



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 Science Participative pour les Données et Corpus Linguistiques 2025
(ParCol)

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

Organisé conjointement par Alice Millour (Université Paris 8) et Mathilde Hutin (Université de Lorraine, CNRS) avec le soutien du GdR LIFT-2, l'atelier "Science Participative pour les Données et Corpus Linguistiques" (ParCoL) vise à rassembler la communauté autour des questions liées à l'usage des sciences participatives pour répondre à des enjeux dans le cadre des sciences du langage. Les sujets d'intérêt incluent (sans y être limités) :

- le développement ou le fonctionnement d'outils de science participative (plateformes de récolte, jeux avec un but, etc.) dédiés aux langues,
- l'engagement des participantes-participants (motivations, nudges, ludification...),
- la science participative dans le contexte des langues peu dotées,
- la science participative au service de l'exploration de l'oralité et de la pluralité des usages,
- les enjeux théoriques des données et corpus participatifs,
- les défis techniques des plateformes de récolte de données ou métadonnées participatives,
- les questions éthiques soulevées par la création ou l'utilisation de corpus participatifs,
- les enjeux politiques ou législatifs des données et corpus participatifs.

L'atelier sera inauguré par une conférence invitée intitulée "Sciences participatives 2.0 : Un changement profond dans la manière de produire des données" présentée par Romain Julliard, professeur affilié à Sorbonne Université et au Muséum national d'histoire naturelle (MNHN), directeur de l'unité de service Mosaic ("Méthodes et Outils pour les sciences participatives") et directeur de la publication de SCIENCE ENSEMBLE, portail de recensement des projets de sciences participatives de l'Alliance Sorbonne Université. S'ensuivra une table ronde ouverte, axée autour de quatre projets de recherche mais pensée comme un espace de partage plus général afin de mettre en commun les expériences des participantes et participants avec les sciences participatives. L'après-midi sera consacrée à un datathon sur les plateformes Wikidata, Wikisource, la Dicothèque, et Lingua Libre.

Nous tenons à exprimer notre profonde reconnaissance envers toutes les personnes qui ont participé à faire vivre l'atelier : les autrices et auteurs des soumissions, les membres du comité de relecture, et le comité d'organisation de CORIA-TALN 2025. Nous remercions également très chaleureusement le GdR LIFT-2. Son soutien financier nous a permis d'inviter des intervenants de qualité et d'ainsi fédérer cette communauté naissante dans un cadre stimulant.

Mathilde Hutin
Alice Millour
Organisatrices de ParCoL

Comités

Comité de Programme

- Mathilde Hutin, Université de Lorraine, CNRS, ATILF
- Alice Millour, Université Paris 8 Vincennes-Saint-Denis, LIASD

Comité de Relecture

- Marc Allasonnière-Tang, Université Paris Cité, CNRS, MNHN
- Berthold Crysmann, Université Paris Cité, CNRS, LLF
- Mathilde Hutin, Université de Lorraine, CNRS, ATILF
- Alice Millour, Université Paris 8 Vincennes-Saint-Denis, LIASD
- Eleonor Chodroff, Université de Zürich, Department of Computational Linguistics
- Karën Fort, Université de Lorraine, CNRS, LORIA
- Loïc Grobol, Université Paris Nanterre, CNRS, MoDyCo
- Mélanie Jouitteau, Université Bordeaux-Montaigne, CNRS, IKER
- Romain Julliard, Sorbonne Université, MNHN, CNRS, CESCO
- Sylvain Loiseau, Université Paris 13, CNRS, LACITO

Table des matières

La science participative en pratique : comment réussir (et ne pas le faire) ?	1
<i>Laure Turcati, Alice Millour, Renaud Debailly, Karën Fort, Asma Steinhausser, Corentin Biets, Anne Dozières</i>	
La science participative et l'ANR DiLSi	3
<i>Pierre Magistry, Ilaine Wang</i>	
LiLA : Outil d'augmentation automatisée des données vocales participatives de Lingua Libre	6
<i>Mathilde Hutin, Marc Allasonnière-Tang, Lucas Prégaldiny, Lucas Lévêque</i>	
Lingua Libre à l'ère de l'automatisation : l'I.A. au service du crowdsourcing d'un corpus oral	11
<i>Camille Lavigne, Florian Cuny</i>	