



## L'HTR des langues peu dotées dans les programmes de recherche et dans les établissements de conservation français

Journée d'étude à la BnF en partenariat avec le consortium HumaNum [DISTAM](#)

14 février 2024, Bibliothèque nationale de France,  
Salle des conférences, 5 rue Vivienne, 75002 Paris

Depuis quelques années, différents projets d'HTR (*handwritten character recognition*) de manuscrits en graphies non-latines (arabe, arménienne, chinoise, hébraïque, etc.), conservés dans des bibliothèques françaises, ont été menés à bien, notamment dans le cadre du programme CollEx Persée, de projets ANR, ou des travaux du DataLab de la BnF.

La plupart parviennent à des taux de reconnaissance supérieurs à 90%, mais avec une marge de progression qui semble limitée en l'état actuel des technologies employées, et un défi qui est celui de la correction des textes extraits.

La journée d'étude se propose de faire un point sur ces avancées et ces défis avec les porteurs actuels de projets de reconnaissance automatique d'écriture en langues peu dotées en matière d'HTR. Il s'agira de s'interroger sur les capacités de progression des outils utilisés, sur ce que recouvrent exactement les taux de reconnaissance obtenus, sur les perspectives que cette mise à disposition massive d'extractions de textes manuscrits ouvre pour la recherche en termes de lecture distante et de traitement automatique de la langue ainsi que sur la normalisation des modèles et la publication des modèles et des données d'entraînement. On examinera enfin la place de l'HTR dans les programmes des établissements de conservation français, ainsi que l'intégration des extractions obtenues dans les bibliothèques numériques.

### Programme

9h00-9h15 *Ouverture*

- Tiphaine Vacqué (BnF)
- Mercedes Volait (DISTAM)

9h15-10h15 *L'HTR des textes en chinois classique*

- Marie Bizais-Lilig (université de Strasbourg) : [Chi-Know-Po Corpus](#) - La poésie dans l'économie des savoirs lettrés en Chine médiévale
- Frédéric Constant (université de Nice) et Colin Brisson (EPHE - Centre de recherche sur les civilisations de l'Asie orientale) : [COREL](#) – Reconstitution de la législation de la Chine impériale tardive à partir de différents corpus

10h15-10h45 *L'HTR des textes en hébreu médiéval*

- Daniel Stoekl ben Ezra (EPHE et Equipex Biblissima +) : [MiDRASH](#), Migrations of Textual and Scribal Traditions via Large-Scale Computational Analysis of Medieval Manuscripts in Hebrew Script

10h45-11h00 *Pause-café*

11h00-12h30 *L'HTR des textes en arabe maghrébin et oriental*

- Noémie Lucas (université d'Edimbourg) : [Rasam](#) - A Dataset for the Recognition and Analysis of Scripts in Arabic Maghrebi
- Antoine Perrier (CNRS-Iremam et Centre Jacques Berque) : [Tarima](#) - Écrire l'histoire au Maghreb aux périodes modernes et contemporaines
- Eric Vallet (université de Strasbourg) : [HUNAI](#) - Humanités numériques pour l'étude de l'arabe islamique

12h30-13h30 *Déjeuner*

13h30-14h30 *Table ronde 1 : Capacités de progression des outils utilisés, signification et implication des taux de reconnaissance obtenus*

Modératrice : Mercedes Volait (CNRS-INHA, InVisu)

Participants.es :

- Peter Stokes (EPHE et Equipex Biblissima +)
- Chahan Vidal-Gorène ([Calfa](#) et École des Chartes)
- Anais Wion (Institut des Mondes Africains, CNRS)

14h30-15h30 *Table ronde 2 : Normalisation et publication des modèles, publication des données d'entraînement*

Modératrice : Marie Bizais-Lilig (université de Strasbourg)

Participants.es :

- Alix Chagué et Thibault Clérice (équipe ALMAnaCH, Inria, projet [HTR United](#))
- Anne-Valérie Schweyer (CNRS-EHESS, Centre Asie du Sud -Est) et Jean-Christophe Burie (Laboratoire L3i, université de La Rochelle) ([ANR ChamDoc](#))
- Daniel Stoekl ben Ezra (EPHE et Equipex Biblissima +)

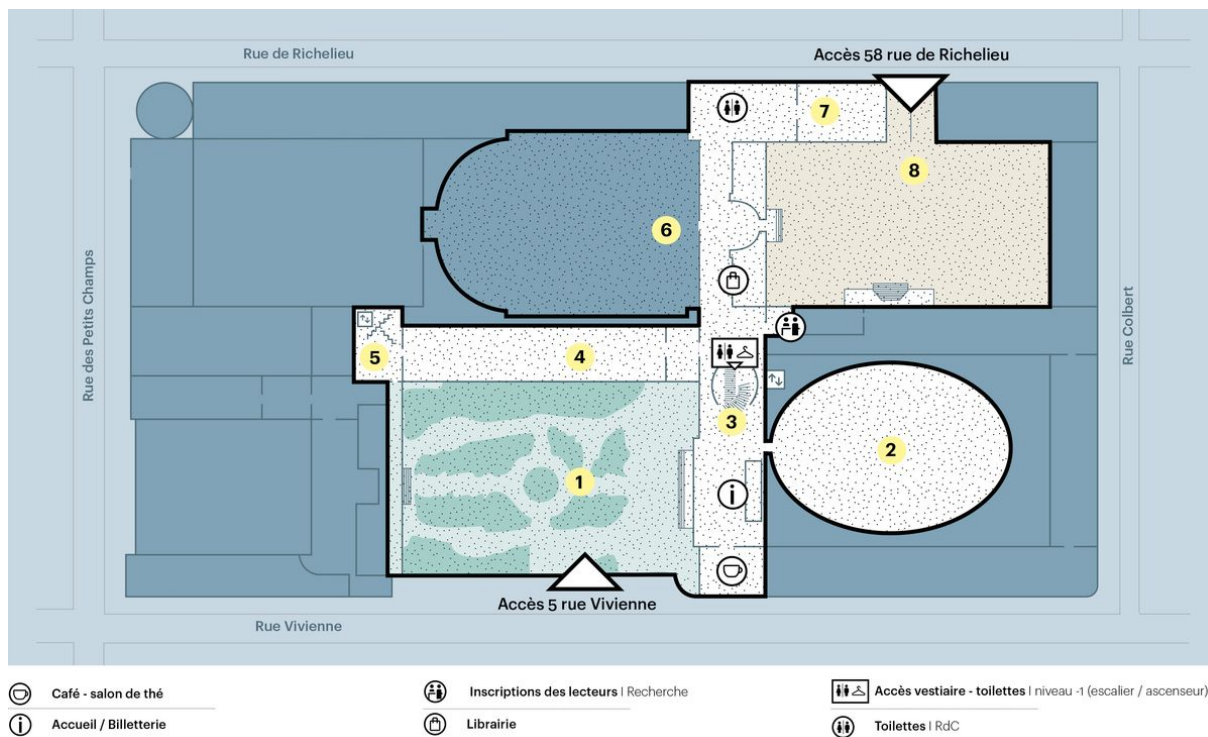
15h30-15h45 *Pause-café*

15h45-16h45 *L'avenir de la collaboration entre les chercheurs et les établissements de conservation autour de l'HRT*

Echanges de clôture entre les intervenants et le public

## Accès

**Plan du RDC : Entrée par le 5 rue Vivienne, puis accès par le hall Roux-Spitz (n°5 du plan ci-dessous)**



**Plan du 1<sup>er</sup> étage : la Salle des Conférences est au numéro 11**

