

7 décembre 2021
DATA SHS

Retour d'expérience autour de données sensibles :
enjeux juridiques, éthiques et techniques de la collecte et la diffusion de
données audiovisuelles en santé

Emilie Jouin & Justine Lascar

emilie.jouin@ens-lyon.fr justine.lascar@ens-lyon.fr

Ingénieures d'étude CNRS

Laboratoire ICAR – UMR 5191

- Interactions, Corpus, Apprentissages et Représentations

Trois tutelles : CNRS/ ENS de Lyon / Université Lyon 2

Sciences du Langage et Sciences de l'Education

Etude de la langue parlée en interaction

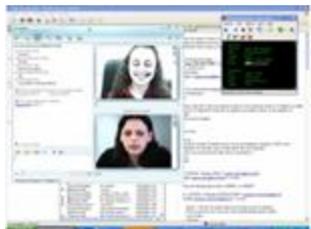
Méthodologie basée sur les corpus d'interactions sociales en « milieu naturel », enregistrés en vidéo et audio.

- Cellule Corpus Ingénierie Audiovisuelle – Justine Lascar
- Référente éthique et juridique pour ICAR – Emilie Jouin

<http://icar.cnrs.fr/>

<http://icar.cnrs.fr/ccf/>

Données écologiques



Corpus disponibles en ligne sur des bases de données :

- CLAPI (Corpus de Langue Parlée en Interaction) <http://clapi.icar.cnrs.fr/>
- ViSA (Vidéos de situations d'enseignement et d'apprentissage) <http://visa-video.ens-lyon.fr/visa-web/>

Projets et thématiques de recherche



Projet CIPSY



Projet
REMILAS



Projet MARTI

Retour d'expériences sur la collecte et la diffusion des données dans le cadre du projet REMILAS.

→ confrontation à des questions juridiques et éthiques et trouver des solutions techniques pour y remédier

Préparation du terrain

Démarches juridiques pour protéger les personnes et les données

Recueil de données **audiovisuelles (images, voix, propos)**

=> Données à caractère personnel et notamment de type sensible

=> Droit à l'image

❖ **Déclaration de traitement des données à caractère personnel**
auprès du Délégué à la protection des données pour enregistrement

❖ **Consentement éclairé et signature d'un formulaire d'autorisation**

- par les responsables des structures accueillantes (ex: collège, hôpital, association...)
- par les personnes enregistrées (majeures, mineurs, soignants, migrants...)

Sur le terrain – les questions éthiques

- ❖ **Consentement éclairé et signature d'un formulaire d'autorisation**
=> Difficile (contextes, lieux de santé, contraintes partage de langue) et pose des questions éthiques
- travail d'explication et de mise en confiance des participants (contexte migratoire et de santé)
- traduction des autorisations (allophones, analphabètes)
- aide apportée par les soignants et les interprètes (figures d'autorité)

Diffusion des données

Partage des données : diffusion restreinte des données

Publication et valorisation : la délicate diffusion des données

- ❖ rendre compte au **monde académique des avancées de la recherche** sur la communication plurilingue et en santé
- ❖ permettre des **retombées sociétales** directement utilisables par les acteurs de terrain (interprètes, intervenants sociaux et en santé) en proposant des ouvrages, formations et outils de formations / auto-formations
- ❖ **valoriser les résultats** des travaux autour de la vulnérabilité et les problèmes de langue **auprès du grand public**



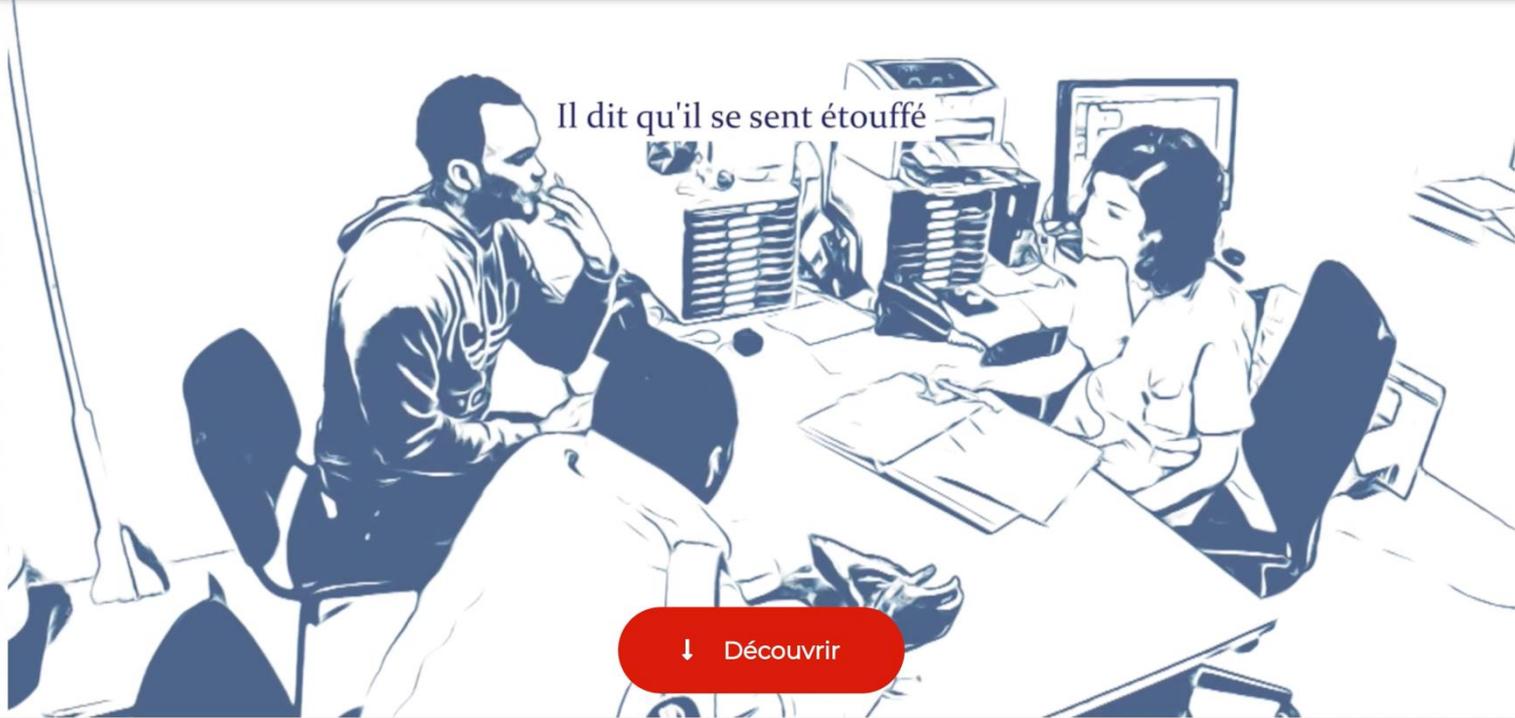
Techniques de pseudonymisation des :

- ❖ **Visages** : flouter ou ajouter un filtre de protection
- ❖ **Voix** : appliquer un traitement de modulation des voix
- ❖ **Autres données identifiantes** :
 - pseudonymiser les personnes et les données identifiantes dans les métadonnées et les transcriptions (attention aux éventuels recoupements d'informations dans les données et via les métadonnées. Ex: personne politique et parle oromos)
 - couvrir l'audio via des beeps pour effacer les données identifiantes (numéros de téléphones, noms...)

Exemples d'ano/pseudo-nymisation vidéo



2 images extraites du **Routledge Handbook on Public Service Interpreting** (à paraître)
Training interpreters in asylum settings: the REMILAS project de Anna Claudia Ticca, Véronique Traverso, Emilie Jouin



Il dit qu'il se sent étouffé

↓ Découvrir



Bienvenue sur **Odimedi**, la formation en ligne pour améliorer la communication entre **interprètes**, professionnels de la santé et patients.

Conclusion

- Diffusion restreinte de données sensibles touchant aux domaines de la santé mentale et somatique, de la migration et de la vulnérabilité au sens large.
- Application de mesures de protection pour la diffusion des données qui nécessitent de s'adapter à chaque projet et aux utilisations qu'on veut faire des données.
- Complexité de travailler sur des données sensibles mais pas impossible de les diffuser pour montrer la recherche.

