



Un réseau de recherche pour l'étude scientifique du français (oral)

Colloque PFC 5 décembre 2019

Gabriel Bergounioux



L'étude patrimoniale des langues

Plusieurs pays ont choisi d'adosser l'étude de leur langue officielle à une institution publique. Quelques exemples :

Institut für Deutsche Sprache (IDS – Allemagne)

National Institute for Japanese Language and Linguistics (NINJAL – Japon)

Instytut Podstaw Informatyki – Polskiej Akademii Nauk (IPI PAN – Pologne)

L'étude patrimoniale des langues

Les missions confiées à ces établissements :

- la collecte patrimoniale et l'archivage documentaire,
- le soutien à la recherche scientifique sur les données,
- la diffusion des ressources,
- la valorisation par des applications pédagogiques ou le traitement automatique des langues

L'étude patrimoniale des langues

Les transformations induites par le développement de l'informatique et le déploiement d'Internet ont modifié les conditions d'accessibilité aux textes et aux enregistrements. Mise en ligne, requêtes et exploitation requièrent :

- (i) des outils accessibles et libres de droit ;
- (ii) une protection des données, des ayants-droit et des utilisateurs ;
- (iii) la préservation pérenne des fichiers.

L'étude patrimoniale des langues

Une difficulté supplémentaire : les données de l'oral

La plupart des instruments disponibles aujourd'hui, dans la continuité de l'investissement engagé dans les bibliothèques et les centres d'archives, ont été conçus pour l'exploration de sources écrites normalisées. La complexité des tâches en reconnaissance de la parole et dans le travail de transcription souligne l'ampleur de la tâche à fournir dans le domaine des données sonores.

Etat des lieux en France

La première institution créée pour traiter des questions de langue, l'Académie française (1635), n'est pas en charge de l'étude scientifique du français.

Les établissements d'enseignement supérieur ne sont pas équipés pour la conservation, la restauration et la diffusion.

Les établissements de conservation et de diffusion (BnF, Archives Nationales) n'ont pas de mission spécifique de recherche sur les contenus qu'ils hébergent.

Initiatives institutionnelles

1930 : une première tentative universitaire (et un échec), l'inventaire lexical de Mario Roques.

1960 : création de l'Institut National de la Langue Française (INaLF) à Nancy avec pour objectif d'éditer un dictionnaire, le *Trésor de la Langue Française*, qui serait au XX^e siècle ce qu'était cent ans auparavant le Littré, sur le modèle de l'*Oxford English Dictionary*.

2001 : l'ATILF (Analyse et Traitement Informatique de la Langue Française – UMR 7118) remplace l'INaLF en fusionnant avec le laboratoire LANDISCO.

Initiatives institutionnelles

2001 : Parallèlement est mise en place une Fédération de Recherche, l'Institut de Linguistique Française (ILF – FR 2393) sous forme d'un réseau des unités du CNRS impliquées dans l'étude du français, avec une extension ultérieure à des équipes d'accueil universitaires (EA).

2005 : création par le CNRS du Centre National de Ressources Textuelles et Lexicales (CNRTL) à Nancy et du Centre de Ressources sur les Données de l'Oral (CRDO) à Aix et Paris pour la gestion des fonds.

Initiatives institutionnelles

2007 : Création du TGE ADONIS (CNRS)

2011 : Création d'IRCOM (Infrastructure de Recherche - Corpus Oraux et Multimodaux) au sein d'IR Corpus (CNRS)

2012 : le CRDO devient COCOON (Collection de Corpus Oraux Numériques) à Paris – LLL et LACITO – et le SLDR (Speech and Language Data Repository) avec le LPL à Aix.

2013 : Fusion d'ADONIS et d'IR Corpus dans une nouvelle structure, la TGIR Huma-Num (Humanités Numériques)

Initiatives institutionnelles

Aux côtés des infrastructures de recherche, la collecte, la conservation, l'archivage et la diffusion des données patrimoniales sont assurés par :

la BnF (en particulier, pour les documents oraux, le service des documents sonores du département de l'audiovisuel, héritier de la Phonothèque nationale),

les Archives nationales,

l'Institut National de l'Audiovisuel (INA).

Initiatives institutionnelles

Le traitement des ressources est réalisé par les laboratoires en lien avec

la TGIR Huma-Num (pour les linguistes, à travers le consortium CORLI – Corpus Langue Interaction – qui a pris la suite des consortiums IRCOM et Corpus écrits),

les plateformes COCOON, SLDR et CNRTL,

l'Equipex ORTOLANG (Open Resources and Tools for Language) qui n'a pas été renouvelé en 2020.

Un Institut pour la Langue Française

L'arrêt en 2018 de l'Institut de Linguistique Française par le CNRS et en 2019 de l'Equipex Ortolang sur décision du Programme Investissements d'Avenir rend nécessaire la création d'un lieu pour préserver les relations entre la conservation, la recherche, l'enseignement et la valorisation de la langue française, en particulier pour les données orales.

Tel est le projet de l'Institut de la Langue Française à Orléans (ILFO).

Un Institut pour la Langue Française

L'ILFO vient en renfort de la politique de l'Etat français conduite à travers la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF) auprès du Ministère de la Culture,

mais aussi des actions conduites par les ministères de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche et des Affaires Etrangères (Instituts Français), notamment pour des tâches d'expertise et de conseil.

Un Institut pour la Langue Française

L'ILFO est surtout complémentaire de la décision présidentielle, prise en mars 2018, d'ouvrir une Cité Internationale de la Langue Française à Villers-Cotterêts, un projet élaboré en marge des universités et du CNRS.

L'ILFO bénéficie déjà de l'engagement financier de la métropole d'Orléans et de la Région Centre Val de Loire, du soutien de l'Université d'Orléans, du CNRS et du Ministère de la Culture.

Un Institut pour la Langue Française

Calendrier prévisionnel

2019 Engagement régional et national (financement)

2020 Lancement des études et organisation en réseau

2021 Travaux immobiliers et programmation scientifique

2022 Ouverture (en même temps que Villers-Cotterêts)

Un réseau à construire

L'ILFO est un outil à construire par et pour les linguistes, afin de poursuivre les études scientifiques sur la langue française, en particulier sur les données de l'oral.

Le projet a une ambition internationale, à l'échelle de toutes les formes de francophonie (FLM, FLS, FLE, FOS, FOU...), avec un focus sur les données orales déjà présent dans FLORAL. La présence de PFC comme l'un des acteurs majeurs dans la réflexion et la participation à cette initiative ouverte à tous les chercheurs est attendue selon des modalités qui sont à discuter de concert.

Un réseau à construire

Quel objectif ? A partir d'un fonds patrimonial centré sur l'oral, approfondir les liens en France et à l'international entre les chercheurs, les différents laboratoires et les institutions de conservation avec

la BnF et ses homologues dans les différents pays francophones,

les services d'appui à la recherche comme la TGIR Huma-Num ou CLARIN au niveau européen,

les plateformes comme le CNRTL, COCOON...

Un réseau à construire

L'apprentissage du français ne relève pas du périmètre d'activité d'ILFO qui n'est pas destiné à devenir un énième acteur sur ce marché.

En relèvent à l'inverse la réflexion sur la diffusion de la langue et la conception d'outils (ceux qui prennent en compte la langue parlée notamment), les technologies de l'EAD et le développement de didacticiels. Dans ce domaine, les recherches conduites dans le cadre d'I-PFC constituent une référence pour la conception des produits.

Un réseau à construire

Concernant la valorisation, on mentionnera, dans le cadre du plan Intelligence Artificielle, le lancement du consortium *Le Voice Lab* qui réunit des start up et des laboratoires universitaires autour d'un centre d'agrégation de contenus oraux et de leur outillage.

A présent, il reste à réfléchir sur les modalités de l'engagement des chercheurs, des équipes et des programmes dans un réseau où les moyens de l'ILFO sont au service de tous les linguistes.

Conclusion

Une affaire à suivre, donc,

et un appel à toutes les bonnes volontés.

gabriel.bergounioux@univ-orleans.fr