

# Table Ronde 1 : Capacités de progression des outils

Anaïs Wion (CNRS, Institut des Mondes Africains)



Journée d'études sur l'HTR des langues peu dotées dans les programmes de recherche et dans les établissements de conservation français

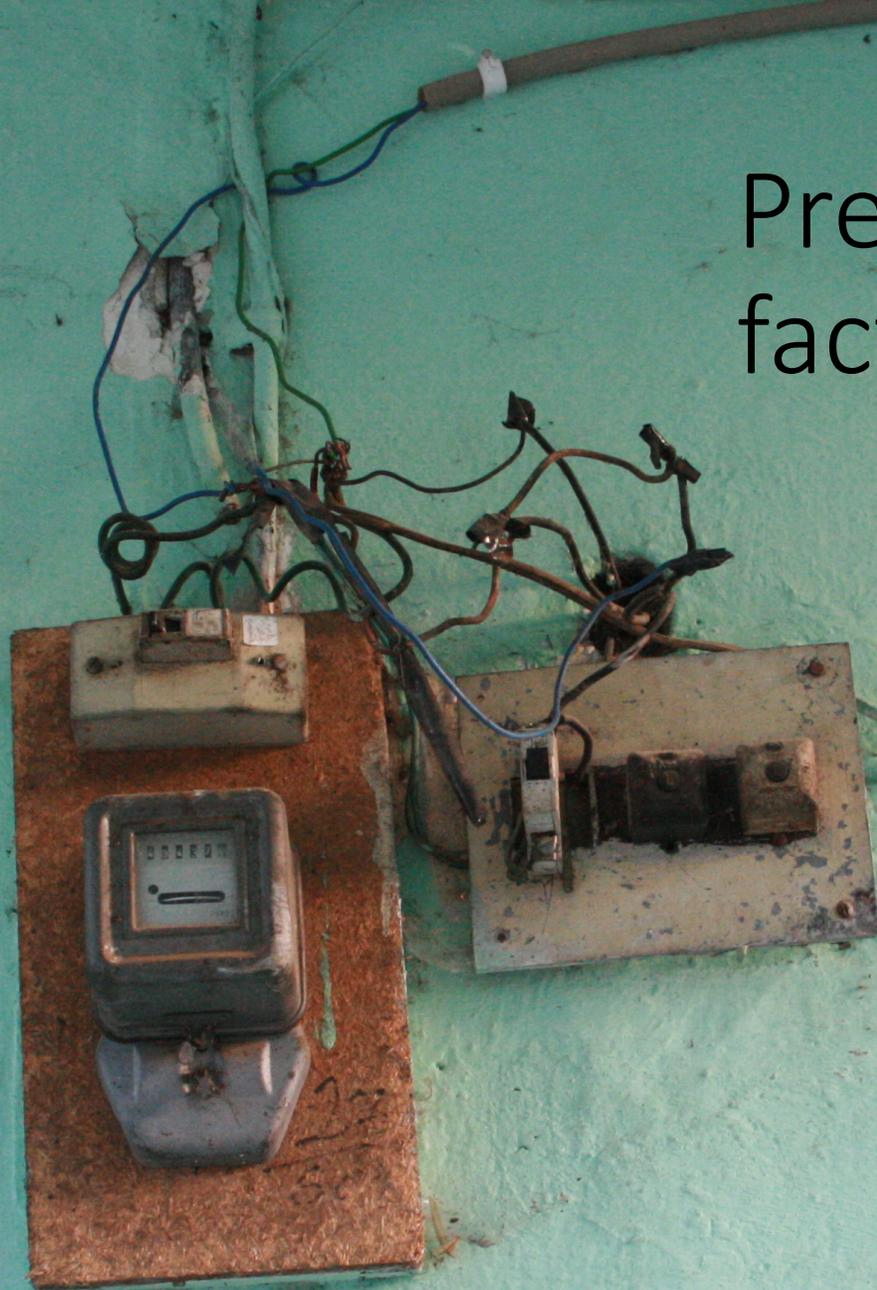
BnF –DISTAM, 14 février 2024

# Quels types de “progression” ?

- Progrès = meilleurs résultats
- Progrès ≠ plus de données, plus de moyens, plus de puissance de calcul
- Langues **peu dotées** → peu de moyens humains en DH dans les aires linguistiques concernées → petits budgets dans les pays concernés

# Prendre en compte des facteurs limitants

- Accès aux ressources énergétiques et à internet
- Budgets faibles : pour la mise en oeuvre des projets et l'usage des outils
- Questions éthiques et juridiques



# Recommandations

- Outils qui peuvent fonctionner en ligne ET en local
- Possibilité de ne pas verser les images sur un serveur
- Outils peu gourmands en ressources
- Outils accessibles en opensource et ré-utilisables pour différents projets

# Deux projets attentifs à ces questions (script éthiopien)

- **Projet Ge'ez-HWR** : see Samuel Grieggs et alii, “Automated Transcription of Gə'əz Manuscripts Using Deep Learning », *Digital Humanities Quarterly*, 2023 17/3

<https://github.com/grieggs/Ge-ez-HWR>

- **Projet HHD-Ethiopic** : see Birhanu Hailu Belay et alii, « A Historical Handwritten Dataset for Ethiopic OCR with Baseline Models and Human-level Performance », 2023 [pré-publication sur Hal-SHS]

<https://www.researchgate.net/profile/Birhanu-Belay-2>