

Compte rendu de la journée d'étude QuECJ (groupe de travail CORLI)

Mardi 2 juillet 2019
Paris Diderot - Site Olympe de Gouges

Conférences

- **Le recueil de données langagières dans le cadre de la loi Jardé et du RGPD** - Muriel Lalin & Gilles Pouchoulin (LPL)
- **La science ouverte : questions juridiques et éthiques que pose le recueil et la diffusion des données audiovisuelles en SHS** - Emilie Jouin (ICAR)

NB : Les supports de présentation seront prochainement disponibles sur la page web du groupe de travail (<https://corli.huma-num.fr/gp4>).

Tour de table : De quoi avons-nous besoin ? Quelles sont les attentes vis-à-vis du GT CorLi ?

(Discussion animée par Katja Ploog)

Parmi les laboratoires représentés : ICAR (Lyon), IIAC (Paris), LACITO (Paris), LIDILE (Rennes), LLL (Orléans), LPL (Aix), LPP (Paris), LRL (Clermont-Ferrand), Modyco (Nanterre), PRISMES (Paris), SFL (Paris), STL (Lille), SPD (CNRS) ...

1) Comité d'éthique autour de la linguistique hors établissement

Les comités d'éthique existent dans un petit nombre d'universités seulement, et ne sont pas spécialisés en SDL. Certains refusent de prendre position.

Il est proposé de créer un comité d'éthique hors établissement et spécifique aux sciences du langage.

2) Réseau d'experts référents éthiques dans le domaine linguistique

La soumission de projets de recherche comprend de plus en plus souvent la nomination d'un référent éthique extérieur au projet, elle est de rigueur pour les projets européens.

Il est proposé de créer un réseau d'experts référents éthiques spécifique aux sciences du langage.

3) Guide données personnelles du CNRS

E. Masson (juriste au Service Protection des Données du CNRS) nous informe de la publication d'un guide pour la protection des données dans la recherche pour les SHS, en nous encourageant à lui faire des retours :

(<https://intranet.cnrs.fr/directions/dpd/Pages/guides-protection-des-donn%C3%A9es-recherche.aspx> et disponible prochainement sur la page web du groupe de travail).

4) Partage des données

Il est proposé que la communauté CORLI affirme comme principe de base pour les dépôts de corpus le partage des données avec la communauté scientifique (NB. tel que le prévoit la charte d'utilisation de la plateforme Ortolang). Or, la volonté de la communauté de partager les données de la recherche se heurte au caractère personnel ou sensible de certaines des données.

Il est proposé de s'inspirer des systèmes d'archivage développés en SHS, p.ex. celui pour les enquêtes ethnographiques (projet Archethno), qui assure une pérennité des données et une sécurité d'accès aux données en intégrant une réflexion sur les niveaux d'anonymisation liés aux types d'accès et types de données.

Réutilisation d'un corpus

L'utilisation de corpus dont on n'est pas l'auteur soulève des questions particulières. Elle est d'autant plus opportune qu'il s'agit de la documentation de langues peu dotées.

De manière générale, il semble important que la procédure respecte les cadres d'utilisation (réutilisation) d'un corpus tels qu'ils ont été définis par son « auteur ». En amont, l'on peut inciter les auteurs à prévoir la chaîne de traitement en pensant aux ré-utilisations, au-delà de la recherche initiale. L'utilisation seconde par des tiers peut être régulée par des conventions d'exploitation qui définissent les processus de récupération des données, le cadre spatio-temporel y compris. En aval, si le locuteur/informateur du corpus est encore en vie, l'on peut faire signer un consentement *a posteriori* qui autorise la mise à disposition des données pour un usage recherche. Lorsqu'il n'est plus possible d'obtenir les autorisations conformes à la loi actuelle de protection des données, c'est la bonne foi de l'auteur du corpus qui devra être démontrée.

Deux problèmes différents liés à la récupération de données anciennes sont pointés : ceux liés aux droits d'auteur (inaliénables) et ceux liés aux droits des données privées (aliénables). La personne qui s'opposerait à la diffusion de données personnelles la concernant, peut légitimement exiger le retrait des données du domaine public.

Pour aiguiller les choix des chercheur/es dans la prise en main de corpus anciens, il est proposé de constituer une "boîte à outils" sur la base des expériences abouties. Le CNRS propose également son soutien à concevoir les procédures adaptées pour la revalorisation de corpus anciens.

5) Formation / information

Le réseau MateSHS (www.mate-shs.cnrs.fr) organise des formations et des journées d'information autour de la production et l'analyse de données.

L'"urgence" des questionnements révèle des besoins de formation importants dans la communauté SDL (p.ex. rédaction du consentement selon le type de public). La diversité des cultures universitaires et des cadres politico-légaux (à l'internationale) empêche d'évoluer vers un protocole "standard".

Il est proposé : de partager les formulaires types (consentements, conventions...) pour faciliter la rédaction, d'établir une check-list des questions à se poser (et où trouver les réponses), de programmer des formations "sur site" pour tirer vers le haut les pratiques éthiques liées aux données.

En outre : de mutualiser les informations et, idéalement, les outils d'anonymisation mêmes, et de mener le travail en collaboration avec les DPO, les membres de comités d'éthique et des experts.

Conclusion : **perspectives pour le groupe de travail**

Objectif 2019 : s'organiser en réseau

Feuille de route :

- **Rassembler les ressources existantes** et les mettre à disposition sur le site CORLI :
 - documents de référence (ex. chaine de traitement) ;
 - exemples de formulaires (ex. consentements (+ retour sur les formulaires utilisés et s'ils ont bien fonctionné)

Envoyez vos documents à Flora Badin (LLL) flora.badin@univ-orleans.fr (pour diffusion sur le site Corli)
- **Définition des “parcours types” des données**
 - Emilie Jouin, ICAR (emilie.jouin@ens-lyon.fr) & Celine Dugua, LLL (celine.dugua@univ-orleans.fr),
- **Inventaire détaillé des outils** existants et des besoins (documents types, etc.)
 - Anne-Lyse Minard, LLL (anne-lyse.minard@univ-orleans.fr) & ???
- **Inventaire des comités d'éthique** dans les établissements en France (info E.Masson) et de la prise en compte des SDL :
 - Muriel Lalain, LPL (muriel.lalain@lpl-aix.fr) ;
 - fonctionnement dans d'autres pays
 - Royaume-Uni : Catrin Bellay, LIDILE (catrin.bellay@univ-rennes2.fr) ;
 - Belgique : ???
- **Constitution d'un réseau d'experts** en corpus SDL : ex. Pour être référents éthiques extérieurs aux projets
 - Aliyah Morgenstern, PRISMES (aliyah.morgenstern@sorbonne-nouvelle.fr)
- **Coordonner la communication avec le réseau SHS** : rassembler les informations et retours sur le guide “données personnelles” publié par le CNRS
 - Caroline Cance, LLL (caroline.cance@univ-orleans.fr) & Emilie Masson, CNRS (Emilie.Masson@cnrs.fr)

Objectif 2020 : concevoir des modules de formation

- en fonction des types de publics (pour promouvoir les bonnes pratiques à moyen terme, visés sont en priorité les doctorant/es et la formation des formateurs)
- en fonction du type / cadre / de l'envergure du projet (ANR vs. doctorat)
- pour différents cadres temporels (école d'été, ANF, demi-journée)