



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 l'atelier Avancement de l'AMR et de l'Analyse Sémantique 2025 (4AS)

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 conjointement par les équipes de recherche LORIA, Télécom-Paris et Orange (DATA&AI/AITT) cet atelier souhaite réunir les chercheuses et chercheurs qui s'intéressent à l'analyse sémantique et la modélisation sémantique. Nous souhaitons :

Faire le point sur les différentes approches d'exploitation de la sémantique du langage, ses différents formalismes (DRS, AMR, Yarn, ...) et leurs usages (modèles d'IA plus frugale, réponses à des problèmes où la sémantique a des avantages clairs par rapport aux LLMs, ...). Un focus sera fait sur les Abstract Meaning Representations (AMR). l'AMR, ses forces, ses faiblesses, ses évolutions potentielles et ses évolutions futures.

Identifier et positionner les différentes ressources utiles à l'analyse sémantique. Si le développement d'un Propbank en français n'apparaît pas comme une stratégie gagnante, porter le développement d'un alignement des mots du français vers Propbank, dans la ligné de VerbNet et Semlink serait une piste pertinente. La question des corpus annotés est aussi largement ouverte.

Echanger les différents points de vue sur l'intérêt de poursuivre cette recherche face aux capacités des très grands modèles de langue (LLM), la meilleure façon de structurer cette recherche (projets collaboratifs par exemple) et de communiquer auprès d'un vaste public.

L'atelier a sollicité des communications qui abordent un ou plusieurs des thèmes suivants :

- Interface syntaxe-sémantique ;
- Les ressources pour la sémantique
- Expansion ou couplage de ressources sémantiques avec des LLM ;
- Conception et annotation des représentations sémantiques ;
- Comparaison des framework de représentations sémantiques ;
- Génération automatique de textes à partir de représentations de sens ;
- Forces et faiblesses des représentations sémantiques existantes ;
- Utilisation des représentations sémantique dans des applications réelles ;
- Application des représentations sémantiques et multilinguismes ;
- multimodalité dans les représentations du sens ;
- La relation entre les représentations symboliques du sens et les représentations sémantiques distribuées ;
- Propriétés formelles des représentations de sens ;

Maxime Amblard

Bruno Guillaume

Maria Boritchev

Johannes Heinecke

Frédéric Herledan

Organisat.eur.rice.s de l'atelier A4S

Comités

- Maxime Amblard, Université de Lorraine
- Bruno Guillaume, Inria
- Maria Boritchev, Télécom-Paris
- Johannes Heinecke, Orange (DATA&AI/AITT)
- Frédéric Herledan, Orange (DATA&AI/AITT)

Table des matières

L'essentiel est invisible pour les représentations sémantiques	1
<i>Amandine Decker, Maxime Amblard</i>	
Prétraitement syntaxique pour enrichir le Bag of Words en Topic Modeling	9
<i>Connor MacLean, Denis Cavallucci</i>	
Ressources lexicales pour la sémantique : WordNet, BabelNet, PropBank, FrameNet, DBpedia et SUMO	17
<i>Ahana Chattopadhyay</i>	