



HTR en pratique : Transkribus Lite

Ljudmila PETKOVIC

Introduction aux humanités numériques (L1HN001)
Mineure « Humanités numériques », licence Lettres
Paris, le 9 novembre 2023, année 2023-2024

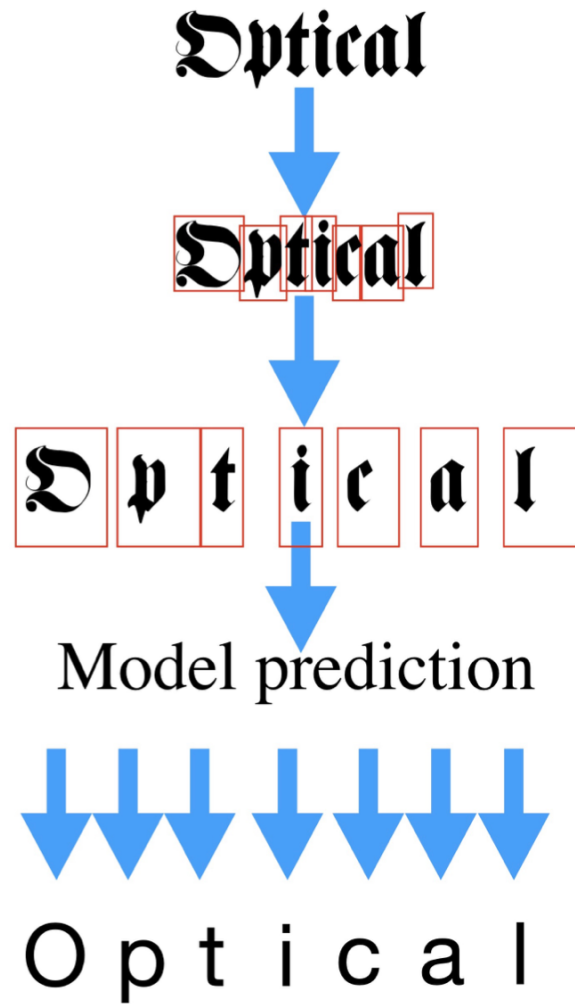
OCR

- reconnaissance optique de caractères (angl. *optical character recognition*)

Processus qui consiste à convertir un ensemble de signes graphiques, le plus souvent alphanumériques (mais aussi les ponctuations, espacements...), encodés sous la forme d'une image, en mode texte. L'OCR désigne à la fois un **processus** (d'OCR) et un **logiciel** (d'OCR).

(Camps & Perreaux, 2021)

OCR



Mécanisme de l'OCR ([Gabay et SciCoS, s.d., pp. 4-5](#)).

HTR

- reconnaissance de l'écriture manuscrite (angl. *handwritten text recognition*)

Comment peut-on segmenter les caractères en écritures cursives, étant donné que les mêmes sont écrits attachés de manière fluide ? Dans ce cas, la segmentation au niveau des caractères s'avère impossible.

(adapté selon Gabay et SciCoS, s.d., p. 5)



Registre du Comité d'administration du Théâtre français de S. M. l'Empereur et Roi, par Nicolas Bernard.

Transkribus

<https://readcoop.eu/transkribus/>

- outil pour la reconnaissance automatique de texte (à l'aide de l'intelligence artificielle – IA) et l'édition numérique des documents d'archives
- potentiel tant pour la recherche que pour l'archivage

Transkribus Lite	Transkribus Expert Client
interface intuitive (navigateur web)	téléchargement requis
destiné aux non-informaticiens	beaucoup plus de fonctionnalités
interface en différentes langues (dont français)	interface en anglais

Envergure du logiciel

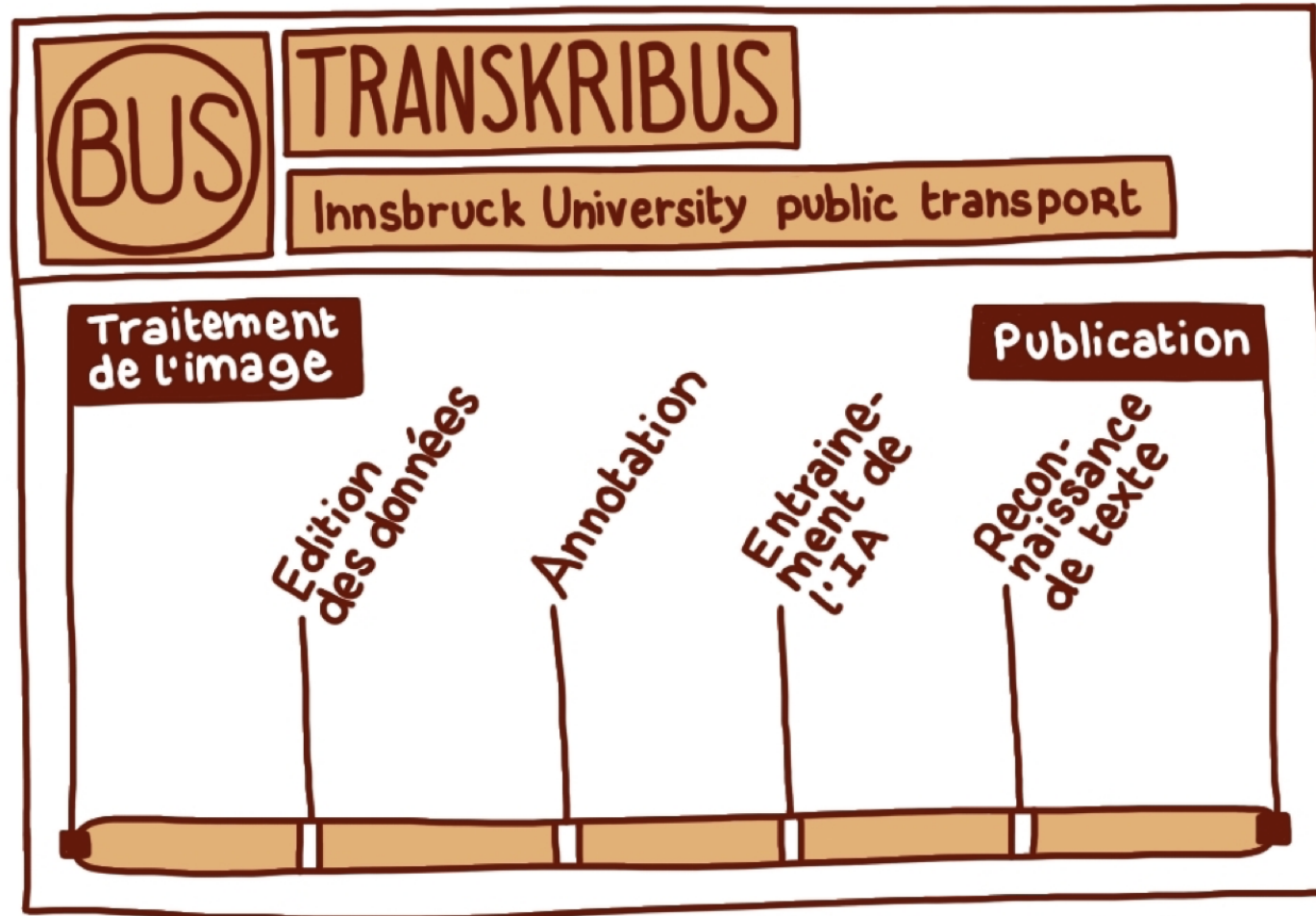


Illustration reprise de [Gautier et al., 2022, p. 2.](#)

Workflow

0. créer un compte sur READ-COOP (**pré-requis**)
1. importer des documents
2. effectuer des analyses de mise en page
3. transcrire automatiquement des documents à l'aide de modèles existants
4. entraîner un modèle spécifique pour vos documents (**vérité terrain nécessaire**)
5. exporter des transcriptions dans différents formats
6. rechercher et compter les mots dans les transcriptions

Capsules vidéo



0. Créer un compte sur READ-COOP [lien YouTube](#)

- Inscription sur le site web de READ-COOP <https://readcoop.eu/>

Télécharger un document de Gallica  **Lien YouTube**

1. Importer des documents [Lien YouTube](#)

2-3. Analyse de mise en page + transcription automatique

lien YouTube

5. Export des transcriptions [Lien YouTube](#)

6. Recherche et comptage des mots [Lien YouTube](#)

Références

- **Camps, J.-B. & Perreaux, N.** (2021). Reconnaissance optique des caractères et des écritures manuscrites · Projet E-NDP [*diapositives*]
https://outils.lamop.fr/lamop/mp3/E-Ndp/JBC-NP_e-NDP_OCR-et-HTR.pdf
- **Gabay et SciCoS.** (s.d.). FoNDUE (FOrmes Numerisees et Detection Unifiee des Ecritures) · An Handwriting Text Recognition infrastructure for Geneva
https://datascience.unige.ch/application/files/2316/3248/7621/3.3._Simon_Gabay__Jean-Luc_Falcone.pdf
- **Galleron, I.** (2021). « Introduction aux humanités numériques (L1HN001) [*diapositives en interne*].