitaires et industriels des technologies de la langue et la recherche d'information. Pour cette édition, nous avons plus de 330 inscrits dont une grande partie des étudiants qui construisent le futur de la recherche francophone et assurent le relais de son développement. La conférence principale compte plus de 260 inscrits et les ateliers plus de 70 inscrits.

Les conférencières et conférenciers invités de la conférence sont Marine CARPUAT, de l'Université du Maryland (USA), Mohamed CHETOUANI de Sorbonne Université, et Owen RAMBOW de Stony Brook University (USA). Ces invités représentent un large spectre de thématiques dans le domaine de la recherche d'information et le traitement automatique des langues, et partageront les dernières avancées dans leur domaine d'expertise.

En termes de soumissions, 101 articles ont été soumis à TALN, 76 acceptés dont 51 contributions scientifiques originales et 25 traductions d'articles déjà publiés dans des conférences internationales majeures du domaine. Il n'y a pas eu de différenciation dans le processus de sélection entre oraux et posters. Un total de 29 articles a été soumis à CORIA dont 25 ont été retenus pour présentation (8 courts, 13 longs et 4 résumés). RECITAL-RJCRI a accepté 17 articles sur 19 soumissions. En complément de la conférence principale, huit ateliers sont présentés : Avancement de l'AMR et de l'Analyse Sémantique (4AS), Accès à l'information basé sur le dialogue et grands modèles de langage (DIAG-LLM), Traitement de données langagières dynamiques par les outils et méthodes du TAL (DYN-TAL), Ethic and Alignment of (Large) Language Models (EALM), Évaluation des modèles génératifs (LLM) et challenge (EvalLLM), Intelligence Artificielle générative et ÉDUcation : Enjeux, Défis et Perspectives de Recherche (IA-ÉDU), Traitement du langage médical à l'époque des LLMs (MLP-LLM) et Science Participative pour les Données et Corpus Linguistiques (ParCoL). Ce programme est complété par une session industrielle ciblant explicitement des contributions qui montrent le développement des technologies associées au langage dans l'industrie. Ces événements illustrent à la fois des tendances nouvelles présentes dans la communauté et des activités récurrentes.

Il convient d'exprimer une profonde reconnaissance envers toutes les personnes qui ont participé à faire vivre la conférence, d'un côté les auteurs de toutes les soumissions et de l'autre les membres de différents comités scientifiques de la conférence. Un remerciement très chaleureux aux relecteurs qui ont accepté une charge importante et qui ont fait des relectures d'urgence afin de faciliter le bon déroulement de la conférence. La bienveillance et l'expertise des comités de programme ont permis la constitution d'un programme riche en thématiques et d'un niveau scientifique correspondant aux attentes de la communauté. Il est également essentiel d'exprimer notre gratitude envers les sponsors et les organisations qui ont subventionné la conférence. Leur soutien financier a permis à cet événement scientifique de se réaliser dans les meilleures conditions, rappelant l'importance des aspects financiers dans la réussite de telles initiatives. Finalement, un grand merci aux différentes équipes présentes pour le bon fonctionnement, notamment des équipes de l'ATALA et de l'ARIA qui nous ont accompagnés dans les différentes étapes

de l'organisation.

- Frédéric BÉCHET et Adrian-Gabriel CHIFU, présidents du comité d'organisation de CORIA-TALN
- Karen PINEL-SAUVAGNAT, présidente du comité de programme de CORIA
- Benoit FAVRE, président du comité de programme de TALN
- Eliot MAES et Diana NURBAKOVA, présidents du comité de programme de RECITAL-RJCRI

# Comités

## Comité de Programme

- Eliot MAËS (AMU, LIS) (président)
- Diana NURBAKOVA (INSA Lyon) (présidente)
- Ismail BADACHE (Computer Science and Systems Laboratory, Aix–Marseille University)
- Frederic BECHET (Aix Marseille Université – LIS)
- Patrice BELLOT (Aix–Marseille Université – CNRS (LIS))
- Florian BOUDIN (Université de Nantes)
- Davide BUSCALDI (LIPN, Université Paris 13, Sorbonne Paris Cité)
- Adrian–Gabriel CHIFU (LIS/AMU)
- Előd EGYED–ZSIGMOND (LIRIS)
- Eric GAUSSIÉ (LIG–UGA)
- Gabriela GONZÁLEZ SÁEZ (Université Grenoble Alpes)
- Stéphane HUET (LIA – Université d’Avignon)
- Guy LAPALME (University of Montreal)
- Gaël LEJEUNE (Paris Sorbonne University)
- Véronique MORICEAU (IRIT Université Toulouse 3)
- Josiane MOTHE (Univ. Toulouse 2 Jean Jaures, IRIT)
- Jian–Yun NIE (University de Montreal)
- Karen PINEL–SAUVAGNAT (Université de Toulouse, IRIT)
- Eric SANJUAN (Laboratoire Informatique d’Avignon– Université d’Avignon)
- Mathias VAST (Sorbonne University)

## Table des matières

<b>Amélioration de la lisibilité de textes via l'utilisation de LLM</b>	<b>1</b>
<i>Baptiste Ramonda, Isabelle Ferrane, Julien Pinquier</i>	
<b>Analyse de la littérature sur les stratégies d'augmentation de données dans des contextes à faible ressources</b>	<b>14</b>
<i>Benedictus Kent Rachmat</i>	
<b>Annotation de Marqueurs Discursifs : le cas de la désambiguïsation de après</b>	<b>31</b>
<i>Paola Herreño Castañeda, Maeva Sillaire</i>	
<b>Annotation et modélisation des émotions dans un corpus textuel : une approche évaluative</b>	<b>47</b>
<i>Jonas Noblet</i>	
<b>Comparaison des approches basées sur BERT et sur l'agent LLM pour la classification hiérarchique de narratifs dans les articles de presse multilingues</b>	<b>64</b>
<i>Yutong Wang, Mohamed-Nour Eldjadiri</i>	
<b>Corpus bilingue sous-titrage et Langue des Signes Française : la problématique de l'alignement automatique des données</b>	<b>91</b>
<i>Julie Halbout, Diandra Fabre</i>	
<b>État de l'art : évaluation, détection et mitigation des hallucinations des LLMs</b>	<b>104</b>
<i>Aygalic Jara-Mikolajczak</i>	
<b>État de l'art sur les marqueurs discursifs en Traitement Automatique des Langues</b>	<b>124</b>
<i>Fatou Sow</i>	
<b>Évaluation Automatique Explicable de l'Écriture Argumentative : État de l'Art, Lacunes et Proposition d'Architecture Modulaire Alignée sur des Grilles Éducatives</b>	<b>143</b>
<i>Marcos Moisés Crisóstomo de Oliveira</i>	
<b>Evaluation de la lisibilité des textes biomédicaux selon le profil du lecteur</b>	<b>160</b>
<i>Anya Nait Djoudi</i>	
<b>Image incomplète : Une étude d'état de l'art sur les biais dans les grands modèles de langage</b>	<b>175</b>
<i>Trung Hieu Ngo</i>	
<b>La traduction automatique dialectale : état de l'art et étude préliminaire sur le continuum dialectal de l'occitan</b>	<b>190</b>
<i>Oriane Nédey</i>	
<b>Normaliser le moyen français : du graphématique au semi-diplomatique</b>	<b>239</b>
<i>Sonia Solfrini, Mylène Dejouy, Aurélia Marques Oliveira, Pierre-Olivier Beaulnes</i>	
<b>Réhabiliter l'écriture Ajami : un levier technologique pour l'alphabétisation en Afrique</b>	<b>253</b>
<i>Samy Ouzerrout, Idriss Saadallah</i>	
<b>VERS : Versification Et Représentation de Séquences</b>	<b>268</b>

*Marceau Hernandez*

**Vers une taxonomie pour l'analyse des intentions dans les interactions textuelles numériques**

**289**

*Senaid Popovic*