

IV. Des données inégalement produites à l'échelle du territoire national

Tout d'abord, il est nécessaire de comprendre que les résultats du recensement, et par conséquent l'analyse qui peut en être faite, sont extrêmement différents en fonction des langues et des territoires concernés. Afin de mieux comprendre la situation, l'analyse va être séparée en trois parties : les COM, les DOM et les langues régionales ou d'immigration en métropole.

1. Les Collectivités des Outre-mer (COM)

Contrairement aux DOM, les Collectivités des Outre-mer peuvent, par dérogation, procéder elles-mêmes à des recensements généraux de leur population tous les cinq ans à la place de l'INSEE¹¹. L'organisation est alors de la responsabilité des instituts statistiques locaux de ces territoires. Il en existe trois à ce jour : l'Institut de la Statistique de la Polynésie Française (ISPF), l'Institut de la Statistique et des Etudes Economiques de Nouvelle-Calédonie (ISEE), et le Service Territorial de la Statistique et des Etudes Economiques du territoire des îles Wallis et Futuna (STSEE).

Ces organismes publient de nombreuses études, et incluent tous les trois des questions linguistiques au recensement de la population. Les données sont également téléchargeables en libre-accès par leurs sites (il ne s'agit pas des données brutes, mais de données déjà retravaillées et croisées avec certaines variables)¹².

Un très bon exemple en la matière est l'ISPF : lors du recensement de la Polynésie française, tous les cinq ans, trois questions linguistiques sont incluses :

- Langue parlée en famille (question ouverte)
- Niveau dans une langue polynésienne (question nivelée)
- Niveau en français (question nivelée)

La première question, en particulier, permet d'avoir un large panorama des langues parlées sur le territoire et de leur évolution, à la fois pour les langues polynésiennes (tahitien, marquisien, paumotu, mangarévien, langues australes), mais aussi pour les langues issues de l'immigration (par exemple, le hakka, les langues chinoises et l'anglais qui sont très présents), et enfin pour les langues issues des autres territoires français ultramarins, tels que les langues kanak ou le wallisien et futunien.

¹¹ Loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité, art. 157 (II).

¹² Voir <https://www.ispf.pf/chiffres>

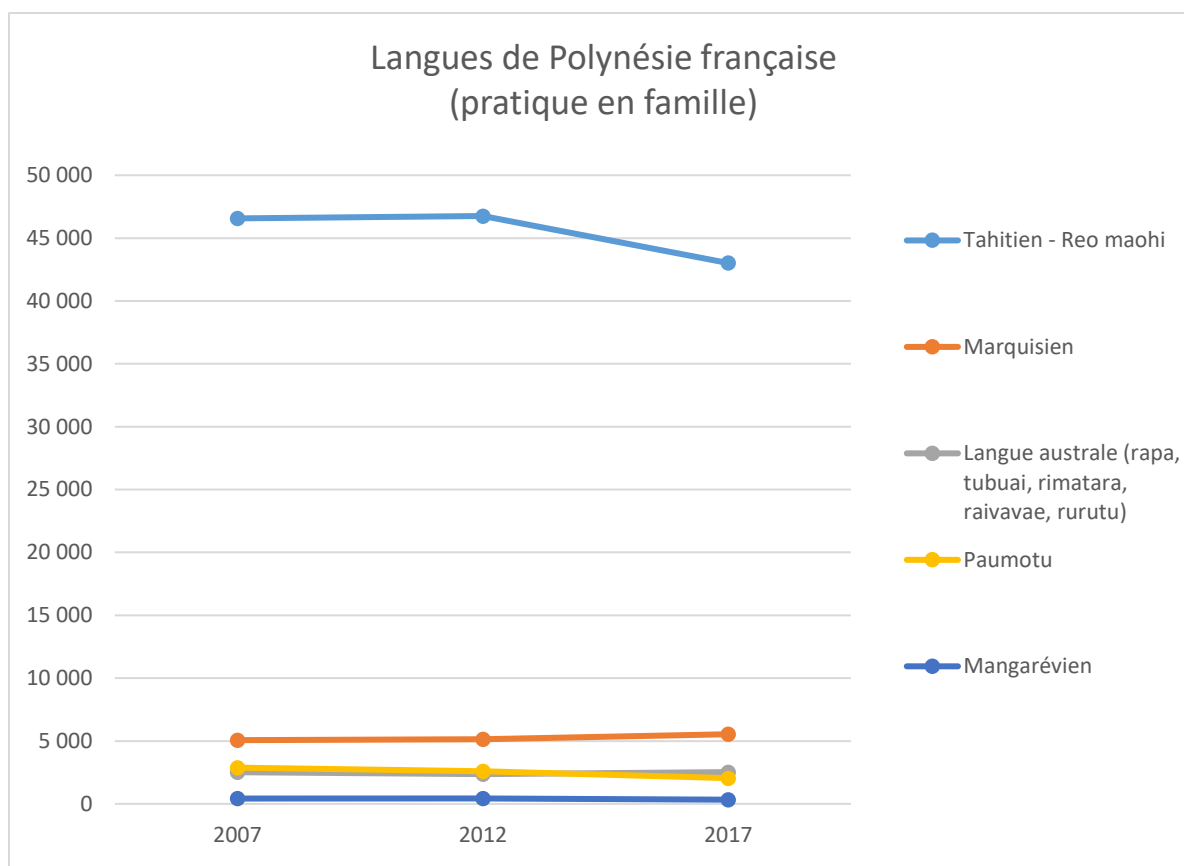


Figure 1. Visualisation construite avec Excel à partir des données ISPF des recensements 2007, 2012 et 2017 de Polynésie Française

La situation est plus disparate concernant les autres COM. Le recensement de Wallis-et-Futuna est effectué tous les 5 ans par le Service Territorial de la Statistique et des Etudes Economiques du Territoire des îles Wallis et Futuna (STSEE)¹³. Il inclut plusieurs questions linguistiques dans le bulletin individuel : la maîtrise du français et la maîtrise d'une « langue locale » (divisés chacun en trois catégories « lire », « parler », « écrire »), la langue la plus couramment parlée en famille (champ libre), et la maîtrise de l'anglais (question fermée)¹⁴.

Sur le site du STSEE, il est possible d'accéder et télécharger les données retravaillées des recensements de 2013 et 2018¹⁵. Sur la liste des variables disponibles, on trouve « Population maîtrisant le français » et « Population maîtrisant une langue polynésienne », ce qui présuppose qu'un travail de regroupement des données a été effectué avec les différentes questions linguistiques.

¹³ URL : <https://www.statistique.wf/recensements-et-enquetes/recensement-general-de-la-population/>

¹⁴ Pour voir le bulletin individuel : https://www.statistique.wf/download/45/recensement-2013/1052/bulletin-individuel_wf-2013-specimen.pdf

¹⁵ <https://wf.poggis.spc.int/#c=home> pour les données quantitatives (libre-accès), pour les autres documents voir : <https://www.statistique.wf/recensements-et-enquetes/recensement-general-de-la-population/>

Wallis-et-Futuna étant un territoire avec un très faible taux d'immigration (correspondant essentiellement à des métropolitains de passage temporaire), on peut supposer que « langue locale » et « langue polynésienne » dans les deux cas correspondent exclusivement au wallisien et futunien.

Par ailleurs, la comparaison entre les deux années est rendue difficile par le fait que ces données linguistiques n'ont pas été regroupées de la même façon malgré des questionnaires identiques : en effet, le recensement de 2013 a séparé les deux variables « Population maîtrisant le français » et « Population maîtrisant une langue polynésienne » en 3 catégories : oui, non et « non-communicé ». Les chiffres de 2013 montrent que le pourcentage de « non communiqué » représentait entre 24,5% et 29,9% des réponses en fonction des districts, soit un total de 3 156 réponses sur le territoire.

En 2018, cette troisième catégorie a été supprimée, et seules les réponses « oui » et « non » ont été prises en compte. En conséquence, le pourcentage de locuteurs de langues polynésiennes parmi les réponses prises en compte a drastiquement augmenté : sur le district de Mua, par exemple, il passe de 70,8 % (2013) à 94,7 % (2018). En revanche, le nombre de locuteurs déclarés au niveau individuel est relativement proche entre ces deux années, et si on regarde la proportion de réponse « oui » en comparaison à la population totale recensée, les proportions sont similaires entre ces deux années. Les chiffres sont donc à manipuler avec précaution car, même si non comptabilisé, le nombre de non-réponse reste important sur l'année 2018. De ce fait, on peut également supposer que le nombre de locuteurs du wallisien et futunien est très largement sous-déclaré par les habitants et sous-estimé pour ces deux années.

2013

District	Oui	Non	Non-communicé	Population totale
Alo	1564	46	546	2156
Sigave	977	45	435	1457
Hahake	2284	309	936	3529
Hihifo	1393	123	493	2009
Mua	2158	142	746	3046
Total	8376	665	3156	12197

2018

District	Oui	Non		Population totale
Alo	1469	59		1950
Sigave	930	38		1275
Hahake	2347	281		3415
Hihifo	1396	140		1942
Mua	2112	119		2976
Total	8254	637		11558

Figure 2. Tableau simplifié de la variable « Population maîtrisant une langue polynésienne » des recensements de la population de Wallis et Futuna en 2013 et 2018, comparé à la population totale recensée ces deux années¹⁶

¹⁶ Données téléchargées sur PopGIS <https://wf.poggis.spc.int/>

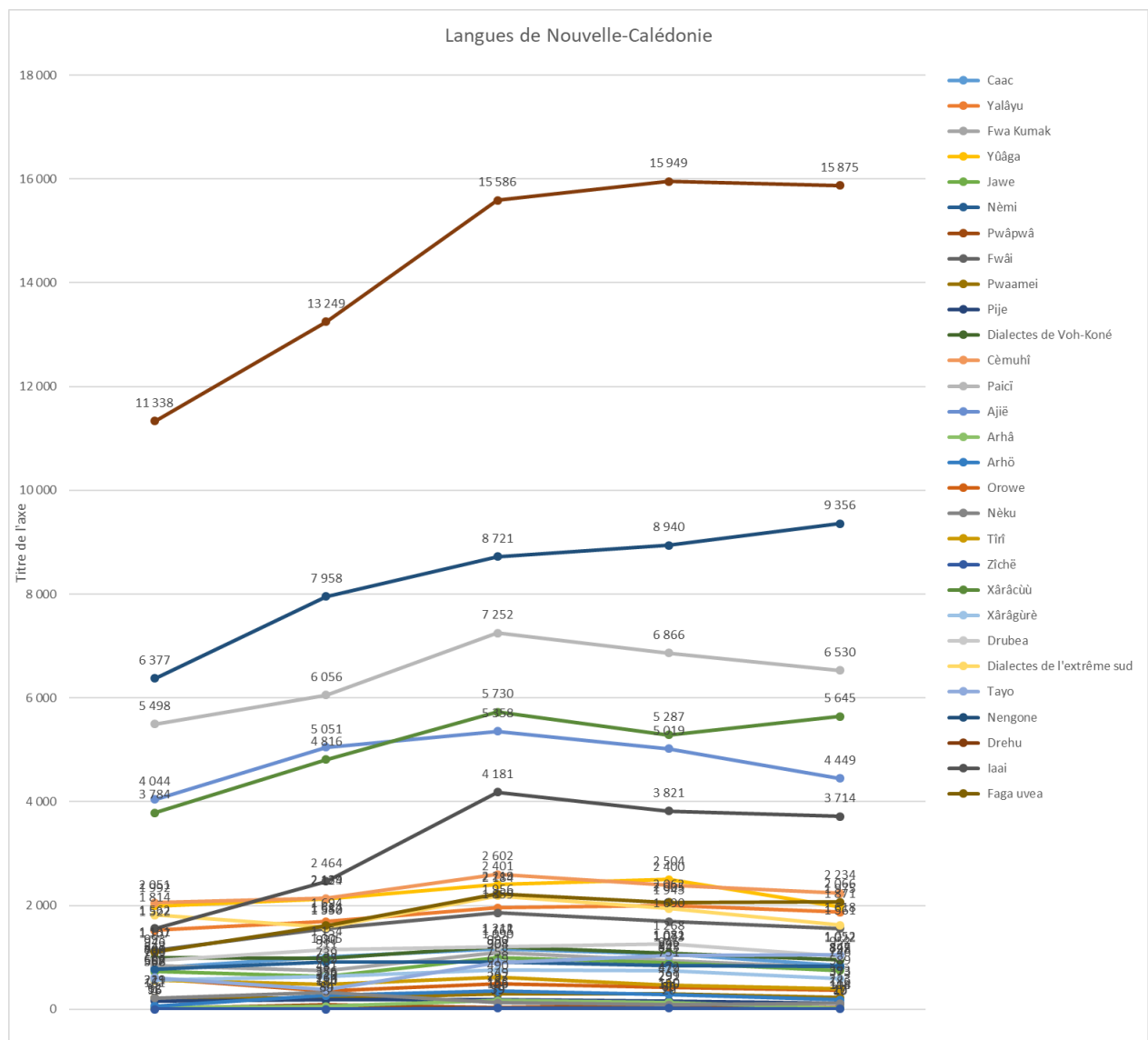
Sur le territoire de Nouvelle-Calédonie, le recensement est de la responsabilité de l'Institut de la Statistique et des Etudes Economiques de la Nouvelle-Calédonie (ISEE)¹⁷. Depuis 1996, celui-ci compote des questions de langue, et plus précisément sur les pratiques des langues « vernaculaires¹⁸ » (28 langues kanak et le créole local, le tayo). Cela permet un aperçu précis dans le temps de l'évolution du nombre de locuteurs pour chaque langue.

Avec des échantillonnages de locuteurs compris entre 53 566 et 75 853 en fonction des années, sur un total de population kanak déclarée de 111 860 personnes en 2019¹⁹, ces recensements donnent un exemple sans précédent de collecte des données linguistiques de façon constante et régulière sur un laps de temps d'une vingtaine d'années.

¹⁷ URL : <https://www.isee.nc/population/recensement>

¹⁸ Il s'agit du terme utilisé dans l'enquête.

¹⁹ ISEE, *Synthèse N° 45 - Recensement de la population 2019 - Nouvelle-Calédonie*.
<https://www.isee.nc/component/phocadownload/category/230-toutes-les-publications?download=1982:rp-2019-la-croissance-demographique-flechit-nettement-en-nc-entre-2014-et-2019>



Néanmoins, dans le cas de la Nouvelle-Calédonie, la collecte des données linguistiques est concentrée sur la pratique des langues locales, et contrairement à la Polynésie française, ne comporte pas d'informations sur les langues issues des flux migratoires et inter-régions ultramarines.

Toutes les Collectivités des Outre-mer cependant ne possèdent pas d'instituts statistiques dédiés à leurs territoires. Les COM des Caraïbes – Saint Barthélemy, St Pierre et Miquelon, St Martin – sont donc sous la tutelle de l'INSEE, qui effectue leur recensement tous les ans²⁰.

Tout comme les recensements en métropole, les recensements de ces trois COM ne comportent pas de questions linguistiques. Les autres enquêtes de l'INSEE et des différents acteurs de la statistique publique ne couvrant pas ces territoires, il y a, en conséquence, un véritable vide en terme de données

²⁰ Loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité, art. 156

linguistiques sur ces trois collectivités. Il s'agit également, à l'issue du travail présenté ici, des seuls territoires de France sur lesquels il existe une absence totale de données quantitatives sur les langues.

Malgré la présence de données linguistiques dans les COM du Pacifique, on peut noter néanmoins l'absence d'informations concernant le sujet des pratiques linguistiques en langues des signes. Il n'y a, en effet, à ce jour, aucune donnée précise sur les modes de communication des populations sourdes sur ces territoires. Dans le recensement de la Polynésie française, une catégorie « sourd et muet » est indiqué parmi la liste des langues parlées en famille (elle représente 357 interrogés en 2012) ; or, c'est un problème méthodologique, car cette catégorie ne correspond pas à une langue et, à l'inverse, renvoie à l'idée que ces populations n'ont aucune langue du fait de leur surdité. Par ailleurs, elle ne dit rien sur les pratiques linguistiques réelles de ces personnes (on ne sait pas si elles utilisent une langue des signes, une variation de la LSF, l'écriture en français ou dans une autre langue, etc.). Sans développement de la recherche et de la collecte d'informations, il est impossible d'en savoir d'avantage sur les pratiques linguistiques des populations sourdes des territoires ultramarins en l'état actuel des choses.

2. Les Départements des Outre-Mer

En France, l'INSEE s'occupe des recensements et conduit des études statistiques relatives à la métropole et aux Départements des Outre-mer (Guadeloupe, Martinique, Réunion, Guyane et Mayotte, ce dernier étant un DOM depuis 2011). Cependant, l'Enquête Famille de 1999 ne couvrait que l'Hexagone. La situation des études linguistiques dans les DOM est paradoxale, car il y a à la fois un historique de prise en compte des questions linguistiques dans les enquêtes, mais parfois aussi des difficultés d'ordre méthodologique, liées par exemple à la pluralité des glossonymes en usage, ce qui peut affecter l'exploitation des données.

Parmi les enquêtes de la statistique publique, on compte principalement deux séries traitant des questions linguistiques dans les départements des Outre-mer : les enquêtes IVQ et MFV.

La première enquête IVQ (Information et Vie Quotidienne) a été lancée en France métropolitaine en 2004. Elle a ensuite été étendue aux départements de la Martinique (2006), Réunion (2007) et Guadeloupe (2008). Ces enquêtes comportaient plusieurs questions linguistiques (langue d'apprentissage de la lecture, langue parlée à la maison, langue parlée vers 5 ans) ; cependant elles n'étaient pas complètement satisfaisantes, car elles avaient tendance à rassembler tous les créoles dans une catégorie unique, sans rendre compte des variations (en particulier, les créoles locaux et ceux issus des migrations). En 2011, l'enquête IVQ a été renouvelée en France métropolitaine, puis étendue à nouveau aux départements de la Réunion (2011) et de la Martinique (2014), et enfin ont été ajoutés la Guyane française et Mayotte (2011). Sur cette deuxième série, les différents types de créole à base lexicale française avaient été précisé dans les questions (créole réunionnais, créole guyanais, etc.). Dans le cas de Mayotte et de la Guyane, les langues locales et les principales langues issues des migrations étaient mentionnées dans le questionnaire (par exemple pour la Guyane : « créole haïtien », « sranan tongo », « portugais-brésilien »).

La deuxième série d'enquêtes comportant des données sur les langues dans les DOM est Migrations-Famille-Vieillesse (MFV)²¹. Réalisée conjointement par l'INED et l'INSEE, la première enquête est

²¹ Informations et résultats des différentes enquêtes sur le site : <https://mfv2.site.ined.fr/fr/trois-enquete-mfv/mfv-1-et-mfv-mayotte/>

sortie en 2009-2010, et concernait les quatre DOM de l'époque (Guadeloupe, Guyane française, Réunion, Martinique). En 2015-2016, MFV a été étendue à Mayotte (devenue un DOM en 2011). Enfin, en 2020-2021, une deuxième version de l'enquête MFV a été lancée, cette fois concernant les cinq DOM, et est toujours en cours de traitement actuellement²². Plusieurs parties des questionnaires de MFV 1 sont dédiées à la pratique du français et du créole. Cependant, « le créole » est utilisé ici comme une entité générale, sans prendre compte des différents types de créole et des spécificités linguistiques de chaque département (la Guyane, en particulier, possède de très nombreuses langues locales en plus du français et du créole guyanais). Le questionnaire de MFV Mayotte est construit différemment, les questions étant ouvertes et centrées sur la transmission et la pratique des langues locales (shimaoré et kibushi) ainsi que des langues de migration. La deuxième version de MFV lancée en 2020-2021 reprend cependant les mêmes catégories que MFV 1 concernant les langues.

En dehors de ces deux séries, quelques enquêtes ponctuelles concernent des DOM mais rarement l'ensemble de ceux-ci. Dans le monde associatif, Lofis La Lang Kreol a réalisé en 2007 et 2021 des enquêtes sur la pratique du créole sur le territoire de la Réunion²³. Dans le monde de la recherche, un certain nombre de publications nous apportent des indications sur les pratiques linguistiques en Guyane française : du fait de sa taille et de la difficulté à accéder à certaines parties du territoire, il est difficile d'avoir un aperçu exhaustif de la situation linguistique guyanaise par les enquêtes publiques. La chercheuse en sociolinguistique Isabelle Léglise, en particulier, a effectué une série d'enquêtes entre 2000 et 2012, auprès d'enfants scolarisés en primaire de différentes villes et villages de l'ensemble de la Guyane, atteignant un échantillonnage total d'environ 2 300 enfants interrogés²⁴. Ces travaux ont fait l'objet de diverses publications et sont utilisés comme références. Plusieurs thèses permettent de compléter ce travail par d'autres enquêtes sociolinguistiques ciblées²⁵.

Un problème récurrent dans les enquêtes concernant les DOM est le choix des glossonymes. Pour illustrer ce point, on peut prendre comme exemple les différents noms utilisés pour désigner le kibushi, une des deux langues locales mahoraises avec le shimaoré. Langue malgache, elle est désignée par ses locuteurs comme « kibushi », ce qui correspond à l'endonyme. En revanche, en shimaoré, cette langue est désignée comme « shibushi », ce qui constitue un exonyme. En dehors de Mayotte, la désignation de la langue varie, les deux premiers termes sont parfois utilisés, mais elle est également désignée par le terme « malgache de Mayotte », ce qui est un autre exonyme utilisé par un autre groupe de personnes. Les trois termes de kibushi, shibushi et malgache de Mayotte sont tous des glossonymes.

Lors d'une enquête, il est attendu que les personnes interrogées utilisent différents glossonymes pour une même langue et l'analyse des données nécessite donc le plus souvent d'effectuer un travail de correspondance entre ces termes.

²² <https://www.insee.fr/fr/statistiques/6676232>

²³ Pour celle de 2021 : <https://lofislalangkreollarenyon.re/sondage/#sondage/1/>

²⁴ Les résultats détaillés de ces enquêtes sont accessibles sur le site : <https://leglise.cnrs.fr/>

²⁵ Par exemple, on peut citer Isabelle Martin, qui a effectué pour sa thèse une enquête macrosociolinguistique à Saint Georges de l'Oyapock auprès des familles Parikwene (2022). <https://theses.hal.science/tel-03632500/document>

En revanche, il est également possible que, au niveau des enquêteurs, la conception des questionnaires ne soit pas suffisamment précise en termes de glossonymes et affecte donc potentiellement l'exploitation des résultats, comme nous allons le voir dans les exemples suivants.

En 2007, Mayotte est encore une Collectivité des Outre-mer. La même année, l'INSEE effectue le recensement de l'île²⁶, et y inclut des questions linguistiques : langues parlées et langues lues ou écrites. Cependant, dans la présentation des résultats, plusieurs problèmes apparaissent : dans la question des langues lues ou écrites, il y a deux catégories : « lit le français » et « lit l'arabe », cette dernière étant précisée par une astérisque comme « *lit l'arabe = arabe et/ou shimaoré* ». Or, l'utilisation du terme arabe ici peut prêter à confusion, car il fait référence à l'utilisation de l'alphabet arabe (ou arabe écrit) et non à la pratique de la langue : le shimaoré, qui est une langue bantoue apparentée au swahili, utilise traditionnellement l'alphabet arabe pour être écrite ; les enfants apprennent à déchiffrer l'arabe au sein des écoles coraniques, sans pour autant maîtriser nécessairement la langue arabe en elle-même.

Dans l'enquête IVQ 2011 à Mayotte, le kibushi est indiqué comme « malgache ». Or, le kibushi peut être considéré comme un dialecte du malgache, mais l'assimilation des deux empêche par exemple de différencier les populations mahoraises parlant la langue locale (qui, de plus, a deux variantes locales, le kibushi kisakalava et le kibushi kiantalautsi) des populations d'origine étrangère parlant le malgache de Madagascar.

Enfin, un autre exemple de difficultés posées par la pluralité des glossonymes peut se trouver dans les enquêtes MFV sur le territoire guyanais : dans les séries MFV, les questionnaires sont harmonisés sur tous les territoires hormis Mayotte, et posent tous un ensemble de question sur la pratique du « créole », sans distinction à l'intérieur de cette catégorie. Or, on sait que dans certains DOM une partie de la population pratique des créoles qui peuvent différer du créole majoritaire local, du fait notamment des mouvements migratoires (par exemple, le créole haïtien en Guadeloupe).

Mais la situation est encore plus complexe dans un territoire comme la Guyane, où cohabitent localement créoles à base française et créoles à base anglaise. Dans la première catégorie, on trouve le créole guyanais, mais aussi le créole haïtien, dont la présence est particulièrement marquée du fait des mouvements migratoires²⁷. La seconde catégorie recouvre principalement les langues dites « businenge », des langues créoles à base lexicale anglaise ou anglo-portugaise, issues des mouvements de marronnage des anciens esclaves africains emmenés au Suriname au XVII^e et XVIII^e siècle. Une partie des descendants de ces communautés se sont installés en Guyane française, en particulier au cours du XIX^e siècle. Les langues businenge sont aujourd'hui considérées comme des langues régionales de Guyane et sont inscrites sur la liste des langues de France²⁸. Ces langues ne sont communément pas désignées par le terme « créole » au sein de la population, mais plutôt par leurs noms spécifiques. Néanmoins, en l'absence d'indications, ce que recouvre le terme « créole » dans l'enquête va donc dépendre de ce que les enquêteurs et enquêtés définissent par celui-

²⁶ <https://www.insee.fr/fr/statistiques/2569525>

²⁷ Le créole haïtien est parlé par plus de 12% des enfants interrogés à l'échelle de la Guyane dans les enquêtes réalisées par Isabelle Léglise. C'est également la langue première de 7% des enfants.

²⁸ Les langues businenge inscrites sur la liste des langues de France sont : le saamaka, et les trois variantes du nenge(e) tongo : aluku, ndyuka, pamaka.

ci, et laisse donc place à différentes interprétations. Cet exemple illustre un manque de précision au niveau de la conception des questionnaires, qui peut amener à des résultats erronés ou imprécis.

Par ailleurs, les langues amérindiennes de Guyane²⁹ sont absentes du questionnaire des enquêtes MFV. Rarement prises en compte dans les enquêtes, on peut les trouver mentionnées dans le questionnaire de l'enquête IVQ Guyane (2011), mais elles ne sont pas comptabilisées parmi les résultats (possiblement à cause du manque de représentativité à grande échelle car ce sont des communautés linguistiques de petite taille). Les données quantitatives sur ces langues ne sont pas issues de la statistique publique, mais plutôt des recherches individuelles ou locales : les enquêtes d'Isabelle Léglise³⁰ permettent notamment de connaître la place de plusieurs langues amérindiennes au sein des 28 localités qu'elle a étudiées. Il existe également quelques thèses ou travaux de recherche concernant plus spécifiquement le parikwaki et les pratiques linguistiques à Saint Georges-de-l'Oyapock³¹. Enfin, le site Cartes des langues³² (réalisé par le Vice-Rectorat de Guyane) indique le nombre d'élèves concernés par le dispositif ILM par langue, ce qui permet d'avoir un aperçu partiel des proportions de locuteurs : sur l'année 2021-2022, par exemple, 548 élèves de maternelle et primaire sont comptabilisés pour les ILMs en langues amérindiennes, distribué de la façon suivante : kali'na (198), wayana (152), parikwaki (106), wayampi (58), teko (34).

Pour finir, une dernière enquête illustre les difficultés méthodologiques au niveau de la conception des enquêtes sur les DOM : en 2003 et 2014, les enquêtes Formation, qualification professionnelle (FQP) sont lancées sur le territoire de la France métropolitaine. Ces deux enquêtes incluent des questions linguistiques (langue maternelle et langue parlée dans le foyer vers l'âge de 5 ans), posées sous forme de QCM à 14 choix : le français, plusieurs langues étrangères fréquentes sur le territoire européen, une catégorie intitulée « Langue régionale » et une catégorie « Autre ».

En 2014, cette enquête est étendue à la Réunion et à la Guadeloupe³³, avec un échantillonnage de 5 000 personnes. Les questionnaires ne sont pas modifiés par rapport à leur version métropolitaine. En conséquence, sur la question de la langue parlée vers l'âge de 5 ans, on observe donc que la moitié des fréquences de réponses (sans critère de pondération et de représentativité) concerne le français, tandis que l'autre moitié est partagée entre « autre » et « langue régionale », en l'absence de catégorie qui représente les langues locales de ces territoires et en particulier les créoles réunionnais et guadeloupéen (les autres catégories n'ont aucune ou pratiquement aucune réponse). Ce dernier exemple illustre les problèmes fréquents de conception des questionnaires au niveau des enquêteurs, en particulier le manque de prise en compte des spécificités des territoires des DOM et les problèmes d'interprétation que cela peut provoquer. En l'absence de précision au niveau de l'enquêteur, l'accès

²⁹ Les langues amérindiennes de Guyane indiquées sur la liste des Langues de France sont : kali'na, wayana, parikwaki, lokono, wayampi, et teko.

³⁰ <https://leglise.cnrs.fr/>

³¹ Par exemple, la thèse d'Isabelle Martin (2022). <https://theses.hal.science/tel-03632500/document>

³² <https://langues-de-guyane.ins.ac-guyane.fr/-Carte-des-langues-.html>

³³ <https://data.progedo.fr/studies/doi/10.13144/lil-1523>

à l'information devient compliqué et n'apporte pas de nouvelle connaissance sur les pratiques linguistiques.

3. Langues régionales et langues d'immigration en métropole

Concernant le territoire métropolitain, une distinction doit être faite entre deux catégories : les « langues de France » (langues historiquement présente sur le territoire métropolitain, d'après la liste de Bernard Cerquiglini de 1999) et les langues issues des flux migratoires.

Au niveau régional, les enquêtes quantitatives sur les langues locales sont généralement commanditées par des collectivités territoriales et par des associations ou avec l'appui de celles-ci.

Parmi les langues régionales de France métropolitaine, il existe tout d'abord une division entre les langues faisant l'objet régulièrement d'études quantitatives et les langues pour lesquelles les données se révèlent inexistantes ou presque. D'après le travail de recensement, les langues régionales ayant fait l'objet d'une ou plusieurs études quantitatives concernant la pratique et/ou le nombre de locuteurs depuis 1999 sont les suivantes :

Langues	Exemples de commanditaire d'enquêtes quantitatives
Basque	Office Public pour la Langue Basque, Gouvernement Basque espagnol
Breton	Office Public de la Langue Bretonne, Conseil régional de Bretagne, Association Bretagne culture diversité, Départements, Ouest-France (journal)
Catalan	Institut Franco-catalan Transfrontalier, Conseil Départemental des Pyrénées-Orientales, Gouvernement Catalan espagnol
Corse	Collectivité territoriale de Corse
Francoprovençal	Région Rhône-Alpes
Alsacien	Office pour la Langue et la Culture d'Alsace, Collectivité européenne d'Alsace, Dernières Nouvelles d'Alsace (journal)
Occitan	Office Public de la Langue Occitane, Conseil régional d'Aquitaine, Conseil régional Midi-Pyrénées, Institut d'Etudes Occitanes, Interreg POCTEFA, Région Rhône-Alpes, Conseil départemental des Pyrénées-Atlantiques
Gallo	Association Bretagne culture diversité, Office Public de la Langue Bretonne, Conseil régional de Bretagne

On observe qu'une écrasante majorité des enquêtes quantitatives régionales concernent la partie Ouest (breton, gallo) et Sud-Ouest (occitan, basque, catalan) de la France, ainsi que l'alsacien dans l'Est de la France, et le corse dans le sud. Les langues d'oïl (à l'exception du gallo), le flamand occidental, et les parlers liguriens n'ont fait l'objet d'aucune étude quantitative identifiée lors du recensement. Le francoprovençal est un cas particulier, car il a fait l'objet d'une seule enquête non-renouvelée³⁴, qui concernait également l'occitan.

Par ailleurs, à l'intérieur même de ces enquêtes, il existe également des disparités parmi les langues et territoires étudiés. Ainsi, les collectivités territoriales et associations de Bretagne produisent un nombre considérable d'études sur la langue bretonne (et, par extension du gallo, du fait de la présence de cette langue d'oïl en Haute-Bretagne) : 10 séries d'enquêtes ont été recensées sur la période, avec la production de nombreux rapports analysant la situation linguistique au sein des différents pays bretons. Cette production est sans équivalent dans le reste du territoire.

Un autre exemple de disparité des territoires concerne l'occitan-langue d'oc. Les territoires historiques de cette langue ne sont pas tous représentés par les enquêtes, qui concernent principalement les territoires des régions Nouvelle-Aquitaine et Occitanie. Une seule enquête concerne l'ancienne région Auvergne, une concerne l'ancienne région Rhône-Alpes, et aucune enquête n'a été réalisée sur le provençal.

Les enquêtes sur le basque et le catalan ont la particularité d'être commanditées ou adossées aux enquêtes des communautés autonomes espagnoles du Pays Basque et de Catalogne. En particulier, le gouvernement basque réalise tous les cinq ans une grande enquête sociolinguistique sur la pratique de la langue, en incluant le pays basque français. Cette configuration est spécifique à ces deux langues et ne se retrouve pas dans le cas des autres langues régionales transfrontalières.

Enfin, on observe une grande multiplicité des acteurs, représentant les différentes strates du mille-feuille administratif français (région, département, collectivités territoriales), mais aussi les acteurs extraterritoriaux (Espagne, Collectivité européenne d'Alsace, INTERREG³⁵). Enfin, le monde associatif des langues locales est très représenté parmi ces enquêtes ; les associations sur les langues régionales sont souvent commanditaires ou associées à des enquêtes quantitatives.

Concernant les langues non-territoriales de la liste des Langues de France (arabe dialectal, arménien occidental, berbère, etc.), il n'existe aucune étude spécifique à ces langues, hormis dans le cas particulier de la Langue des Signes Française (voir partie V).

Langues issues des flux migratoires

Les grandes enquêtes de la statistique publique à l'échelle nationale métropolitaine telles que l'Enquête Famille de 1999 ou l'Enquête IVQ de 2011 concernent toutes les langues potentiellement mentionnées par les interrogés, et donc également les langues issues des flux migratoires. Ces enquêtes ont été plus ou moins mobilisées dans le domaine de la recherche. Alexandra Filhon analyse

³⁴ Enquête FORA (francoprovençal et occitan) en Région Rhône-Alpes (2009)

³⁵ INTERREG est un programme de l'Union Européenne visant à promouvoir la coopération entre les régions européennes sur des sujets tels que la gouvernance, la mobilité, le développement urbain, etc. Par exemple, INTERREG Poctefa concerne la zone Espagne – Andorre – France. Voir <https://www.interregeurope.eu/>

par exemple une partie des résultats du volet linguistique de l'Enquête Famille 1999 concernant les langues d'immigration, dans le cadre du livre *Histoires de Familles, Histoires familiales*³⁶. Concernant plus spécifiquement les primo-arrivants, l'enquête ELIPA (Enquête Longitudinale sur l'Intégration des Primo-Arrivants)³⁷, réalisée en 2010 et 2019, pose également des questions de langue, mais l'étude est limitée aux ressortissants étrangers majeurs possédant un titre de séjour (de plus, la deuxième version d'ELIPA de 2019 ne concerne même que les étrangers hors de l'Union Européenne et des quatre pays de l'AELE).

En dehors de ces exemples, la principale enquête centrée spécifiquement sur les langues d'immigration en France métropolitaine et unique enquête régionale à ce sujet est ELIRA, « Etude Langues de l'Immigration en Rhône-Alpes (2014). Commanditée par le Conseil régional de l'ancienne région Rhône-Alpes, elle se base sur l'exploitation de 303 questionnaires en ligne et diffusés auprès de d'acteurs variés (associations, consulats, services culturels, etc.). L'étude présentait 4 objectifs :

- *Apporter des connaissances sur les langues pratiquées en Rhône-Alpes par les populations immigrées*
- *Déterminer si certaines méritent une attention particulière de la part des pouvoirs publics*
- *Dresser un état des lieux des formes d'action associatives et institutionnelles en faveur des langues de l'immigration dans le domaine de la transmission et de la valorisation*
- *Répertorier les politiques territoriales de soutien aux actions de valorisation et de transmission des langues de l'immigration*³⁸

Les langues considérées comme langues de références au moment de la préparation de l'enquête ont été les suivantes : arabes littéraire et dialectaux, kabyle/berbère, linguala, wolof, comorien, turc, kurde, arménien occidental, vietnamien, mandarin, portugais, espagnol, langues d'Italie (général), polonais, yiddish, et rromani (le hmong, grec, « BCMS » bosniaque-croate-monténégrin-serbe et russe avaient initialement été retenus mais ont été abandonnés par la suite du fait du manque d'interlocuteurs).

L'une des conclusions apportées par l'enquête ELIRA sont que, sur la zone étudiée, au moins une centaine de langues ont été répertoriées. De plus, « *une immigration n'est donc pas à associer à une seule langue en fonction de sa citoyenneté (c'est le cas notamment du berbère, du fé-fé, du kurde, du peul, rromani ou des dialectes italiens). De plus, une part importante des immigrés est plurilingue avant d'arriver en France.* ». Il existe donc une diversité linguistique bien plus importante qu'estimée à première vue, y compris sur un échantillon limité, et d'autant plus que les populations issues de l'immigration ont une forte propension au plurilinguisme (21% des 303 interrogés du questionnaire déclarent parler 3 langues ou plus³⁹).

³⁶ Institut national d'études démographiques (INED), *Histoires de familles, histoires familiales: les résultats de l'enquête Famille de 1999*, Les Cahiers de l'INED, No. 156, 2005.

³⁷ Commanditée par le Département des Statistiques, des Études et de la Documentation (DESD) du Ministère de l'Intérieur

³⁸ Rapport ELIRA Phase 1, Introduction Partie 2

³⁹ Annexe 12 du Rapport ELIRA – Phase 1.

Test PISA

Le test PISA permet de combler partiellement ce manque d'information en donnant une estimation au niveau national du nombre d'élèves parlant une autre langue maternelle que le français. Le test PISA (*Program for International Student Assessment*, Programme international pour le suivi des acquis des élèves) est un programme de l'OCDE. Effectué tous les 3 ans auprès des élèves de 15 ans des 34 pays membre de l'OCDE et des pays partenaires, a pour but d'évaluer les niveaux en compréhension de l'écrit, sciences et mathématiques des élèves à partir de tests standardisés. Ce test a la particularité d'inclure également un questionnaire de contexte sur l'environnement familial des élèves et leurs habitudes d'apprentissage. Parmi ces questions, il est notamment demandé le pays de naissance de l'élève et la langue parlée dans le foyer. Le but est de déterminer combien d'élèves passant le test PISA ont une langue maternelle différente de la langue d'éducation du pays (dans laquelle est passé le test) ; autrement dit, les élèves « non-natifs » dans la langue officielle (« *non-native-speaking students* »), afin d'évaluer si ce facteur a une influence sur les résultats. Les statistiques concernant la France ont été extraits du document *The Resilience of Students with an Immigrant Background : Factors that Shape Well-being*, publié en 2018 par l'OCDE⁴⁰ :

	Students with an immigrant background (1)		Immigrant students (2)		Immigrant students with at least one native-born parent (3)		Students who speak a language at home different from the one in which they sat the PISA assessment (4)	
	%	Erreur-type	%	Erreur-type	%	Erreur-type	%	Erreur-type
France	26,6	1,31	13,18	0,97	13,08	0,52	8,48	0,64

(1) Elèves d'origine immigrée⁴¹, (2) Elèves immigrés⁴², (3) Elèves immigrés dont au moins un des parents est natif du pays, (4) Elèves qui parlent à la maison une langue différente de celle dans laquelle ils ont passé le test PISA (i.e. langue d'enseignement du pays)

Figure 4. Extrait du tableau « Percentage of students with an immigrant background and non-native-speaking students » [*Pourcentage d'élèves d'origine immigrée et des élèves non-natifs de la langue d'enseignement*], résultats du test PISA 2018, réalisé en France métropolitaine et Outre-mer

⁴⁰ OCDE (2018), « Graphique 5.2 - Percentage of students with an immigrant background and non-native-speaking students », dans *The Resilience of Students with an Immigrant Background : Factors that Shape Well-being*, OECD Reviews of Migrant Education, Éditions OCDE, Paris, <https://doi.org/10.1787/9789264292093-graph50-en>.

⁴¹ Immigrés de 2^{ème} génération (élève né en France dont les parents sont nés à l'étranger)

⁴² Immigrés de 1^{ère} génération (nés à l'étranger)

Ainsi, les résultats de PISA 2018 nous apprennent que 8,48% des élèves ayant passé le test en France ont une langue maternelle autre que le français.

Ce chiffre n'est pas nécessairement corrélé à la question migratoire : le pourcentage d'élèves eux-mêmes « immigrés » (c'est-à-dire nés dans un autre pays que la France) est de 13,18%, et le pourcentage d'élèves « d'origine immigrée » (« *students with an immigrant background* », c'est-à-dire un élève né en France de parents nés à l'étrangers, ou immigré de 2^{ème} génération) est de 26,26%. Autrement dit, une grande partie des élèves immigrés ou de parents immigrés ont pour langue maternelle le français.

En se basant sur les données du test PISA 2018, le rapport Eurydice⁴³ 2023 a croisé les données entre langue maternelle et migration, afin de déterminer dans quelle mesure les langues d'immigration sont concernées ⁴⁴ :

	Immigrant students who mainly speak a different language at home from the language of schooling (1)		Immigrant students who mainly speak the language of schooling at home (2)		Non-immigrant students who mainly speak a different language at home from the language of schooling (3)		Non-immigrant students who mainly speak the language of schooling at home (4)	
	%	Erreur-type	%	Erreur-type	%	Erreur-type	%	Erreur-type
France	6,5	0,49	7,8	0,66	5,0	0,39	80,8	1,02

- (1) Elèves immigrés parlant principalement à la maison une langue différente de la langue d'enseignement
- (2) Elèves immigrés parlant principalement à la maison la langue d'enseignement
- (3) Elèves non-immigrés parlant principalement à la maison une langue différente de la langue d'enseignement
- (4) Elèves non-immigrés parlant principalement à la maison la langue d'enseignement

Figure 5. Extrait du tableau « Percentages of immigrant and non-immigrant 15-year-old students, by language spoken at home » [*Pourcentage des élèves de 15 ans immigrés et non-immigrés, par langue parlée à la maison*], résultats du test PISA 2018, réalisé en France métropolitaine et Outre-mer

Ainsi, on constate que seuls 6,5% des élèves ayant passé le test en France sont des élèves « immigrés » avec une langue maternelle différente du français. A l'inverse, les élèves immigrés parlant principalement le français à la maison correspondent à 7,8% du total des élèves. Selon la notice du tableau, le terme d'« élève immigré » désigne ici les élèves nés à l'étranger ou ceux nés en France dont

⁴³ Eurydice est un réseau européen sur les systèmes éducatifs, piloté par la Commission Européenne.

⁴⁴ Chiffres basés sur le test PISA 2018, in : European Education and Culture Executive Agency, Eurydice, « Table A3: Percentages of immigrant and non-immigrant 15-year-old students, by language spoken at home (2018) », Table des annexes *Key data on teaching languages at school in Europe – 2023 edition*, Publications Office of the European Union, 2023, <https://eurydice.eacea.ec.europa.eu/sites/default/files/2023-03/Annex%20I%20Statistical%20tables.xlsx>

les parents sont nés à l'étranger (immigrés de 1^{ère} et 2^{ème} génération, les deux catégories étant combinées)⁴⁵.

Par ailleurs, le rapport Eurydice indique également que dans ces chiffres, les dialectes d'une langue officielle sont considérés comme assimilés à la langue officielle dans la plupart des pays où les informations ont été collectées⁴⁶. Avec ces seules informations, il est impossible de savoir si les langues régionales de France ont été considérées comme assimilées au français ou comme langues à part entière, et cela dépend donc du traitement des données qui a été fait à posteriori. Du point de vue linguistique, seules les langues dites d'« oïl » sont considérées comme des dialectes du français (par opposition aux langues comme par exemple l'occitan ou le breton, qui sont des langues séparées du français), mais elles sont néanmoins toutes intégrées à la liste des langues de France du rapport Cerquiglini.

De cette façon, il est impossible de savoir si les 5% d'élèves non-immigrés ayant une autre langue maternelle que le français sont des élèves parlant une langue régionale, ou bien des immigrés de 3^{ème} génération, ou encore d'autres configurations spécifiques telles que les élèves de nationalité française des territoires des Outre-mer parlant une langue maternelle différente du français, les élèves dont la langue maternelle est la LSF, etc.

De façon générale, les données du test PISA permettent d'avoir une estimation du nombre d'élèves dont la langue maternelle n'est pas le français, mais ne permettent pas d'avoir d'avantage d'informations sur les langues concernées à l'intérieur de ce pourcentage. Peu d'enquêtes travaillant par ailleurs sur les données quantitatives liées aux langues d'immigration, il est donc difficile de pouvoir estimer quelles sont les langues concernées.

⁴⁵ Dans la note explicative du tableau : « *The category 'immigrant students' corresponds to students whose parents were born abroad. This category merges two distinct PISA categories of immigrant students, namely (1) the student and both parents were born abroad (i.e. first-generation immigrants); and (2) the student was born in the country of the test, but both parents were born abroad (i.e. second-generation immigrants).* »

⁴⁶ Dans la note explicative du tableau : « *Speaking a dialect of a certain language at home is considered speaking the standard language. This approach has been used in most countries participating in the PISA survey. As the approach has not yet been applied to the German-speaking Community of Belgium and Italy, dialects have been recoded accordingly.* »