

Utiliser des données quantitatives dans les recherches sur le langage : l'exemple de trois approches disciplinaires

PLAN :

1 – Sources de données linguistiques

2- La méthode d'analyse combinée des représentations sociales des langues (MAC) : limites liées à la collecte de données en ligne - Marie SARRAUTE ARMENTIA

3 - Dictionnaires vs. Données orales : Le cas de la gémination de /r/ en anglais - Quentin DABOUIS

4 - L'approche bayésienne, une approche statistique alternative en psycholinguistique. Un exemple d'application : l'étude du rôle des syllabes dans la production écrite - Cyril PERRET

1 - Sources de données linguistiques

- Trouver des données linguistiques à travers les micro-données de recensement : **le projet IPUMS** → Claire KERSUZAN – PUD de **Bordeaux**
- **Elfe** : Etude Longitudinale Française depuis l'Enfance → Axel DELARUE – PUD de **Clermont-Ferrand**
- Trouver des données sur les langues via le portail **Quetelet-Progedo-Diffusion** → Gaëlle COZ – PUD de **Poitiers**

Les données sur les langues dans les recensements

**Le projet IPUMS
(Integrated Public Use Microdata Series)**

Trouver des données linguistiques à travers les micro-données de recensement : le projet IPUMS

Lien: <https://international.ipums.org/international/index.shtml>

IPUMS INTERNATIONAL
HOME | SELECT DATA | MY DATA | SUPPORT

IPUMS INTERNATIONAL
ABOUT
INTERNATIONAL PARTNERS
REGISTER
DONATE TO IPUMS

DATA
BROWSE AND SELECT DATA
ANALYZE DATA ONLINE
DOWNLOAD OR REVISE MY DATA

SUPPLEMENTAL DATA
GEOGRAPHY & GIS
FERTILITY, MORTALITY, MIGRATION
RESEARCH DATA ENCLAVE
LINKED HISTORICAL CENSUSES

DOCUMENTATION
REVISION HISTORY
SAMPLE DESCRIPTIONS
QUESTIONNAIRES
NAPP PROJECT
WORLD CENSUS FORMS

HARMONIZED INTERNATIONAL CENSUS DATA FOR SOCIAL SCIENCE AND HEALTH RESEARCH

IPUMS International is dedicated to collecting and distributing census microdata from around the world. The project goals are to collect and preserve data and documentation, harmonize data, and disseminate the harmonized data free of charge.

103 COUNTRIES - 547 CENSUSES AND SURVEYS - OVER 1 BILLION PERSON RECORDS

SOURCE DATA FOR IPUMS INTERNATIONAL ARE GENEROUSLY PROVIDED BY PARTICIPATING NATIONAL STATISTICAL OFFICES

CREATE AN EXTRACT
Browse Data

CREATE AN ACCOUNT
Register

- IPUMS (Integrated Public Use Microdata Series)
 - ✓ collecter, harmoniser et diffuser des échantillons de microdonnées de recensement anonymisées de tous les pays du monde, pour qu'elles soient utilisables par les chercheurs
 - ✓ Variables harmonisées (pour faciliter la comparaison entre pays) et données non harmonisées (variables brutes)
 - ✓ Métadonnées et documentation complète de chaque recensement



SELECT SAMPLES

SELECT HARMONIZED VARIABLES

HARMONIZED VARIABLES
 SOURCE VARIABLES

DISPLAY OPTIONS

HELP
[COUNTRY ABBREVIATIONS](#)

Select samples and variables to

- HOUSEHOLD
- PERSON
- A-Z
- SEARCH
- TECHNICAL PERSON
- CONSTRUCTED FAMILY INTERRELATIONSHIP
- DEMOGRAPHIC
- FERTILITY AND MORTALITY
- NATIVITY AND BIRTHPLACE
- ETHNICITY AND LANGUAGE
- EDUCATION

CREATED BY: [NATIONAL INSTITUTES OF HEALTH](#), [NATIONAL SCIENCE FOUNDATION](#), [STAT/TRANSFER](#), AND [UNIVERSITY OF MINNESOTA](#).

		BY COUNTRY															BY YEAR				Variable								
		ARGENTINA					ARMENIA					AUSTRIA					BELARUS				BENIN				BOLIVIA				Variable
		1970 1980 1991 2001 2010					2001 2011					1971 1981 1991 2001 2011					1999 2009				1979 1992 2002 2013				1976 1992 2001 2012				Variable
+	SPEAKENG	Speaks English															P				X	SPEAKENG							
+	SPEAKFIL	Speaks Filipino															P					SPEAKFIL							
+	SPEAKIRSH	Speaks Irish															P					SPEAKIRSH							
+	SPEAKRUS	Speaks Russian															P					SPEAKRUS							
+	SPEAKIND	Speaks indigenous language															P					SPEAKIND							
+	SPEAKPORT	Speaks Portuguese, Mozambique															P					SPEAKPORT							
+	LANGAM2	Second language spoken, Armenia															P				X X	LANGAM2							
+	LANGBJ	Primary language spoken, Benin															P					LANGBJ							
+	LANGBY	Language spoken at home, Belarus															P					LANGBY							
+	LANGBY2	Second language spoken, Belarus															P				X X	LANGBY2							
+	LANGBO	Language spoken, Bolivia															P					LANGBO							
+	LANGBW	Language spoken at home, Botswana															P					LANGBW							
+	LANGBF	Language spoken, Burkina Faso															P					LANGBF							
+	LANGCA	Language spoken at home, Canada															P					LANGCA							
+	LANGCR	Language taught to speak, Costa Rica															P					LANGCR							
+	LANGEC1	Languages spoken, Ecuador															P					LANGEC1							
+	LANGEC2	Indigenous language spoken, Ecuador															P					LANGEC2							
+	LANGHN	Language spoken, Honduras															P					LANGHN							
+	LANGSV	Non-Spanish language spoken, El Salvador															P					LANGSV							
+	LANGGN	Language of literacy, Guinea															P					LANGGN							
+	LANGGN2	National language spoken, Guinea															P					LANGGN2							

Documentation détaillée pour chaque variable: l'exemple de la variable harmonisée « Langue maternelle » dans les recensements canadiens

MTONGCA ADD TO CART SELECT SAMPLES

Mother tongue, Canada
[Return to Ethnicity and Language variables list](#)

CODES DESCRIPTION COMPARABILITY UNIVERSE AVAILABILITY QUESTIONNAIRE TEXT SOURCE VARIABLES

Codes and Frequencies

Category availability view
 Case-count view

An 'X' indicates the category is available for that sample

Code	Label	canada 1971	canada 1981	canada 1991	canada 2001	canada 2011
00	NIU (not in universe)	-	-	-	X	-
01	English single responses	X	X	X	X	X
02	French single responses	X	X	X	X	X
03	English and French	-	-	X	X	X
04	Aboriginal languages	X	-	X	X	X
05	German	X	X	X	X	X
06	Netherlandic languages	X	X	X	X	-
07	Scandinavian languages	X	-	-	-	-
08	Italian	X	X	X	X	X
09	Spanish	-	-	X	X	X
10	Portuguese	-	X	X	X	X
11	Polish	X	X	X	X	X
12	Ukrainian	X	X	X	X	-
13	Greek	-	X	X	X	-
14	Chinese	-	X	X	X	X
15	Austro-Asiatic languages	-	-	X	X	X
16	Arabic	-	-	X	X	X

Modalités /effectifs

Description de la variable

Comparabilité de la variable entre les différents recensements

Variables brutes utilisées pour construire la variable harmonisée

Libellé de la question (et instructions aux enquêteurs)

Années des recensements pour lesquelles la variable est disponible

Champ de la variable (restriction d'âge...)

Les pays pour lesquels des données linguistiques sont disponibles à travers les recensements IPUMS



- Trois types de questions à travers les recensements :
 - connaissance des langues officielles ou indigènes,
 - langues parlées à la maison (parfois au travail),
 - langue maternelle
 - Mais variabilité entre pays et dans le temps concernant:
 - le choix et la formulation des questions (question ouverte ou non, langue maternelle: première langue apprise/langue parlée par la mère, ...),
 - l'univers des répondants (critère d'âge,...),
 - le choix de la nomenclature (dénominations et délimitations des langues et des variétés linguistiques, ...)
- Regarder la façon dont les données sur les langues sont collectées (métadonnées) et la distribution des variables brutes



Elfe : Etude Longitudinale
Française depuis l'Enfance

Acquisition du langage à l'enfance

Elfe : Etude Longitudinale Française depuis l'Enfance

Lien : <https://www.elfe-france.fr/fr>



Étude Longitudinale Française depuis l'Enfance



À LA UNE

► **Août 2022 - [Résultats]** L'écart de réussite en mathématiques entre les filles et les garçons inexistant avant le Cours préparatoire

► **Juin 2022** - Thierry Lhermitte et la communauté scientifique remercient les familles Elfe !

► **Mai 2022** - Santé des enfants et inégalités : un rapport de la Cour des comptes documenté par les données de l'étude Elfe

► [Toutes les actualités](#)

LES GRANDES ÉTAPES DE L'ÉTUDE

- Suivi des enfants jusqu'à l'âge adulte
- 18 000 enfants ont été inclus dans l'étude en 2011, soit 1 naissance sur 50 cette année
- 150 chercheurs mobilisés sur l'étude
- Thèmes d'étude : **santé des enfants, leur scolarité, leur alimentation, leur vie familiale et sociale ou encore leur environnement**
- Exemple de question :
- **Quels sont les effets du mode d'accueil du jeune enfant sur ses relations avec les autres enfants, son intégration à la maternelle et son acquisition du langage ?**

Elfe : Etude Longitudinale Française depuis l'Enfance

Lien : <https://www.elfe-france.fr/fr>

ACCÈS AUX DOCUMENTS ET AUX DONNÉES



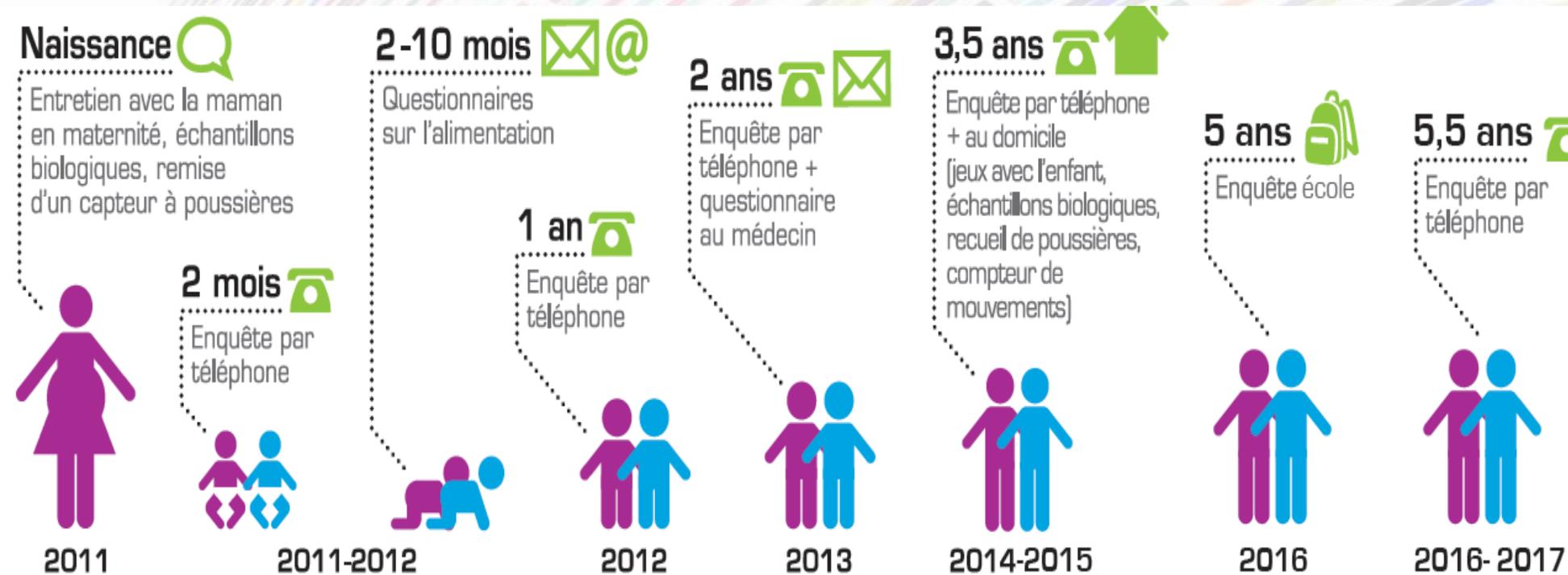
➤ [Plateforme d'accès aux données](#)

Nous contacter : ➤ contact@elfe-france.fr

➤ [Voir les projets de recherche en cours](#)

Données disponibles

- Enquête en maternité
- Enquête 2 mois
- Alimentation de 2 à 10 mois
- Enquête 1 an
- Enquête 2 ans
- Enquête 3,5 ans
- Enquête École maternelle
- Enquête ELFE-PMI



Lien : <https://www.elfe-france.fr/fr>

Plusieurs variables sont alors construites grâce aux données de l'enquête

Voir le détail : https://pandora-elfe.inserm.fr/public/?page_html&id=6

DESCRIPTION DES VARIABLES CONSTRUITES

Vous trouverez ci-dessous des informations sur une liste de variables construites par les utilisateurs Elfe.
Ces informations correspondent à :

1. Programmes vous permettant de construire la variable :

- [Score de dépression maternelle à 2 mois](#)

2. Description de la construction de variables mises à disposition sous forme de bases appariables à votre panier de données Elfe :

Information : Entre parenthèse, figure le nom de la variable à sélectionner dans votre panier de variables pour obtenir l'appariement. Elle se situe sur la page « Recherche par enquêtes : variables construites ».

- [Codage des pathologies maternelles avant la grossesse \(EQR15\)](#)
- [Alimentation de la mère pendant les 3 derniers mois de la grossesse \(EQR1\)](#)
- [Score de qualité de l'alimentation de la mère pendant la grossesse \(EQR21\), \[Annexe 1\]\(#\), \[Annexe 2\]\(#\)](#)
- [Activité physique de la mère pendant la grossesse \(EQR9\)](#)
- [Codage des professions et secteurs d'activité des mères pendant la grossesse pour caractériser les expositions professionnelles \(EQR2\)](#)
- [Note méthodologique codage profession père \(EQR5\)](#)
- [Indicateurs de croissance fœtale \(EQR13\)](#)
- [Apurement des données anthropométriques de 0 à 9 ans et mise à disposition d'indicateurs de croissance \(EQR20\)](#)
- [Codage des malformations congénitales \(EQR11\)](#)
- [Alimentation lactée du nourrisson \(EQR3\)](#)
- [Diversification alimentaire \(EQR16\)](#)
- [Variables sociodémographiques construites 0-2 ans, 3.5 ans \(EQR12\)](#)
- [Langues parlées à l'enfant à 2 mois, 1 an, 2 ans \(EQR29\)](#)
- [Traitement des données concernant le sommeil et les habitudes de sommeil de 1 à 5 ans \(EQR38\)](#)

[Cliquer ici pour voir plus](#)

Lien : <https://www.elfe-france.fr/fr>



Santé des enfants et inégalités : un rapport de la Cour des comptes documenté par les données de l'étude Elfe

2022

À la demande de la commission des affaires sociales de l'Assemblée nationale, la Cour des comptes a réalisé une enquête relative à la santé des enfants, en s'appuyant en partie sur les données de l'Étude longitudinale française pour l'enfance (Elfe).

[Lire la suite](#)



Les préparations infantiles : facteurs sociaux et de santé

2018

Un allaitement maternel exclusif est recommandé jusqu'à l'âge de 6 mois. Lorsque celui-ci n'est pas possible ou souhaité, ou alors en relai ou accompagnement de l'allaitement maternel, il existe une grande variété de préparations infantiles disponibles sur le marché sous forme de « laits » en poudre ou liquides pour nourrissons [...]

[Lire la suite](#)



Déterminants de la conduite de la diversification alimentaire

2017

Le Programme National Nutrition Santé (PNNS) conseille l'initiation de la diversification alimentaire à partir du 6e mois révolu, mais pas avant le 4e mois révolu, avec une introduction progressive des groupes d'aliments. Cependant, les caractéristiques de cette initiation sont encore mal connues en France. [...]

[Lire la suite](#)



Les pratiques de l'allaitement

2014

Les résultats de l'étude Elfe confirment la tendance à l'augmentation de l'initiation de l'allaitement maternel en maternité documentée depuis plusieurs années, en parallèle des nombreuses actions de santé publique qui ont été initiées pour promouvoir l'allaitement maternel. [...]

[Lire la suite](#)



Durée de l'allaitement en France

2015

Parmi les 70 % de mères ayant initié un allaitement à la maternité, la médiane de la durée totale d'allaitement est de 17 semaines et celle de l'allaitement prédominant de 7 semaines. Seuls 19 % des enfants reçoivent encore du lait maternel à 6 mois. [...]

[Lire la suite](#)

L'alimentation des femmes enceintes : respect des recommandations

2015

L'alimentation des femmes enceintes : respect des recommandations

Par Manik Kadawathagedara, chercheuse à l'Inserm, CRESS

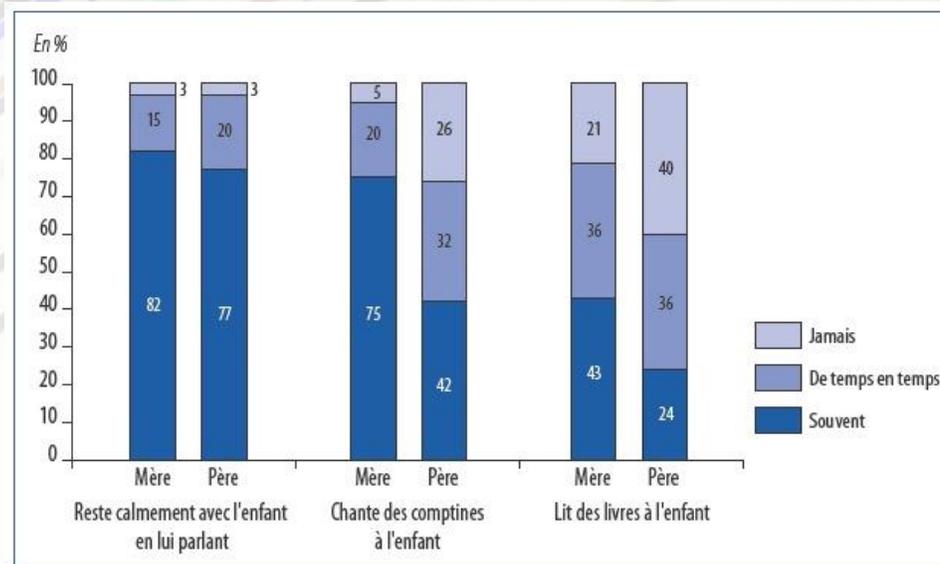


[Lire la suite](#)

Tableau n° 2 : retard de développement chez les enfants de 3,5 ans (en %)

	En couple	Foyer mono-parental	Mère en activité	Mère sans activité	Père en activité	Père sans activité
Difficultés de langage	4,5	8,0	2,5	10,2	4,0	10,9
Difficultés de développement psychomoteur	8,4	11,5	6,7	13,5	7,9	14,4

Source : Enquête Elfe-PMI



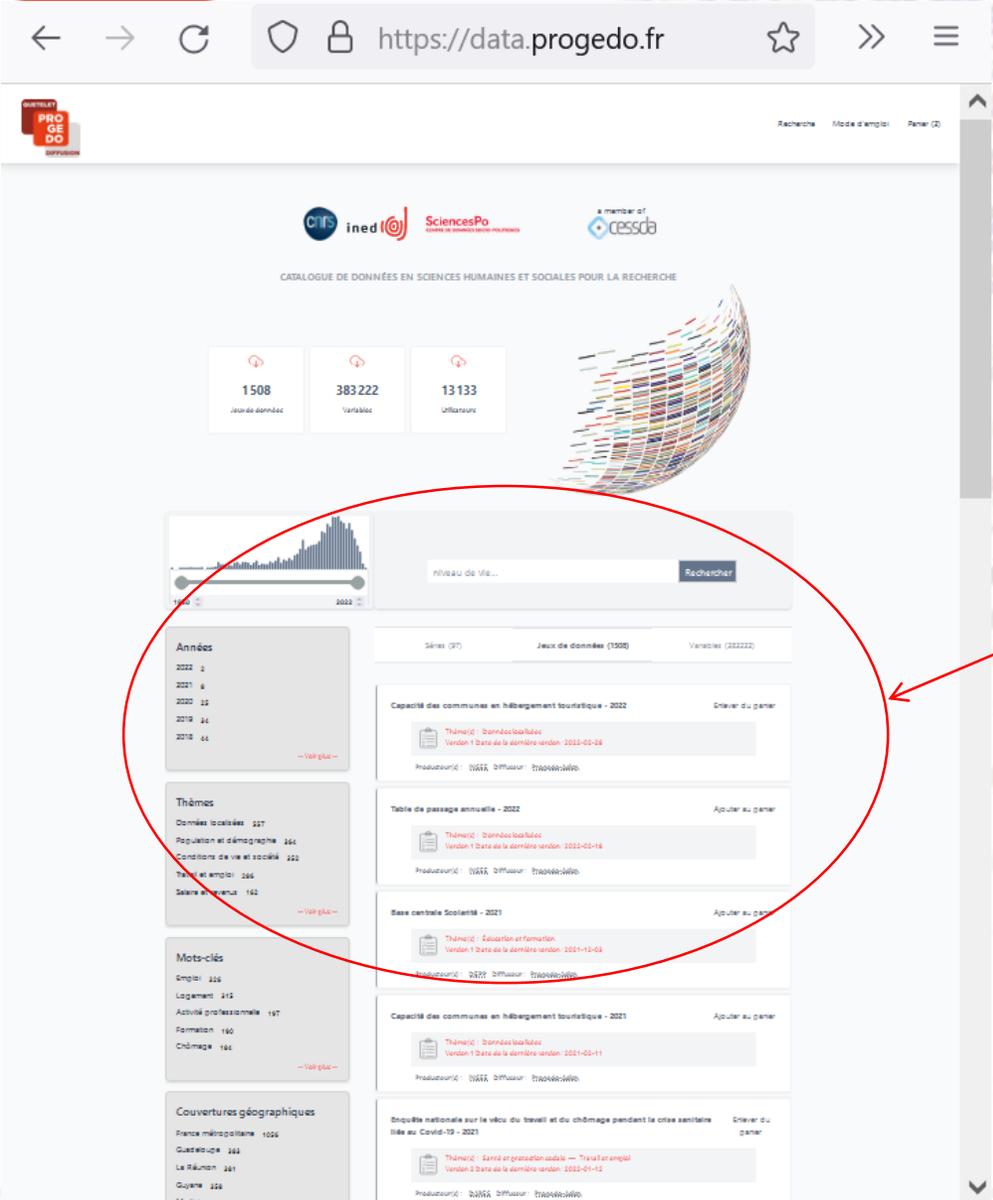
Activités langagières de la mère et du père avec l'enfant à 1 an (11035 enfants)



**Les données sur les langues
accessibles via le portail
Quetelet-Progedo-Diffusion**

Trouver des données sur les langues dans le catalogue Quetelet-Progedo-Diffusion

- En recherchant par mot clé et en affinant par des filtres
- Les résultats sont présentés sous la forme de séries, de jeux de données ou de variables



This block shows a detailed view of search results for the keyword 'langue'. At the top, there's a search bar with 'langue' and a 'Rechercher' button. Below it, a timeline slider shows the years 1950 to 2022. The results are categorized into 'Séries (25)', 'Jeux de données (146)', and 'Variables (1457)'. A list of years is shown: 2021 (3), 2020 (3), 2019 (5), 2018 (4), 2017 (5). Below this, there are sections for 'Thèmes' and 'Mots-clés'. The 'Thèmes' section lists: Éducation et formation (59), Conditions de vie et société (37), Travail et emploi (22), Santé et protection sociale (11), and Population et démographie (9). The 'Mots-clés' section lists: Enseignement supérieur (34). On the right, there are three data series cards: 'Cosmopolitisme culturel chez les jeunes - 2015', 'Enquête permanente sur les conditions de vie des ménages, partie variable : éducation et famille - octobre 2003', and 'Enquête Etude de l'histoire familiale - 1999'. Each card includes a 'Ajouter au panier' button and details about the theme and version.

L'enquête EHF – Étude de l'Histoire Familiale (1999)



JEU DE DONNÉES

Identifiants :
Ili-0173
doi:10.13144/lii-0173

Thème :
Population et démographie

Séries : Famille

Couverture géographique :
France métropolitaine

Producteurs :
INSEE
Ined

Diffuseur :
Progedo-Adisp

Enquête Etude de l'histoire familiale - 1999 Ajouter

Version 2 : ajout d'une table contenant un jeu de pondération pour les études au niveau « famille ». Elle permet également disposer de deux variables sur le type de famille en 1999 correspondant à deux approches différentes pour définir la re familiale. L'une de ces approches permet d'obtenir les évolutions observées à méthodologie constante entre 1999 et 20 (TYPFAM_COMPARABLE2011). La seconde se base sur les seules données disponibles en 1999 (variable TYPFAM), date :

Résumé Détails Variables **Données** Documentation Accès

Nom du fichier ▼

Adu_mdif

Contenu du fichier

Le niveau « adultes EHF » : l'unité statistique est la personne interrogée lors de l'enquête "Etude de l'Histoire Familiale" (EHF). Les variables sont issues du questionnaire : "Etude de l'histoire familiale" (auto-administré), elles décrivent un ensemble de caractéristiques décrivant la vie professionnelle de l'individu, les étapes de la vie familiale, ainsi que la transmission familiale des langues et des parlers pour l'ensemble des individus interrogés à l'enquête EHF. 380 481 adultes ont été décrits.

Nom du fichier ▼

EHFD0022

Contenu du fichier

Le niveau « données brutes » : l'unité statistique est la personne interrogée lors de l'enquête "Etude de l'Histoire Familiale" (EHF). Les variables sont issues des différents questionnaires : "Etude de l'histoire familiale" (auto-administré), elles décrivent un ensemble de caractéristiques décrivant la vie professionnelle de

- Autrefois nommée « **Enquête Famille** »
« L'Enquête Famille » associée au recensement depuis 1954 consacre à chaque édition une batterie de questions à l'étude d'un thème particulier
- En 1999, le thème est **la transmission des langues et parlers d'une génération à l'autre**
- **380 000 individus de 18 ans et plus** sont interrogés sur leur histoire familiale : leurs origines, leurs enfants, leurs beaux-enfants (enfants du conjoint), leurs périodes de vie en couple et leur parcours social, ainsi que sur les **leurs langues d'usage (nationales et régionales) au sein de leur famille.**
- Enquête représentative au niveau régional
- Le fichier FPR diffusé par Progedo-Adisp comprend des données du recensement

Questionnaire EHF, volet Langues

TRANSMISSION FAMILIALE DES LANGUES ET DES PARLERS

19 EN QUELLES LANGUES, DIALECTES OU «PATOIS», VOS PARENTS VOUS PARLAIENT-ILS D'HABITUDE QUAND VOUS ÉTIEZ ENFANT, VERS L'ÂGE DE 5 ANS ?

VOTRE PÈRE ou l'homme qui vous a élevé

1 - vous parlait d'habitude en...

D19PL1

2 - et aussi en...

D19PL2

Exemples : ALSACIEN, BASQUE, BRETON, CATALAN, CORSE, CRÉOLE, FLAMAND, GALLO, OCCITAN, PICARD, PLATT, PROVENÇAL, ARABE, ESPAGNOL, KABYLE, PORTUGAIS, LANGUE DES SIGNES (LSF)...

- Pour le FRANÇAIS, notez simplement « F »
- En cas de dialecte ou «patois», précisez de quel endroit (PICARDIE, BÉARN, ROUERGUE, MOSELLE...)
- Pour les langues étrangères, ne notez pas la nationalité mais la langue.
Exemples : ne dites pas Algérien, Marocain, Sénégalais, mais ARABE, KABYLE, WOLOF, etc.

VOTRE MÈRE ou la femme qui vous a élevé

1 - vous parlait d'habitude en...

D19ML1

2 - et aussi en...

D19ML2

20 EN QUELLES LANGUES, DIALECTES OU «PATOIS», PARLIEZ-VOUS À VOS JEUNES ENFANTS, QUAND ILS AVAIENT 5 ANS (ou leur parlez-vous maintenant s'ils sont plus jeunes) ?

C20NOENF

Si vous n'avez pas eu d'enfant, cochez cette case :

1 - Vous leur parliez d'habitude en...

D20JL1

2 - Et aussi en...

D20JL2

21 ET ACTUELLEMENT, VOUS ARRIVE-T-IL DE DISCUTER AVEC DES PROCHES (conjointe, parents, amis, collègues, commerçants...) DANS D'AUTRES LANGUES QUE LE FRANÇAIS ?

OUI 1

NON 2

Si oui : laquelle ou lesquelles ?

1. D21DCU1

2. D21DCU2

C21DSCU

22 SOUHAITEZ-VOUS RETOURNER VIVRE UN JOUR DANS LA RÉGION DE VOTRE ENFANCE (que ce soit en France ou ailleurs) ?

C22RT

1
OUI

2
NON

3
Vous ne savez pas

4
Vous y vivez déjà

Proc format ;

value \$fethl

"aav" = "Langue ou dialecte allemand s. i. (autre que l'allemand standard)"

"aax" = "Langue ou dialecte d'Angola s.i."

"aay" = "Berrichon- bourbonnais"

"aba" = "Abbey"

"abb" = "Abo, Abou, Aboo"

"abi" = "Abidji"

"abu" = "Aboure, Abure"

"abv" = "Arabe, marocain, tunisien..."

"acf" = "Bete"

"aci" = "Cari"

"adj" = "Adioukrou, adjukru"

"adl" = "Andalou, Patois du sud de l'Espagne"

"afk" = "Afrikaans"

"afr" = "Afar"

"aii" = "Assyrien"

"aji" = "Ajie"

"aju" = "Judéo-arabe, judeo-tunisien"

"akw" = "Akwa"

"aln" = "Albanais, kosovar"

"amh" = "Amharique"

"any" = "Agni"

"arm" = "Arménien"

"ati" = "Attié"

"aub" = "Bable, Asturien"

"auv" = "Auvergnat"

"axx" = "Aragonais"

"azb" = "Azerbaïdjanais, Azeri"

"baa" = "Bassa, Mbele"

"bag" = "Sanaga"

"ban" = "Dschang"

L'Enquête Cosmopolitisme culturel chez les jeunes (2015)

- Enquête réalisée pour le ministère de la Culture et de la Communication
- Objectif de l'enquête : mieux comprendre les rapports des jeunes à la culture dans le contexte de la globalisation
- Thèmes abordés portant sur les langues :
 - les consommations médiatiques (films, musique, lecture, sites web, réseaux sociaux etc.)
 - les compétences linguistiques (langues d'usage, langues d'origine du jeune et de sa famille proche)
- Enquête réalisée auprès d'un échantillon représentatif de jeunes de 18 à 29 ans résidant en France métropolitaine (1 665 jeunes interrogés)

4- Regardez-vous des films étrangers ? Nous parlons bien ici de l'origine du film, quelle que soit la langue dans laquelle vous regardez le film. Un film étranger est un film réalisé par un réalisateur étranger, que le film ait été vu en langue française ou étrangère.

Enquêteur : ne pas citer

- 1- Oui
- 2- Non

Filtre : si Q4=1

5- Quelle est la part de films étrangers que vous regardez par rapport aux films français ?

Enquêteur : Montrez la carte n°3 - une seule réponse possible

- 1- Principalement des films français
- 2- Autant de films français qu'étrangers
- 3- Principalement des films étrangers
- 4- Uniquement des films étrangers
- 5- [Vous ne savez pas]

Filtre : si Q4=1

6 Quand vous regardez des films étrangers, les regardez-vous principalement en VO ou doublés ? Regarder un film étranger en VO signifie donc que vous le regardez dans la langue d'origine du film, avec ou sans sous-titre.

Enquêteur : Montrez la carte n°4

- 1 Doublé en français exclusivement
- 2 Doublé en français majoritairement
- 3 Doublé en français et en VO à égalité
- 4 En VO majoritairement
- 5 En VO exclusivement
- 6 **Le français est la langue originale (pays francophones)**
- 7 [Vous ne savez pas]

Filtre : si Q6=2 ou 3 ou 4 ou 5 ou 6

7 Lorsque vous regardez des films dans une ou des autres langues que le Français, de quelle(s) langue(s) s'agit-il? (liste précodée)