



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 session industrielle de CORIA-TALN 2025

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

La session industrielle de CORIA-TALN 2025 est la première édition de ce type pour cette conférence, elle met en avant les interactions entre recherche académique et applications industrielles des technologies du langage. Elle vise à réunir chercheurs, ingénieurs et utilisateurs pour discuter des défis et innovations liés au développement et au déploiement de ces technologies dans des environnements concrets. Les articles sélectionnés illustrent la mise en application des avancées en TALN dans le monde réel, en traitant des aspects tels que l'efficacité, la maintenabilité et l'évolutivité des systèmes. Sont particulièrement encouragées les contributions portant sur des méthodologies pratiques, des cas d'usage innovants et des défis techniques liés au déploiement de systèmes linguistiques. Cette session s'adresse tant aux chercheurs qu'aux acteurs industriels, sans exigence de mise à disposition des données propriétaires utilisées. Les thèmes incluent notamment l'optimisation des systèmes, les infrastructures à grande échelle, les approches interactives et l'éthique dans le déploiement des technologies linguistiques.

Comités

Comité d'organisation

- Géraldine Damnati – Orange Innovation, Lannion
- Luce Lefeuvre – SNCF, Paris
- Alexis Nasr – Aix-Marseille Université
- Frédéric Béchet – Aix-Marseille Université

Comité de programme

- Jeremy Auguste, Nunki, Paris
- Ismail Badache, LIS CNRS/Aix Marseille Université
- Frederic Bechet , LIS CNRS/Aix Marseille Université
- Patrice Bellot, LIS CNRS/Aix Marseille Université
- Laurent Besacier, Naver Labs, Grenoble
- Adrian-Gabriel Chifu, LIS CNRS/Aix Marseille Université
- Géraldine Damnati, Orange Innovation, Lannion
- Gaël Guibon, LIPN CNRS/Université Sorbonne Paris Nord
- Catherine Kobus, Airbus, Toulouse
- Luce Lefeuvre, SNCF, Paris
- Alexis Nasr, LIS CNRS/Aix Marseille Université
- Mickael Rouvier, LIA Université d'Avignon
- Anastasia Shimorina, Orange Innovation, Lannion

Table des matières

Apprentissage Actif à l'ère des Grands Modèles de Langue (LLMs)	1
<i>Shami Thirion Sen, Rime Abrougui, Guillaume Lechien, Damien Nouvel</i>	
Backtesting des signaux de sentiment pour le trading : évaluer la viabilité de la génération d'alpha à partir de l'analyse de sentiment	17
<i>Elvys Linhares Pontes, Carlos-Emiliano González-Gallardo, Georgeta Bordea, Jose G Moreno, Mohamed Ben Jannet, Yuxuan Zhao, Antoine Doucet</i>	
COLaF : Corpus et Outils pour les Langues de France et variétés de français	33
<i>Benoît Sagot, Slim Ouni, Sam Bigeard, Lucence Ing, Thibault Clérice, Rachel Bawden, Emmanuel Vincent, Malek Yaich, Panagiotis Tsolakis, Juliette Janès, Rasul Dent, Oriane Nédey, Vincent Colotte, Mostafa Sadeghi</i>	
Les modèles multimodaux peuvent-ils aider à l'interprétation de cartes ? Une étude exploratoire avec GPT-4o	48
<i>Edith Galy, Ahmed Moubtahij, Azur Handan, Marc Queudot</i>	
SIMI v3 : Une liste de cas patients similaires pour la télé expertise médicale	59
<i>Pierre Jourlin, Marc-Antoine Sulmon, David Bensoussan, Émilie Mercadal</i>	
SPARK : Exploiter les échanges techniques passés pour améliorer le support client	69
<i>Steve Bellart, Arnaud Deleruyelle</i>	