

## Vers la numérisation et l'exploration automatisée du *Dictionnaire Universel François et Latin de Trévoux (1704-1771)*

Journée d'étude organisée à l'ENS de Lyon le 4 juillet 2024  
dans le cadre de l'Appel à Projet Pluridisciplinaire Lyon 2 (2024-2026)

**9h00-17h00, ENS Lyon, salle D4 179**

(Bâtiment Recherche, 1er étage, bout du couloir ICAR).

La journée sera aussi accessible par visio-conférence.

Inscription nécessaire auprès de [denis.vigier@ens-lyon.fr](mailto:denis.vigier@ens-lyon.fr) (contact)

**8h30:** Accueil, café

### MATIN

**D. Vigier** (U. Lyon 2) **9h00-9h05.** *Mot d'introduction.*

**I. Turcan** (U. Angers) **9h05 - 9h30.** *Pour une grammaire des marques de domaines : l'exemple de la marine.*

**J.- F. Bianco** (U. Angers) **9h30-9h55.** *Recherches sur la marque « terme de relation » dans le Trévoux (éd. 1743-1752). Un domaine transversal.*

**G. Kassiteridis** (U. Avignon) & **Ch. Wionet** (U. Avignon) **9h55-10h20.** *L'emploi de TEI Lex 0 au Dictionnaire universel de Trévoux (1721)*

Pause 10h20-10h35

- **I. Galleron** (U. Paris Sorbonne Nouvelle) & **G. Williams** (U. Bretagne Sud) **10h35 - 11h.** *Annoter le Dictionnaire universel: les leçons d'un projet.*

- **C. Rey** (U. Paris Cergy) **11h - 11h25.** *Le Grand Vocabulaire Français (1767-1774) de Charles-Joseph Panckoucke : un anti-Trévoux notoire.*

- **G. Roe** (U. Paris Sorbonne) **11h 25 - 11h 50.** *La présence du Chambers dans l'Encyclopédie: expériences de traduction automatique.*

**Discussion: 11h50 – 12h 30**

**Déjeuner : 12h 30-14h 00**

- **H. Bohbot** (U. Montpellier 3) **14h00 -14h25.** *Numérisation d'éditions successives : retour d'expérience avec le projet Nénufar et le Petit Larousse.*

- **J. Chazalon** (EPITA Paris) **14h25 - 14h50.** *Un exemple d'extraction et de représentation des données : les annuaires du commerce et cadastre napoléonien*

- **V. Eglin** (INSA Lyon) **14h50 – 15h 15.** *Apport de la multimodalité texte et image pour l'analyse des documents anciens imprimés et manuscrits.*

- **T. Joliveau** (Université St. Étienne) et **A. Taroni** (CNRS). **15h15- 15h 40.** *Un projet cartographique pour l'Encyclopédie. Localiser et géovisualiser les articles géographiques de l'Encyclopédie de Diderot et d'Alembert.*

- **L. Moncla** (INSA Lyon), **J. Velcin** (U. Lyon 2), **D. Vigier** (U. Lyon 2) **15h40-16h05.** *Pistes pour une classification automatique des domaines de connaissance dans le Trévoux*

Discussion finale. **16h05-17h00**

## Titres et résumés (par ordre de passage)

### I. TURCAN (Université Angers)

#### Pour une grammaire des marques de domaines : l'exemple de la marine.

Dans la continuité de nos premières réflexions sur le développement des marques de domaines dans les dictionnaires généraux de langue aux XVII<sup>e</sup> et XVIII<sup>e</sup> siècles et sur leur importance métalexigraphique, nous proposons une réflexion sur les enjeux afférents à ces marques et sur leurs modalités d'exploitation automatisée en prenant l'exemple du vocabulaire de la marine dans le *DUFLT*.

### J.-F. BIANCO (Université Angers)

#### Recherches sur la marque *terme de relation* dans le *Trévoux* (éd. 1743-1752). Un domaine transversal ?

Le statut de la marque *terme de relation* pose question. Les *Relations* des voyageurs, genre florissant au dix-septième et au dix-huitième siècle, offrent une source aux lexicographes et ouvrent un champ de connaissances aux lecteurs. C'est une sorte de *domaine* qui croise plusieurs domaines. C'est un nouveau monde lexicologique qui se greffe sur l'ancien monde. Nous proposons, dans la perspective d'un traitement automatisé, quelques recherches sur l'usage et la grammaire de cette marque dans le *Trévoux*, en comparaison avec d'autres ouvrages, surtout le *Dictionnaire* de Furetière, évidemment, et l'*Encyclopédie*.

### G. KASSITERIDIS (Université d'Avignon) et C. WIONET (Université d'Avignon)

#### L'emploi de TEI Lex 0 au *Dictionnaire universel de Trévoux* (1721)

Il s'agit d'une présentation du modèle de balisage TEI Lex 0 et de son application à un échantillon de la deuxième édition du *Dictionnaire universel de Trévoux*. Les avantages du modèle seront discutés, notamment, la possibilité d'interopérabilité qu'il permet entre différents dictionnaires informatisés, ainsi que ses limites par rapport au texte du *Trévoux*. Certaines difficultés liées à la transcription du texte, comme les ligatures, seront également abordées.

### I. GALLERON (Université Paris Sorbonne Nouvelle) et G. WILLIAMS (Université Bretagne Sud)

#### Annoter le *Dictionnaire universel* : les leçons d'un projet

Le projet BASNUM a mené à bien l'océrisation, puis l'annotation massive du *Dictionnaire universel de Basnage de Beauval*. Toutefois, entre le point de départ du projet et le point d'arrivée, une distance palpable se constate, à la fois en ce qui concerne la finesse des informations annotées et leur précision. Notre présentation se concentrera sur trois aspects:

- 1° Ce que l'annotation initialement prévue avait de trop ambitieux.
- 2° Ce que GROBID a pu faire.
- 3° Comment il faudrait s'y prendre, si c'était à refaire.

### C. REY (Université Paris Cergy)

#### Le *Grand Vocabulaire Français* (1767-1774) de Charles-Joseph Panckoucke : un anti-Trévoux notoire.

En raison de sa filiation étroite avec l'*Encyclopédie* de Diderot et d'Alembert, le *Grand Vocabulaire Français* (1767-1774) de Charles-Joseph Panckoucke s'est d'emblée inscrit dans la position d'un adversaire avéré et opiniâtre du *Dictionnaire Universel français et latin* de Trévoux. Nous nous proposons de revenir ici sur l'opposition entre ces deux ouvrages, en faisant état des nombreux griefs et attaques formulés au fil de leurs pages, nous plongeant ainsi dans une querelle lexicographique stimulante.

**G. ROE** (Université Paris Sorbonne)

### La présence du Chambers dans l'*Encyclopédie*: expériences de traduction automatique

L'*Encyclopédie*, comme on le sait, a été lancée comme un simple projet de traduction du *Cyclopaedia* anglais d'Ephraïm Chambers. Vite abandonné, ce projet laisse néanmoins ses traces dans les articles encyclopédiques ; traces parfois difficiles à identifier d'une manière systématique. Cette communication explore une expérience avec des méthodes de traduction automatique et d'alignement de séquences pour comparer le texte de l'*Encyclopédie* à la *Cyclopaedia*. En utilisant des techniques avancées de traitement du langage naturel pour analyser les similitudes et les différences entre les textes, nous mettons en lumière les processus de traduction et d'adaptation au XVIII<sup>e</sup> siècle.

**H. BOHBOT** (Université Montpellier 3)

### Numérisation d'éditions successives : retour d'expérience avec le projet *Nénufar* et le *Petit Larousse*

Le projet *Nénufar* vise à rendre accessible toutes les éditions du *Petit Larousse* dans le domaine public (actuellement, 47 éditions de 1906 à 1953). Nous aborderons la stratégie mise en place pour n'avoir à numériser qu'un certain d'éditions clés, le développement d'interface de comparaisons et de corrections ainsi que l'automatisation autant que possible du balisage en XML-TEI et les difficultés rencontrées.

**J. CHAZALON** (EPITA Paris)

### Annuaire du commerce et cadastre napoléonien : extraction et représentation des données.

Cette présentation rapide résumera les principaux enseignements tirés du projet ANR SODUCO concernant l'extraction massive d'entrées dans des annuaires du commerce parisiens du 19<sup>e</sup> siècle et leur structuration, puis identifiera les limites de cette approche pour l'analyse d'états de sections (tables) du cadastre napoléonien, et présentera finalement les derniers outils utilisés pour ce nouveau projet.

**V. ÉGLIN** (INSA Lyon)

### Apport de la multimodalité texte et image pour l'analyse des documents anciens imprimés et manuscrits

Dans cet exposé seront présentés des résultats d'analyse de documents anciens, imprimés, manuscrits et cartographiques qui exploitent le texte et l'image de manière conjointe pour mieux détecter et reconnaître les informations présentes. Les techniques abordées permettront de montrer l'évolution des méthodes allant de l'OCRisation des images à leur compréhension par des transformers et LLMs. On ciblera également les enjeux liés à l'analyse de corpus rares, pour lesquelles il est nécessaire de compenser les données difficiles à annoter."

**T. JOLIVEAU** (Université Saint Étienne) et **A. TARONI** (CNRS)

Sur les 74 000 articles contenus dans l'*Encyclopédie* de Diderot et d'Alembert (et Jaucourt), plus de 15 000 décrivent des entités de lieu : localités, rivières, montagnes ... Or l'*Encyclopédie* ne comprend ni index des lieux ni carte. Un des objectifs du projet Géode a été de localiser et cartographier ces entités en testant différentes méthodes : extraction des coordonnées anciennes du texte, recherche des coordonnées modernes, rattachement à des entités plus vastes ...

**L. MONCLA** (INSA Lyon), **J. VELCIN** (Université St. Étienne), **D. VIGIER** (Université Lyon 2)

### Pistes pour une classification automatique des domaines de connaissances dans le Trévoux

À partir d'une réflexion critique autour de la question des « domaines » dans le Trévoux (tensions notamment entre évolutions au sein de la série et possibilités d'une catégorisation), nous esquisserons quelques pistes possibles pour une classification automatique des domaines de connaissances associables aux articles du Trévoux.