

## II. Méthodologie

Pour effectuer ce travail de recensement, une première recherche a tout d'abord été effectuée en collaboration avec l'infrastructure de recherche Progedo afin d'identifier les enquêtes issues de la statistique publique. Quetelet-Progedo-Diffusion est un entrepôt et catalogue de données qui a pour but de stocker et de donner accès à la communauté scientifique à des jeux de données issus de l'INED et de l'ADISP (Archives des Données Issues de la Statistique Publique<sup>6</sup>).

« PROGEDO (pour PROduction et Gestion des DONnées) est une très grande infrastructure (TGIR) visant à développer la culture des données en sciences humaines et sociales. Son champ est centré sur les données d'enquête. Elle soutient la création et la diffusion de grandes enquêtes internationales et l'usage des statistiques publiques par les chercheurs.

À l'échelle française, PROGEDO contribue, par le biais du portail Quêtelet-PROGEDO Diffusion à répondre aux problématiques d'archivage, de documentation et de diffusion des données françaises issues d'enquêtes ou de la statistique publique pour la recherche en sciences humaines et sociales. Le portail héberge les données issues des Archives de Données Issues de la Statistique Publique (plus de 1500 enquêtes), du Centre de Données Socio-Politiques de Sciences Po et du service d'enquête de l'Ined. PROGEDO soutient en outre le Centre d'Accès Sécurisé aux Données qui permet l'accès dans des conditions de sécurité élevées à des données individuelles très détaillées. »<sup>7</sup>

L'accès aux données ne se fait pas directement, mais après demande des chercheurs ou des étudiants, qui est par la suite examinée par Progedo : les données brutes sont alors remaniées de façon à remplir les critères d'anonymisation, afin que les enquêtés ne puissent pas être identifiés par les informations déclarées au cours de l'enquête.

Dans le catalogue de Progedo, on peut donc trouver un grand nombre d'enquêtes de l'INSEE, l'INED, mais aussi des services statistiques ministériels<sup>8</sup> tels que la DEPP (Ministère de l'Education), le DEPSD (Ministère de la Culture), le DSED (Ministère de l'Intérieur), etc. Dans un premier temps, la recherche parmi les bases de données du catalogue Progedo a permis d'identifier 29 études quantitatives de la statistique publique comportant des questions linguistiques, certaines d'entre elles n'étant d'ailleurs pas connues pour leur volet linguistique car ces données n'avaient pas été analysées ou fait l'objet de publications. Cette étape a permis d'identifier des enquêtes à l'échelle de la France métropolitaine, mais aussi des Départements des Outre-mer (DOM).

Dans un second temps, d'autres sources plus spécifiques ont été explorées. Tout d'abord, j'ai focalisé ma recherche sur les instituts statistiques des Collectivités des Outre-mer (COM). En effet, la Nouvelle Calédonie, la Polynésie française, ainsi que Wallis-et-Futuna, ont par dérogation l'autorisation d'avoir des instituts statistiques locaux en charge du recensement de leur propre territoire (cette compétence étant sinon attribuée à l'INSEE). Ces trois instituts statistiques des COM disposent d'une certaine liberté d'action dans leurs travaux et incluent notamment des questions linguistiques à leurs recensements. Il est ainsi possible de trouver des statistiques linguistiques très détaillées sur ces territoires, avec des échantillonnages quasi exhaustifs par rapport à la population totale.

---

<sup>6</sup> <http://www.progedo-adisp.fr/>

<sup>7</sup> <https://www.inshs.cnrs.fr/fr/progedo>

<sup>8</sup> Liste des services statistiques ministériels : <https://www.insee.fr/fr/information/2013075>

Par la suite, en partant de la liste des langues de France, j'ai effectué pour chacune d'elles une recherche des acteurs susceptibles de produire des enquêtes (régions, offices publiques de la langue, associations linguistiques), ainsi que des recherches bibliographiques afin d'identifier les sources des statistiques citées dans les différents travaux de recherche. Dans ce but, les DRAC ont également été contacté systématiquement afin de les interroger sur la présence d'enquêtes linguistiques sur leur territoire.

Pour finir, j'ai effectué une recherche de données parmi les acteurs extraterritoriaux, plus spécifiquement l'Union Européenne (Eurostat, euro-régions, Réseau Eurydice), l'OCDE (test PISA) et les pays voisins à la France (pour les langues régionales transfrontalières).