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 Traitement de données langagières dynamiques par les outils et méthodes du TAL 2025 (DYN-TAL)

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

Préface

L'atelier DYN-TAL porte sur l'étude des données dynamiques, c'est-à-dire les données qui préservent les indices de production et qui, de ce fait, présentent un caractère évolutif et incrémental. Ces données sont hétérogènes et multi dimensionnelles. Elles présentent un vrai défi pour les outils du TAL : comment les traiter tout en conservant les informations dynamiques, avec tout ce qu'elles ont d'instable et de contingent ?

L'objectif de l'atelier est de contribuer à la mise en commun de pratiques de développement ou d'adaptation d'outils et de méthodes pour le traitement des données dynamiques.

- les 35ème Journées d'Études sur la Parole (JEP),
- la 31ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN),
- la 26ème Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL).

L'écriture enregistrée en temps réel à l'aide d'outils d'enregistrement des frappes (keyloggers) ou de tablettes tactiles constitue un exemple intéressant de ce type de données. Celles-ci contiennent des informations de nature variée sur les caractères tapés, sur le décours temporel du processus et les événements d'écriture tels que les révisions. Par ailleurs, ces données comportent des phénomènes regroupés ici sous le terme de contingence, tels que les éléments qui ne sont pas récurrents et qui sont imprévisibles : faux départs, hésitations, retour au début du paragraphe, suppressions, etc. Le caractère dynamique de ces données peut être attesté à travers les différentes étapes de production dont la dernière ne comporte aucune trace de révision et représente un document « propre » alors que les étapes intermédiaires contiennent les indices apparaissant et disparaissant au fur et au mesure de l'écriture. Comment intégrer de manière cohérente l'ensemble de ces données sans perdre d'informations ni au niveau langagier ni au niveau chronologique ? Cela soulève de nombreuses questions méthodologiques et pratiques liées à leur traitement automatique. Sont concernées des tâches variées comme l'annotation (voir Miletic et al., 2022), la segmentation (voir Cislaru et al., 2023), l'analyse syntaxique (Ulasik et al., 2025), etc.

Les recherches en TAL sur les données dynamiques concernent d'abord le traitement de la parole, un exemple emblématique de ces données, et ses nombreuses tâches dont la Reconnaissance Automatique de la Parole (RAP) (Mariani 2002 ; Gelly, 2017 ; Haton, 2018 ; Elloumi, 2019). Ces recherches témoignent des avancées significatives dans le domaine, en intégrant des approches de pointe, notamment les réseaux de neurones, et en prenant en compte les indices de production de la parole afin d'optimiser les performances des systèmes. Si des travaux se sont intéressés à l'adaptation dynamique des outils TAL (Upadhyaya, 2024) ou au traitement en temps réel de données du Web (Archi Saloot & Nghia Pham, 2021), les applications des outils TAL aux données dynamiques telles que l'écriture en temps réel n'en sont qu'à leurs débuts (Eshkol-Taravella et al., 2024 ; Qian, 2024).

Nous aurons le plaisir d'accueillir, en tant que conférencier invité, M. Mickael Zock, Directeur de recherche au Laboratoire d'Informatique et Systèmes (LIS – CNRS). Il présentera une communication intitulée intitulée "Creation of a Cognitively Motivated Ecosystem to Assist Authors in Text Composition".

Six communications seront présentées, portant sur la préparation des données dynamiques en vue de leur traitement automatique, la segmentation de ces données, ainsi que la prédiction automatique de certains phénomènes de contingence.

Iris Eshkol-Taravella
Georgeta Cislaru

Comités

Comité de Programme

— Georgeta Cislaru, Université Sorbonne Nouvelle, Clesthia
— Iris Eshkol-Taravella, Université Paris Nanterre, MoDyCo

Comité de Relecture

— Nicolas Ballier, Université Paris Cité, ALTAE
— Guénaël Cabanes, Université de Lorraine, LORIA
— Géraldine Damnati, Orange Labs
— Claire Doquet, Université de Bordeaux, LAB-E3D
— Gaëtanelle Gilquin, Université Catholique de Louvain
— Nistor Grozavu, CYU, ETIS
— Cerstin Mahlow, ZHAW, Zürich
— Alexandra Miletič, MoDyCo, CNRS
— Claude Ponton, Université Grenoble Alpes, LIDILEM
— Nicoleta Rogovschi, Université Paris Cité, LIPADE
— Didier Schwab, Université Grenoble Alpes, LIG-GETALP
— Nadi Tomeh, Université Paris Nord, LIPN
— Biagio Ursi, Université d'Orléans, LLL

Table des matières