



HAL
open science

Accroître la vitalité et la visibilité numérique des langues de France

Delphine Bernhard, Pascale Erhart, Dominique Huck, Pablo Ruiz Fabo,
Carole Werner

► To cite this version:

Delphine Bernhard, Pascale Erhart, Dominique Huck, Pablo Ruiz Fabo, Carole Werner. Accroître la vitalité et la visibilité numérique des langues de France. Humanités numériques en pédagogie et en recherche à la faculté des langues, Sep 2023, Strasbourg, France. , 2024, 10.34847/nkl.deafs14t . hal-04686233

HAL Id: hal-04686233

<https://hal.science/hal-04686233v1>

Submitted on 3 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DiViTal

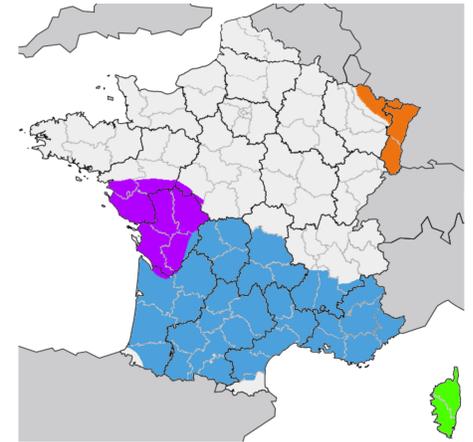
Accroître la vitalité et la visibilité numérique des langues de France

Delphine Bernhard, Pascale Erhart, Dominique Huck, Pablo Ruiz Fabo, Carole Werner

Faculté des Langues, Université de Strasbourg

Partenaires et langues régionales

- **Alsacien** : Université de Strasbourg, UR LiLPa
- **Corse** : Università di Corsica Pasquale Paoli, UMR LISA
- **Occitan** : Université Toulouse Jean Jaurès, UMR CLLE
- **Poitevin-Saintongeais** : Université de Poitiers, UR FoReLLIS



Constat

Les langues régionales de France font partie des langues dites « peu dotées » et accusent un retard important dans le développement de ressources linguistiques pour le traitement automatique des langues : corpus bruts et annotés, monolingues et parallèles, lexiques.

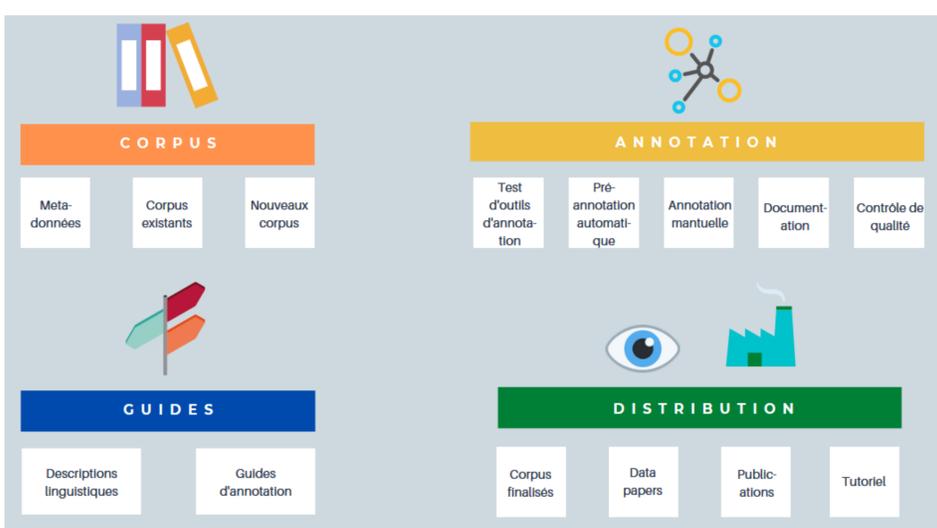
Défis

① Faible usage dans l'espace public et en ligne, limitant les données accessibles facilement

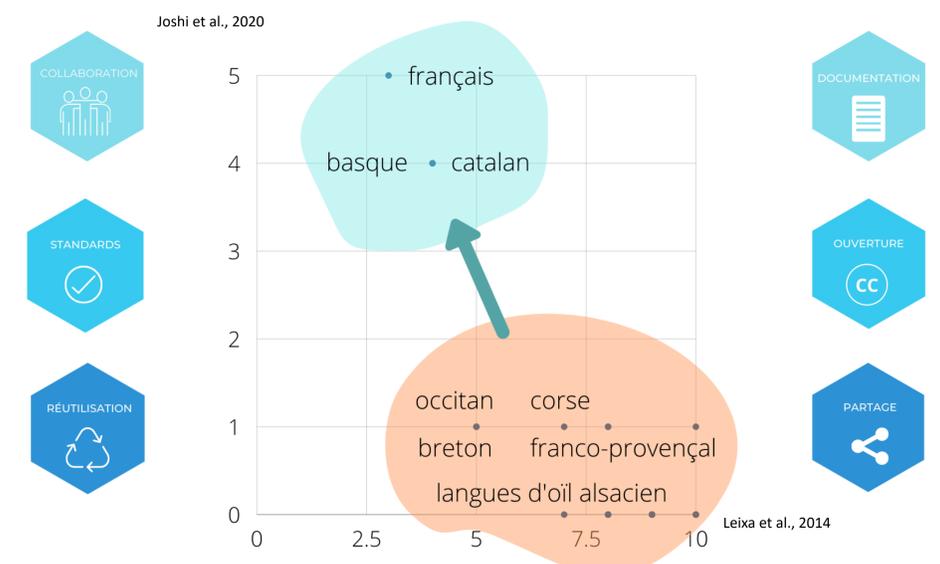
② Forte variation à l'écrit



Organisation et tâches 2022-2025



Objectifs



Références utiles

- "Inventaire des ressources linguistiques des langues de France". Leixa, Mapelli et Choukri, ELDA/DGLFLF – 2014
- "The State and Fate of Linguistic Diversity and Inclusion in the NLP World". Joshi, Santy, Budhiraja, Bali & Choudhury, ACL 2020

Légende :

- Axe vertical, classification selon Joshi et al. (2020) : 0 = langues extrêmement peu dotées, 5 = langues les mieux dotées
- Axe horizontal, classification selon Leixa et al. (2014) : 1 = excellent volume de ressources, 10 = volume faible