

APPEL A PROJETS - AMORÇAGE 2021

FICHE BILAN ACTION

Adresse de dépôt : katel.lochet@univ-nantes.fr

Annexes à joindre :

- **Programme** des manifestations scientifiques organisées dans le cadre du projet
- **Bilan financier** (budget détaillant l'utilisation des financements et co-financements)

ACRONYME : TALL

Titre du projet : Transcriptions d'autoportraits de lecteurs en Pays de la Loire

1 - Bilan des actions réalisées

Description des principales actions (objectifs, acteurs impliqués...). Pour les manifestations scientifiques organisées, préciser le programme, les dates, le(s) lieu(x) et (si possible) le, nombre de participants. (1,5 pages max.)

Le programme de travail qui avait été proposé dans la réponse à l'appel à projet a été globalement respecté.

La phase de collecte des données s'est étalée de mars à septembre 2021 et a nécessité une dizaine de déplacements aux Archives nationales à Pierrefitte-sur-Seine. Elle a consisté en la prise de clichés HD de toutes les lettres de candidature du corpus (environ 300).

La phase de préparation des fichiers a commencé parallèlement à la collecte. Elle a consisté comme prévu en une OCRisation pour les lettres tapuscrites qui a donné des résultats satisfaisants. Toutes les lettres manuscrites (environ 130) ont été lues par des étudiants dans le Userlab UDP. Cela a représenté environ 50 heures de travail. Elles ont été ensuite transcrites sur la plateforme de traitement automatique de la parole développée par le LIUM. L'ensemble des transcriptions a enfin été corrigé manuellement. Cette étape s'est achevée en novembre 2021. Les corrections ont été directement intégrées à la plateforme LIUM et ont donc contribué à l'amélioration des performances des algorithmes. Les enregistrements sonores réalisés dans le UserLab UDP ont également permis d'accélérer le processus de calibration des outils de captation déployés sur place. Dans le même temps, un protocole strict d'anonymisation a été implémentée et toutes les lettres du corpus, conformément aux demandes émises par les Archives nationales et Radio-France. Les transcriptions

anonymisées sont stockées et gérées dans une instance Heurist sécurisée, déployée par la TGIR Huma-Num.

L'analyse statistique des lettres de candidature a été menée au cours de l'automne 2021. Elle a porté sur le profil des candidats (âge, sexe, profession, lieu de résidence) et les lettres (nombre de signes, forme, primo-candidature). L'analyse qualitative repose sur l'analyse du contenu des lettres. Ont été plus particulièrement analysés ici les aspects liés à la motivation des candidats et à l'expression de leurs souvenirs de lecture. Les outils mobilisés, notamment l'interface de détection et d'annotation du projet READ-IT ont été d'un apport très précieux. Cette phase est en cours d'achèvement.

Autres résultats

Publications académiques, volet numérique, actions de formation, médiation grand public, nouveaux partenariats... (0,5 page)

- Une présentation partielle des résultats a eu lieu dans le cadre de la Journée REUSE organisée par le réseau SHN le 19 octobre 2021 à Nantes.
- Elena PRAT présentera une synthèse plus complète lors de la conférence en ligne IGEL 2022 qui se tiendra du 14 au 16 juillet 2022.
- Une action de présentation des résultats est également envisagée avec le réseau Mobilis en direction des professionnels du livre et de la lecture des Pays-de la Loire.
- Prise de contact avec Radio-France pour la valorisation des archives
- Présentation du corpus en cours de Master ECI
- Séminaire de présentation à Mayence (Gutenberg Forschungskolleg) prévu en octobre 2022 (dans le cadre de Lexilect).

2 - Bilan scientifique

(1,5 pages max.)

Ce projet avait pour ambition notamment d'apporter des éléments de connaissance sur une population de lecteurs ordinaires à travers les lettres de candidatures pour participer au jury du Prix du Livre Inter.

Le travail statistique sur le profil des candidat.e.s a pu être mené à son terme. Sur le plan de l'analyse qualitative, il est intéressant de noter que la plupart des candidats partagent des souvenirs et des détails de leur vie personnelle avec les journalistes de France Inter. Les livres servent souvent de prétexte pour évoquer des éléments autobiographiques et vice versa. L'analyse qualitative a fait apparaître des thèmes récurrents comprennent les souvenirs de lecture d'enfance, les émotions ressenties

pendant la lecture, les descriptions des lieux de lecture privilégiés, l'évocation d'un sentiment de communauté avec d'autres auditeurs de France Inter, les prises de position (généralement négatives) sur les pratiques de lecture numérique émergentes, et enfin l'expérience presque cathartique de la rédaction de la lettre, qui permet aux candidats de mieux se connaître en tant qu'individus et lecteurs.

Ensuite, en raison de la taille limitée de cet échantillon, ce projet avait pour but de tester et valider l'organisation et la robustesse d'une chaîne de travail en vue du traitement de données du même type mais en volume nettement plus important. La qualité de la transcription automatique de la parole a été très satisfaisante une fois les problèmes de prise de son réglés et la plateforme de transcription développée par le LIUM a pu ainsi profiter des corrections manuelles des fichiers pour améliorer la qualité du traitement final.

Le protocole d'anonymisation proposé aux propriétaires des archives a pu être implémenté sans difficulté et n'a pas eu de conséquences sur le traitement statistique et sur l'analyse qualitative.

L'utilisation de la plateforme développée par le projet READ-IT pour la reconnaissance des entités nommées et la détection automatique des expériences de lecture contenues dans les lettres de candidatures s'est révélée décisive et TALL a pu apporter ses commentaires et, là aussi, contribué à l'amélioration de la performance des algorithmes, notamment en matière d'extraction de vocabulaire et d'annotations. Ceci permettra aux données produites par TALL d'être très largement réutilisées par le projet LEXILECT.

3 - Perspectives

Quelles perspectives à l'issue du financement « Amorçage » ?

Nouveaux partenariats envisagés, dépôt de candidature(s) à des appels à projets de plus grande envergure (le cas échéant, indiquer spécifiquement les AAP visés), publications à venir... (0,5 page).

Sur le plan de la dissémination académique, une publication est en cours de rédaction pour une revue internationale (Digital Humanities Quarterly, DHQ).

Une demande a été faite à Radio-France (propriétaire des archives) afin d'envisager une collaboration plus large qui porterait sur l'ensemble des lettres de candidature depuis 1976 provenant de l'ensemble des régions. Ceci pourrait déboucher sur le dépôt d'une réponse à un appel à projet ANR.

Les données produites par TALL seront réutilisées par le projet LEXILECT.