



© Julien Fournigault

Atelier Digit_Hum 2019

Les humanités numériques en langues

JEUDI 17 OCTOBRE 2019

ÉCOLE NORMALE SUPÉRIEURE | SALLE DUSSANE





Atelier Digit_Hum 2019

Les humanités numériques en langues

Cette journée s'inscrit dans le cycle des Ateliers Digit_Hum qui explorent depuis 5 ans, selon une thématique choisie, les nouvelles pratiques et les nouveaux outils développés en humanités numériques.

Cette année, l'Atelier souhaite rendre compte de la multiplication des plateformes de transcription et de traduction automatiques et collaboratives et s'interroger sur ce que peuvent apporter les humanités numériques pour transformer les textes et permettre leur transmission et leur analyse.

A l'issue de cette journée, une séance de mise en pratique des outils présentés (environnement collaboratif de traduction TRADUXIO - plateforme de transcription collaborative TACT - module de transcription TRANSCRIPT utilisé par la plateforme EMAN - logiciel de transcription automatique TRANSKRIBUS) sera proposée pendant une heure, de 16h30 à 17h30.

Journée organisée par :

Marie-Laure Massot (CAPHÉS, UMS 3610, CNRS-ENS)

Julien Cavero (LGP, UMR 8591, CNRS-Paris 1-UPEC)

 #Digit_Hum2019

 digit_hum@ens.fr

9:30 | **Introduction**

Isabelle Kalinowski (directrice de l'EUR Translitterae)

Modération : Thierry Poibeau (UMR 8094 Lattice – ENS Paris)

9:45-10:15 | **De l'usage de la pelle et de la pioche en traduction littéraire**

Olivier Mannoni (Directeur de l'École de Traduction Littéraire – CNL/Asford)

Au mois de juillet encore, j'ai eu une nouvelle occasion de tester avec des élèves le plus récent et, selon ses promoteurs, le plus perfectionné des logiciels de traduction automatique utilisant des « systèmes neuronaux ». Le résultat était consternant : à peine compréhensible, criblé de faux sens et de contresens d'élève de sixième (« Il appela le taxi » pour « Le taxi se mit à crier ») et surtout d'une platitude extrême, comme si l'on avait passé à la va-vite un rabot à la surface du texte sans jamais chercher à voir ce que dissimulaient ses veinures.



Or la traduction littéraire est justement l'art de creuser : creuser derrière l'apparence des mots, derrière les expressions figées, derrière les artifices du style et même, au moins à titre conservatoire, derrière les intentions de l'auteur. Contexte, sous-texte, non-dits

constituent l'essentiel de ce qu'il faut traduire et de ce que, par définition, aucune machine ne peut appréhender. J'essaierai de montrer au cours de ce bref exposé pourquoi, au moins en matière littéraire, la traduction automatique est un oxymore.

10:15-10:45 | **La traduction outillée au cœur de l'interaction Homme-Machine**

Hanna Martikainen (CLILLAC-ARP – Université de Paris)

La traduction professionnelle pragmatique telle qu'elle est pratiquée aujourd'hui s'inscrit pleinement dans le champ des humanités numériques notamment à travers les corpus qui sont à la base de tous les outils d'aide à la traduction (ex. mémoires de traduction, moteurs de traduction automatique (TA) statistique et neuronale).

Dans cette interaction constante entre l'humain et la machine, les frontières entre les deux se brouillent de plus en plus : citons par exemple les usages hybrides où les segments issus d'une mémoire sont complétés par la TA, le recyclage des solutions de TA via les mémoires de traduction, ou encore l'exploitation de corpus traduits automatiquement pour l'entraînement des moteurs de TA. L'impact de l'intervention de la machine se manifeste dans ces textes hybrides par

exemple à travers des effets de calque et d'uniformisation, et le traducteur humain conserve toute sa place dans cette interaction pour garantir la cohésion textuelle et la cohérence discursive des textes traduits.



10:45-11:00 | *Pause*

Alors que le miracle de la traduction automatique que l'on appelle « neuronale » nous est de plus en plus souvent conté, on ne se penche que trop rarement sur ses usages réels et sur les perceptions qu'en ont les utilisateurs. Qui sont aujourd'hui les utilisateurs des systèmes de traduction automatique, et quelles sont leurs pratiques ? Que savent-ils de la traduction automatique et comment la perçoivent-ils ? A partir de données disponibles dans la littérature ou issues de sondages récents, j'essayerai de définir les usages actuels de la traduction automatique, afin de montrer les risques encourus par des utilisateurs non avertis, et les avantages que peuvent présenter certains usages bien encadrés.



11:30-12:00 | **TraduXio : un environnement numérique pour la traduction collaborative et multilingue de textes culturels**
Philippe Lacour (Département de Philosophie de l'Université de Brasilia)

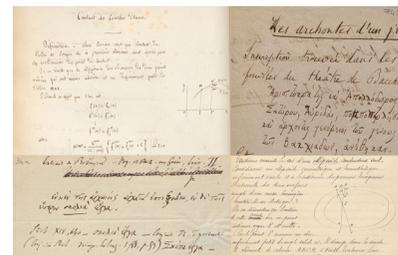
TraduXio est un environnement numérique gratuit, open source, basé Web, collaboratif, destiné à la traduction multilingue assistée par ordinateur. Il utilise une technologie innovante, adaptée aux défis des textes culturels (non commerciaux, non répétitifs). Il promeut la création de biens communs, suivant une logique d'enrichissement progressif de la base textuelle. Son originalité est triple. Tout d'abord, la traduction n'y consiste pas en un transfert entre deux langues



(source/cible), mais en un passage d'un texte singulier vers d'autres textes singuliers, dans plusieurs langues. Ensuite, la concordance multilingue permet de comparer les différentes versions d'un même texte, stimulant la créativité des traducteurs : des filtres permettent de rechercher des suggestions de façon pertinente, en fonction de catégories introduites par les utilisateurs pour exploiter les corpus de façon intelligente. Enfin, les utilisateurs peuvent collaborer en ligne, de façon synchrone ou non, dans un environnement intuitif et pratique (avec des fonctionnalités « sociales » du type : chat, forum, glossaire, etc.).

13:30-14:00 | **Transcript dans la chaîne de production de la plateforme EMAN**
Richard Walter (UMR 8132 ITEM, – ENS Paris)

Le plugin *Transcript* pour Omeka classic est un outil de transcription et d'annotation utilisant le balisage TEI et mis à disposition des projets de la plate-forme EMAN. Le module propose à un utilisateur non expert de saisir et d'annoter une transcription, en



vis-à-vis de l'image du texte, dans un bloc accompagné d'une barre d'outils pour l'encodage de la transcription avec des balises XML/TEI. Ces balises et leurs règles d'utilisation sont gérées en fonction du projet et respectent les préconisations du schéma TEI, notamment en ce qui concerne l'imbrication des balises. Pour la visualisation de la transcription, nous utilisons pour l'instant l'outil TEI Boilerplate mais la personnalisation du schéma TEI par chaque projet fragilise l'affichage et nous travaillons à une nouvelle solution de visualisation. Notre intervention présentera un bilan de l'utilisation de la version actuelle du module par plusieurs projets EMAN, en particulier les difficultés rencontrées par la nécessaire adaptation entre les besoins scientifiques de transcription et les nécessités de respecter en même temps le standard TEI et une saisie rapide et efficace des données.

15:45-16:30 | **TACT, une plateforme collaborative multi-projets pour la transcription de corpus textuels**
Célia Marion, Elisabeth Greslou et Anne Garcia-Fernandez (ELAN, Litt&Arts – UGA/CNRS)

Transcrire est une étape clef dans le processus de numérisation de sources, indispensable pour des analyses, réutilisations et autres exploitations ultérieures. La

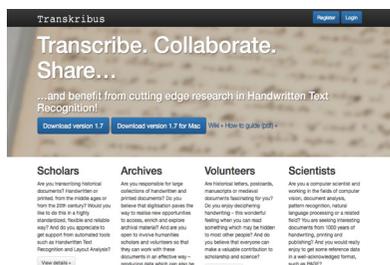


plateforme TACT permet aux projets de créer leur propre espace de transcription en l'adaptant à leurs besoins : modèle de transcription, structuration du corpus, public de transcrip-teurs, visualisation des transcriptions, flux de validation, etc. TACT a été conçue pour favoriser l'autonomie et la gestion des projets, de la création à l'export des données.

12:00-12:30 | *Echange avec la salle : Les nouveaux enjeux de la traduction*

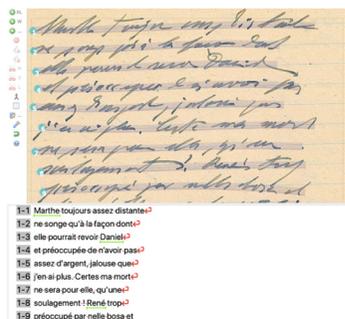
12h30-13h30 | *Déjeuner / Buffet*

14:30-15:00 | **Transkribus : Transcribe. Collaborate. Share...**
Günter Mühlberger and Johanna Walcher (Read - Transkribus)



Transkribus is the READ project's comprehensive platform for the automated recognition, transcription and searching of historical documents. The main objective of *Transkribus* is to support users who are engaged in the transcription of printed or handwritten documents, namely humanities scholars, archives, volunteers and computer scientists.

15:15-15:45 | **Transcription assistée par reconnaissance optique – Transkribus : l'expérience du journal intime d'Eugène Wilhelm (1885-1951)**
Régis Schlagdenhauffen (EHESS-IRIS)



Cette contribution propose de restituer une « expérience utilisateur » du logiciel *Transkribus*. Elle s'inscrit en lien avec le projet de transcription du journal intime du juriste Eugène Wilhelm (1866-1951). Ce journal comporte deux défis principaux : le premier est lié à la durée de sa rédaction, 66 années, ce qui engendre de grande variation dans la forme de l'écriture, car cette dernière devient de moins en moins lisible le temps passant. Le second défi est lié à l'emploi de deux alphabets : romain pour tout ce qui relève du quotidien et grec pour le for privé.

L'expérience dont je souhaite rendre compte s'articule autour de deux aspects. Dans un premier temps, après avoir présenté brièvement les avancées du projet, je synthétiserai les principaux obstacles rencontrés et les solutions apportées pour y remédier : transcription manuelle, premiers errements, difficultés de créer un modèle, création d'un dictionnaire. Dans un second temps, je reviendrai sur l'expérience de transcription collaborative menée avec des étudiant·e·s de l'EHESS. En ouverture, je synthétiserai les enjeux en cours et à venir afin d'améliorer le modèle et donc l'automatisation de la transcription.

15:45-16:15 | *Echange avec la salle : Les nouveaux enjeux de la transcription*

16h30-17h30 | *Possibilités de tester TraduXio, TACT, Transcript et Transkribus (sur inscription à digit_hum@ens.fr)*



DIGIT_HUM

Humanités numériques
entre savoirs et savoir-faire

Retrouvez toute l'actualité de l'initiative Digit_Hum sur le site

<https://digithum.huma-num.fr/>

Vidéos des Ateliers Digit_Hum

Enquête sur les activités en humanités numériques à l'ENS

Portraits d'acteurs en humanités numériques

Portfolio par l'artiste Saint-Oma

Digit_Hum est une initiative visant à échanger sur les différents usages qui sont faits des humanités numériques et à structurer ces nouvelles pratiques à l'échelle de l'École normale supérieure et de l'Université Paris Sciences & Lettres.



Cette initiative est portée par le CAPHÉS (Centre d'Archives en Philosophie, Histoire et Édition des Sciences) et le pôle Humanités Numériques de l'EUR Translitteræ.

