



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) ¹*

Actes de CORIA-TALN-RJCRI-RECITAL 2025.

Actes de la 20e Conférence en Recherche d'Information et Applications (CORIA)

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



LABORATOIRE
D'INFORMATIQUE
& SYSTÈMES



Soutiens académiques



AFIA

Association française
pour l'Intelligence Artificielle



ATALA



ILCB

Institute of
Language, Communication
and the Brain



amU

Institut Laënnec, sciences
numériques et IA pour la santé
Aix Marseille Université

Sponsors privés

Sponsor « GOLD »



Smart
Tribune

EVERDIAN

aday



POSTLAB

Sponsors « SILVER »

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 CORIA

- Ismail BADACHE (Aix Marseille Univ, LIS)
- Frédéric BÉCHET (Aix Marseille Université – LIF)
- Patrice BELLOT (Aix-Marseille Université – CNRS (LIS))
- Catherine BERRUT (Université Joseph Fourier Grenoble I, LIG)
- Romaric BESANÇON (CEA LIST)
- Mohand BOUGHANEM (Université de Toulouse, IRIT)
- Davide BUSCALDI (Université Paris 13, LIPN)
- Sylvie CALABRETTO (INSA Lyon, LIRIS)
- Adrian-Gabriel CHIFU (LIS/AMU)
- Vincent CLAVEAU (AMIAD)
- Romain DEVEAUD (Qwant)
- Antoine DOUCET (Université de La Rochelle, L3I)
- Elöd EGYED-ZSIGMOND (INSA Lyon, LIRIS)
- Liana ERMAKOVA (Université de Bretagne Occidentale, HCTI)
- Sébastien FOURNIER (Université d’Aix-Marseille, LIS)
- Eric GAUSSIER (Univ. Grenoble Alpes, LIG)
- Mathias GÉRY (Université Jean Monnet, Saint-Etienne, LaHC)
- Lorraine GOEURIOT (Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG)
- Gaël GUIBON (Université de Lorraine – LORIA)
- Gilles HUBERT (Université de Toulouse, IRIT)
- Gaël LEJEUNE (Paris Sorbonne Université)
- José MORENO (Université de Toulouse, IRIT)
- Véronique MORICEAU (Université de Toulouse, IRIT)
- Josiane MOTHE (Université de Toulouse 2 Jean Jaurès, IRIT)
- Diana NURBAKOVA (INSA Lyon, LIRIS)
- Karen PINEL-SAUVAGNAT (Université de Toulouse, IRIT)
- Benjamin PIWOWARSKI (Sorbonne Université, CNRS / ISIR)
- Mathieu ROCHE (Cirad, TETIS)
- Eric SANJUAN (Université d’Avignon – Laboratoire Informatique d’Avignon)
- Laure SOULIER (Sorbonne Université-ISIR)
- Lynda TAMINE (Université de Toulouse, IRIT)
- Haïfa ZARGAYOUNA (Université Paris 13, LIPN)

Table des matières

Analyse Textuelle et Extraction Géospatiale pour la Surveillance des Crises Alimentaires en Afrique de l'Ouest	1
<i>Charles Abdoulaye Ngom, Maguelonne Teisseire, Sarah Valentin</i>	
Application de Transformers multimodaux à l'extraction d'informations des documents de sondage des sols	11
<i>Stanislas Bagnol, Killian Barrere, Véronique Eglin, Elöd Egyed-Zsigmond, Jean-Marie Côme, David Pitaval</i>	
Approche méthodologique pour la génération de question-réponse portant sur plusieurs documents	21
<i>Hui Huang, Julien Velcin, Yacine Kessaci</i>	
AutoCluster : Un agent pour le clustering basé sur les grands modèles de langue	31
<i>Erwan Versmée, Youcef Remil, Mehdi Kaytoue, Julien Velcin</i>	
Cadre d'évaluation pour les systèmes de génération augmentée (RAG) : combinaison des performances de recherche d'informations et de LLM	50
<i>Mohamed-Amine El-Yagoubi, Philippe Mulhem, Jean-Pierre Chevallet, Eric Gaussier</i>	
Clarification des Ambiguïtés : Sur le Rôle des Types d'Ambiguïté dans les Méthodes d'Amorçage pour la Génération de Clarifications	61
<i>Anfu Tang, Laure Soulier, Vincent Guigue</i>	
Clustering de résumés LLM guidés par l'utilisateur : vers une approche constructiviste et réaliste unifiée	82
<i>Carl Hatoum, Catherine Combes, Virginie Fresse, Christophe Gravier, Mathieu Orzalesi</i>	
Comprendre la Nature des Signaux de Correspondance dans les Modèles Neuronaux pour la RI	96
<i>Mathias Vast, Basile Van Cooten, Laure Soulier, Benjamin Piwowski</i>	
Entraînement avec solveur pour l'intégration de contraintes logiques dans l'extraction de relations d'événements	107
<i>Baptiste Brunet de la Charie, Abdallah Arioua, Elöd Egyed-Zsigmond, Thomas Veran</i>	
Évaluation des capacités des grands modèles de langue à comprendre les dossiers médicaux de patients : Une étude approfondie de l'extraction et la recherche de données des patients	126
<i>Jesús Lovón-Melgarejo, Martin Mouysset, Jo Oleiwan, Jose G. Moreno, Christine-Damase-Michel, Lynda Tamine</i>	
Génération augmentée de récupération multi-niveau pour répondre à des questions visuelles	128
<i>Omar Adjali, Olivier Ferret, Sahar Ghannay, Hervé Le Borgne</i>	
Génération augmentée de récupération pour les journaux historiques	131
<i>Trung Tran, Carlos-Emiliano González-Gallardo, Antoine Doucet</i>	
Optimisation de la Recherche d'Information Juridiques à travers l'Agrégation des Si-	

gnaux Contextuels Multi-niveaux des Modèles de Langue Préentraînés	135
<i>Eya Hammami, Mohand Boughanem, Taoufiq Dkaki</i>	
Prédiction des préférences et génération de revue personnalisée basées sur les aspects et attention	151
<i>Ben Kabongo, Vincent Guigue, Pirmin Lemberger</i>	
Quand les Bots Déjouent l'Apprentissage : Enjeux et Défis de la Détection	171
<i>Mohsine Aabid, Patrice Bellot, Simon Dumas Primbault</i>	
Rapido, interopérabilité et fouille de textes : vers un alignement des publications scientifiques en archéologie	180
<i>Justine Revol, Agnieszka Halczuk, Lucas Anki, Pascal Cuxac</i>	
Restructuration de la Littérature Biomédicale dans une Architecture RAG pour la Génération de Réponse	183
<i>Mael Lesavourey, Gilles Hubert</i>	
SEBRAG : Vers l'Utilisation des LLM pour une Tâche de Questions-Réponses Extractive	201
<i>Quentin Signé, Thiziri Belkacem, Jose G. Moreno, Mohand Boughanem</i>	
Seval-ex : Un paradigme basé sur les phrases atomiques pour une évaluation explicable de la qualité des résumés	217
<i>Tanguy Herserant, Vincent Guigue</i>	
Simplification de Textes Scientifiques (et Rien de Plus). Rapport sur l'Action CLEF 2025 SimpleText	230
<i>Liana Ermakova, Hosein Azarbondy, Jan Bakker, Benjamin Vendeville, Jaap Kamps</i>	
Transfert de modèles de langue pour la classification rhétorique des citations à travers les disciplines	233
<i>Anne-Sophie Foussat, Vincent Guigue, Nicolas Sauvion, Robert Bossy, Claire Nédellec</i>	
UC-FIRE : Approche efficace pour la recherche d'informations non supervisée	249
<i>Maxime Hanus, Quentin Guignard, Christophe Rodrigues, Léonard De Vinci</i>	
Utilisation de mécanismes inférentiels dans le processus d'explication automatique de la métaphore à une inconnue	265
<i>Jérémie Roux, Hani Guenoune, Mathieu Lafourcade, Richard Moot</i>	
Vers des interfaces favorisant l'engagement critique des utilisateurs : un prototype utilisant RAG	283
<i>Petra Dadić, Liana Ermakova</i>	
Vers un élagage de tokens sans coût dans les modèles de récupération à interaction tardive	294
<i>Yuxuan Zong, Benjamin Piwowarski</i>	