

Étude sur la situation des langues parlées au Québec 2024

Rapport méthodologique



Pour tout renseignement concernant l'Institut de la statistique du Québec (ISQ) et les données statistiques dont il dispose, s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy, 3^e étage
Québec (Québec) G1R 5T4

Téléphone :
418 691-2401
1 800 463-4090 (Canada et États-Unis)

Site Web : statistique.quebec.ca

Ce document est disponible seulement en version électronique.

Dépôt légal
Bibliothèque et Archives nationales du Québec
2^e trimestre 2025
ISBN 978-2-555-01037-6 (en ligne)

© Gouvernement du Québec, Institut de la statistique du Québec, 2025

Toute reproduction autre qu'à des fins de consultation personnelle est interdite sans l'autorisation du gouvernement du Québec.
statistique.quebec.ca/fr/institut/nous-joindre/droits-auteur-permission-reproduction

Mai 2025

Publication réalisée à l'Institut de la statistique du Québec par :	Vanessa Djapa, Luc Côté, Baptiste Beck et Marie-Eve Tremblay
Avec la collaboration de :	Mbuyi Kelelekela et Leila Nombo
Sous la direction de :	Éric Gagnon
Révision linguistique et édition :	Direction de la diffusion et des communications
Comité de lecture interne :	Christine Routhier, Paul Berthiaume et Patricia Caris
Enquête sous la responsabilité de :	Direction des statistiques sociodémographiques Institut de la statistique du Québec
Enquête financée par :	Ministère de la Langue française
Pour tout renseignement concernant le contenu de cette publication :	Direction de la méthodologie Institut de la statistique du Québec 200, chemin Sainte-Foy, 3 ^e étage Québec (Québec) G1R 5T4 Téléphone : 418 691-2401 1 800 463-4090 (Canada et États-Unis) Site Web : statistique.quebec.ca

Notice bibliographique suggérée

DJAPA, Vanessa, et autres (2025). *Étude sur la situation des langues parlées au Québec 2024. Méthodologie de l'enquête*, [En ligne], Québec, Institut de la statistique du Québec, 29 p. [statistique.quebec.ca/fr/fichier/eslpq2024-methodologie.pdf].

Avertissement

Les proportions estimées contenues dans le présent rapport sont arrondies à une décimale dans les tableaux et à l'unité dans le texte. Exceptionnellement, dans le texte, les proportions inférieures à 5 % et celles qui ne sont pas reportées dans les tableaux sont présentées avec une décimale. Les proportions dont la décimale est ,5 sont arrondies à l'unité inférieure ou supérieure selon la seconde décimale. En raison de l'arrondissement, la somme des proportions peut être légèrement supérieure ou inférieure à 100 %.

Table des matières

Introduction	6
1 Plan d'échantillonnage	7
1.1 Population visée	7
1.2 Base de sondage	7
1.3 Méthode de sélection, taille et répartition de l'échantillon	8
2 Collecte de données	12
2.1 Prétest	12
2.2 Méthodes de collecte	12
2.3 Résultats de la collecte	14
3 Traitement des données	15
3.1 Validation et saisie	15
3.2 Pondération	15
3.3 Non-réponse totale	17
3.4 Non-réponse partielle	17
4 Analyse des données, précision des estimations et tests statistiques	18
4.1 Source et construction des indicateurs sur les langues le plus fréquemment utilisées dans différentes situations	18
4.1.1 Question sur la fréquence d'utilisation des langues dans différentes situations	18
4.1.2 Question sur la connaissance des langues (intitulé, choix de réponses et consignes)	19
4.1.3 Construction de l'indicateur des langues le plus fréquemment utilisées	19
4.2 Indicateur de niveau de revenu du ménage	21
4.3 Précision des estimations et tests statistiques	22

5	Présentation des résultats	23
6	Portée et limites de l'enquête	24
	Annexe 1 – Caractéristiques de la population visée	26
	Annexe 2 – Imputation du revenu du ménage	27
	Références bibliographiques	28

Introduction

Afin de faire une utilisation adéquate des données et des résultats issus de l'*Étude sur la situation des langues parlées au Québec* (ESLPQ 2024)¹, il importe de connaître la méthodologie d'enquête utilisée. En effet, le plan de sondage, les procédures de collecte et le traitement des données sont tous des éléments qui ont une incidence sur les résultats d'une enquête. La connaissance des aspects méthodologiques aide à interpréter adéquatement les résultats et à en apprécier la qualité, la portée et les limites.

L'*Étude sur la situation des langues parlées au Québec* (ESLPQ) est une enquête populationnelle d'envergure qui vise à recueillir des données sur l'utilisation des langues dans différents contextes par la population québécoise de 15 ans et plus. C'est une première édition pour cette enquête qui pourrait être réalisée à nouveau.

Objectifs de l'ESLPQ

L'ESLPQ est réalisée à la demande du ministère de la Langue française et vise à fournir des statistiques précises et fiables sur l'utilisation des langues par l'ensemble de la population du Québec âgée de 15 ans et plus ainsi que par les populations de certains territoires du Québec en particulier.

Le but de l'enquête est d'obtenir un portrait relatif aux langues qu'utilisent les individus dans divers contextes de leur vie. Certains des contextes examinés impliquent des conversations avec d'autres personnes (p. ex. des membres de la maisonnée, des amis ou amies et des commerçants ou commerçantes) et d'autres contextes impliquent plutôt la consultation, par l'individu, d'un type de contenu langagier donné (p. ex. naviguer sur Internet, écouter des émissions d'information et écouter des chansons). Par ailleurs, certains contextes relèvent de la sphère de la vie privée et d'autres, de la sphère publique (p. ex. l'usage des langues dans les commerces ou au travail).

Plus précisément, l'ESLPQ cherche à connaître la proportion de la population de 15 ans et plus qui utilise telle ou telle langue ou combinaison de langues dans un contexte donné. L'enquête vise la

production de tels résultats non seulement pour l'ensemble de la population et en fonction de certains territoires de résidence, mais aussi en fonction d'autres paramètres comme le groupe d'âge des personnes ou leur lieu de naissance.

Structure du rapport méthodologique

Les six sections du présent document portent sur les principaux éléments de la méthodologie de l'ESLPQ. La section 1 présente le plan d'échantillonnage de l'enquête, alors que la section 2 décrit les stratégies utilisées et les résultats obtenus lors de la collecte des données. La section 3 porte sur le traitement des données : le processus de validation, la méthode de pondération utilisée afin que les résultats puissent être inférés à la population visée, de même que l'examen de l'ampleur de la non-réponse partielle. La section 4 traite des méthodes utilisées pour l'analyse des données, y compris la création de certains indicateurs, l'estimation de la précision et les tests statistiques. Les normes de présentation des résultats sont précisées à la section 5, alors qu'un aperçu de la portée et des limites de l'enquête est offert à la section 6.

1. Notons que le terme « étude » est utilisé dans le titre du projet et les produits de diffusion pour assurer l'uniformité avec le questionnaire de l'ESLPQ. Toutefois, pour le rapport méthodologique, le terme « enquête » est majoritairement utilisé pour désigner l'étude.

1 Plan d'échantillonnage

Cette section comprend une description de la population visée et de la base de sondage, ainsi que toutes les informations utiles sur la sélection de l'échantillon à partir de cette base.

1.1 Population visée

La population visée par l'ESLPQ 2024 correspond à l'ensemble des personnes de 15 ans et plus vivant dans un logement non institutionnel au Québec². Elle inclut donc les personnes qui ne comprennent pas le français ni l'anglais³. Les personnes vivant dans un logement collectif institutionnel (hôpital, centre d'hébergement de soins de longue durée, établissement pénitentiaire, centre de réadaptation, etc.) sont exclues, ainsi que celles résidant dans les régions sociosanitaires du Nunavik (17) et des Terres-Cries-de-la-Baie-James (18)⁴. On estime la population visée par l'ESLPQ 2024 à environ 7 427 160 personnes, soit environ 99,2 % de la population québécoise âgée de 15 ans et plus. Les principales caractéristiques sociodémographiques de ces personnes sont présentées à l'annexe 1.

1.2 Base de sondage

La base de sondage utilisée pour sélectionner l'échantillon de l'enquête a été élaborée à partir des données du Fichier d'inscription des personnes assurées (FIPA) de la Régie de l'assurance maladie du Québec (RAMQ). Ce fichier présente plusieurs avantages : il permet notamment l'exclusion de la majeure partie des personnes non visées par l'enquête tout en assurant une excellente couverture de la population cible. Il permet également la sélection d'individus en fonction de l'âge, du sexe⁵ et du lieu de résidence. Finalement, il contient les variables nécessaires pour communiquer avec les individus sélectionnés.

Les principaux inconvénients du FIPA découlent de la qualité de la mise à jour des numéros de téléphone, qui sont dans certains cas inexacts et parfois même absents, ce qui complique du même coup la collecte de données. Cette situation est plus fréquente chez les jeunes de 18 à 24 ans. L'utilisation d'autres sources administratives, notamment celle du ministère de l'Éducation du Québec, a permis de pallier cette lacune en conformité avec la *Loi sur l'Institut de la statistique du Québec*.

2. Cela inclut les personnes qui vivent dans un ménage privé et celles qui vivent dans certains logements collectifs non institutionnels, tels que les résidences pour aînés et les résidences de religieuses ou de religieux. Selon le recensement (Statistique Canada), un ménage collectif est constitué d'une personne ou d'un groupe de personnes occupant un logement collectif et n'ayant pas de domicile habituel ailleurs au Canada.
3. Les personnes qui déclarent ne pas être en mesure de soutenir une conversation en français ou en anglais représentaient 1,0 % de la population du Québec selon le Recensement de 2021 (Institut de la statistique du Québec, Tableau [Répartition de la population selon la connaissance du français et de l'anglais et selon la langue maternelle, régions métropolitaines de recensement \[RMR\] du Québec, 2011, 2016 et 2021](#)).
4. Les personnes résidant dans les réserves indiennes situées dans le reste du Québec font partie de la population visée par l'enquête.
5. Pour les personnes transgenres ayant obtenu un changement de la mention du sexe figurant à leur acte de naissance auprès du Directeur de l'état civil, le sexe indiqué dans le FIPA est celui indiqué dans l'acte de naissance après le changement.

Par ailleurs, la base de sondage présente une légère sous-couverture, puisque certaines personnes admissibles à l'enquête ne sont pas inscrites au régime québécois d'assurance maladie (p. ex. les personnes n'ayant pas renouvelé leur carte de la RAMQ) ou n'y sont pas admissibles (p. ex. certains groupes de résidents non permanents) et ne figurent donc pas au FIPA. De même, la base de sondage pourrait compter quelques personnes inadmissibles à l'enquête, par exemple celles n'ayant pas encore informé la RAMQ d'un déménagement à l'extérieur du territoire visé par l'enquête. Il est toutefois difficile de quantifier de manière précise la couverture nette de la base de sondage, étant donné que la taille de la population visée ne peut pas être établie de façon exacte. En comparant les effectifs de la base de sondage aux plus récentes estimations de population produites par l'ISQ (qui englobent l'ensemble des personnes résidant au Québec), ajustées pour tenir compte de l'exclusion des personnes en institution de la base de sondage, on peut établir que la couverture de la population visée par le FIPA est d'au moins 94 %.

1.3 Méthode de sélection, taille et répartition de l'échantillon

L'un des objectifs principaux de l'ESLPQ est de fournir des estimations fiables sur les langues utilisées dans divers contextes de vie à l'échelle de l'ensemble du Québec et des territoires géographiques visés, à savoir : région métropolitaine de recensement (RMR) de Montréal subdivisée en région administrative (RA) de Montréal et reste de la RMR de Montréal, RA de la Capitale-Nationale, municipalité de Gatineau, et ailleurs au Québec. La taille d'échantillon totale a été fixée à 75 048 personnes, pour un nombre de personnes répondantes attendu d'environ 44 580 personnes qui est calculé sur la base d'un taux de réponse visé de 60 % et un taux d'admissibilité de 99 % sur le plan provincial. Le tableau 1.1 présente la taille d'échantillon totale ainsi que le nombre de personnes répondantes attendu par territoire géographique.

Tableau 1.1
Taille d'échantillon et nombre attendu de personnes répondantes par territoire géographique, Québec, 2024

	Taille d'échantillon	Nombre attendu de personnes répondantes
	n	
RMR de Montréal	46 535	27 642
RA de Montréal	29 129	17 303
Reste de la RMR de Montréal	17 406	10 339
RA de la Capitale-Nationale	6 910	4 105
Municipalité de Gatineau	4 885	2 902
Ailleurs au Québec	16 718	9 930
Ensemble du Québec	75 048	44 579

Source : Institut de la statistique du Québec, *Étude sur la situation des langues parlées au Québec 2024*.

Cette taille d'échantillon permet d'obtenir, pour les indicateurs principaux de l'enquête, des estimations avec une bonne précision pour l'ensemble du Québec et pour chaque territoire géographique (voir tableaux 1.2 et 1.3). Par exemple, elle permettra d'estimer la proportion de personnes ayant le français comme langue le plus souvent parlée à la maison (seule ou en combinaison avec une autre langue) avec une marge d'erreur de 0,4 % à l'échelle du Québec et une marge d'erreur d'au plus

1,0 % à l'échelle des territoires géographiques visés par l'ESLPQ, sauf pour la municipalité de Gatineau (marge d'erreur de 1,6 %).

De plus, avec le même plan de sondage, il sera possible de détecter avec une bonne puissance statistique des écarts ténus dans les indicateurs principaux entre deux éditions de l'enquête. Le tableau 1.4 donne un aperçu des proportions attendues et des écarts pouvant être

Tableau 1.2

Nombre de personnes répondantes, proportion et marge d'erreur attendues selon le territoire pour le français comme langue parlée le plus souvent à la maison, Québec, 2024

	Le français comme langue parlée le plus souvent à la maison		
	Nombre de personnes répondantes attendues	Proportion estimée attendue ¹	Marge d'erreur attendue
	n	%	
RMR de Montréal	27 642	68,5	0,6
RA de Montréal	17 303	54,1	0,8
RA de la Capitale-Nationale	4 105	96,4	0,6
Municipalité de Gatineau	2 902	77,0	1,6
Ailleurs au Québec	9 930	93,9	0,5
Ensemble du Québec	44 579	80,8	0,4

1. Statistique Canada, [Recensement de la population](#). Adaptation par l'Institut de la statistique du Québec.

Source : Institut de la statistique du Québec, *Étude sur la situation des langues parlées au Québec 2024*.

Tableau 1.3

Nombre de personnes répondantes, proportion et marge d'erreur attendues selon le territoire pour le français comme langue utilisée le plus souvent au travail, Québec, 2024

	Le français comme langue utilisée le plus souvent au travail		
	Nombre de personnes répondantes attendues	Proportion estimée attendue ¹	Marge d'erreur attendue
	n	%	
RMR de Montréal	16 792	77,8	0,7
RA de Montréal	10 184	66,9	1,0
RA de la Capitale-Nationale	2 483	97,1	0,7
Municipalité de Gatineau	1 754	65,4	2,3
Ailleurs au Québec	5 647	94,2	0,6
Ensemble du Québec	26 676	85,0	0,5

1. Statistique Canada, [Recensement de la population](#). Adaptation par l'Institut de la statistique du Québec.

Source : Institut de la statistique du Québec, *Étude sur la situation des langues parlées au Québec 2024*.

détectés avec une bonne puissance statistique⁶ (80 % ou plus) selon le territoire géographique pour deux indicateurs de langues⁷. Pour une caractéristique mesurée sur l'ensemble de la population visée, comme la proportion de personnes ayant le français comme langue le plus souvent parlée à la maison, l'enquête permet de détecter, avec une bonne puissance, tout écart entre deux éditions de 1,0 point de pourcentage et plus au niveau de l'ensemble du Québec et de 1,6 point de pourcentage et plus à l'échelle des territoires, sauf pour la municipalité

de Gatineau, avec un écart détectable de 3,5 points de pourcentage et plus. Pour la proportion de personnes utilisant le plus fréquemment le français comme langue d'usage au travail, l'enquête permet de détecter, avec une bonne puissance, tout écart entre deux éditions de 1,1 point de pourcentage et plus au niveau de l'ensemble du Québec et de 2,0 points de pourcentage et plus à l'échelle des territoires, sauf pour la municipalité de Gatineau, avec un écart détectable de 4,9 points de pourcentage et plus.

Tableau 1.4

Proportions attendues et écarts pouvant être détectés lors de la prochaine édition pour le français comme langue le plus souvent parlée à la maison et utilisée au travail, selon le territoire géographique, Québec, 2024

	Le français comme langue le plus souvent :			
	Parlée à la maison		Utilisée au travail	
	Proportion attendue en 2024 ¹	Écart détectable avec une puissance de 80 % ou plus	Proportion attendue en 2024 ¹	Écart détectable avec une puissance de 80 % ou plus
	%	points de %	%	points de %
RMR de Montréal	68,5	≥ 1,3	68,5	≥ 1,5
RA de Montréal	54,1	≥ 1,6	54,1	≥ 2,0
RA de la Capitale-Nationale	96,4	≥ 1,4	96,4	≥ 1,7
Municipalité de Gatineau	77,0	≥ 3,5	77,0	≥ 4,9
Ailleurs au Québec	93,9	≥ 1,1	93,9	≥ 1,4
Ensemble du Québec	80,8	≥ 0,9	80,8	≥ 1,1

1. Statistique Canada, [Recensement de la population](#). Adaptation par l'Institut de la statistique du Québec.

Source : Institut de la statistique du Québec, *Étude sur la situation des langues parlées au Québec 2024*.

6. Par exemple, pour la RA de Montréal, supposons que la proportion estimée du français comme langue le plus fréquemment utilisée au travail est d'environ 67 % en 2024. Si cette proportion a varié de 2,0 % ou plus entre 2024 et 2025, l'enquête permettra de détecter un tel écart avec une probabilité de 80 % ou plus. Autrement dit, si l'enquête était répétée un grand nombre de fois, notamment avec des échantillons différents, on conclurait à un écart significatif au moins 8 fois sur 10. En revanche, si le changement dans la proportion du français comme langue le plus fréquemment utilisée au travail dans la RA de Montréal est plus ténu que 2,0 %, l'enquête offre une puissance statistique inférieure à 80 %, c'est-à-dire qu'il y a plus de risque qu'une modification si subtile passe inaperçue.

7. Tiré du Recensement de 2021, le premier indicateur porte sur la langue le plus souvent parlée à la maison. Cet indicateur est basé sur deux questions : « Quelle(s) langue(s) cette personne parle-t-elle régulièrement à la maison ? » Dans le cas où cette personne mentionnait plus d'une langue, la question suivante était posée : « Parmi ces langues, laquelle cette personne parle-t-elle le plus souvent à la maison ? Indiquez plus d'une langue seulement si elles sont parlées aussi souvent l'une que l'autre à la maison. ».

Le deuxième indicateur tiré du recensement est la langue le plus souvent utilisée au travail. Ce dernier indicateur concerne la sous-population des travailleurs et travailleuses définie par les réponses aux questions suivantes lors du Recensement de 2021 : « Pendant la semaine du dimanche 2 mai au samedi 8 mai 2021, combien d'heures cette personne a-t-elle travaillé à un emploi salarié ou à son compte ? » et « Pendant la semaine du 2 mai au 8 mai 2021, cette personne était-elle mise à pied temporairement ou absente de son emploi ou de son entreprise ? ». L'indicateur de la langue le plus souvent utilisée au travail est basé sur deux questions : « Dans cet emploi, quelle(s) langue(s) cette personne utilisait-elle régulièrement ? ». Dans le cas où cette personne mentionnait plus d'une langue, la question suivante était posée : « Parmi ces langues, laquelle cette personne utilisait-elle le plus souvent dans cet emploi ? Indiquez plus d'une langue seulement si elles étaient utilisées aussi souvent l'une que l'autre au travail ».

Afin de pouvoir détecter des écarts aussi ténus entre les éditions à l'échelle territoriale, la RA de Montréal et la municipalité de Gatineau ont été suréchantillonnées, c'est-à-dire que la part relative de chacun des deux territoires est plus importante dans l'échantillon que dans la population visée.

Étant donné la taille d'échantillon relativement élevée, la sélection de l'échantillon s'est faite en trois vagues selon la répartition suivante : 40 % pour la vague 1 (env. 30 000 individus), et 30 % pour la vague 2 et pour la vague 3 (env. 22 500 individus). Le découpage de l'échantillon en vagues permet de tirer parti d'une base de sondage comportant les informations de contact (p.

ex. lieu de résidence et coordonnées) les plus à jour au début de la collecte de données de chaque vague, et de mieux répartir l'échantillon sur l'ensemble de la période de collecte. Le tirage de l'échantillon de chaque vague s'est fait selon la stratification de la base de sondage en vue d'assurer un contrôle de la composition de l'échantillon selon le territoire géographique, la catégorie d'âge (15-24 ans, 25-44 ans, 45-64 ans, 65-74 ans, 75 ans et plus) et le sexe de la personne. Un échantillon a ensuite été sélectionné par l'ISQ de manière aléatoire et indépendante dans chacune des strates.

2

Collecte de données

Cette section porte sur les méthodes et les résultats de la collecte des données. On y présente notamment les instruments et le mode de collecte, ainsi que les taux de réponse obtenus.

2.1 Prétest

Le prétest de l'ESLPQ 2024 s'est déroulé du 23 octobre au 10 novembre 2023 auprès d'un échantillon aléatoire de 499 personnes de 15 ans et plus. Les entrevues ont été effectuées uniquement au moyen d'interviews téléphoniques assistées par ordinateur (ITAO). L'objectif était de valider l'intelligibilité des questions et la fluidité du questionnaire, et d'estimer le temps requis pour répondre aux différentes sections au téléphone.

Après l'analyse des résultats, des ajustements majeurs ont été apportés aussi bien dans la structure du questionnaire que dans la formulation de certaines questions. Compte tenu de ces modifications, les données fournies par les 165 répondantes et répondants au prétest n'ont pas été conservées pour les estimations produites en fin d'enquête. Quelques personnes échantillonnées au prétest, qui n'ont pas répondu ni refusé de participer, ont été échantillonnées à nouveau pour l'enquête proprement dite et ont donc été de nouveau invitées à répondre à l'enquête.

2.2 Méthodes de collecte

Instruments de collecte

Lors de l'élaboration du questionnaire, plusieurs critères ont été pris en compte : la pertinence au regard des objectifs poursuivis par l'enquête, la perspective démolinguistique, la précision pour les estimations à l'échelle territoriale, la complémentarité avec d'autres sources de données, la cohérence et l'ergonomie du questionnaire pour le répondant ou la répondante (durée du questionnaire facilité à le remplir, etc.). Lors de l'élaboration du questionnaire, l'Institut a été accompagné, comme c'est généralement le cas dans ses enquêtes, d'un comité d'orientation de projet regroupant des membres du ministère de la Langue française et de l'Office québécois de la langue française ainsi que des chercheurs universitaires experts dans l'étude de l'utilisation des langues par les populations.

Le questionnaire de l'enquête est subdivisé en sections qui abordent les thématiques suivantes :

- Section 1 : connaissance et maîtrise des langues ;
- Section 2 : usage des langues avec l'entourage ;
- Section 3 : usage et fréquence d'utilisation des langues dans différentes situations ;
- Section 4 : langue d'enseignement ;
- Section 5 : questions sociodémographiques.

Le déroulement du processus de réponse au questionnaire varie selon les réponses que la personne fournit à la première question, soit celle sur la ou les langues connues suffisamment pour tenir une conversation simple. Ainsi, les choix de réponses proposées à plusieurs questions subséquentes sont déterminés par les langues que la personne a déclaré connaître. Par ailleurs, si une personne déclare connaître une seule langue, certaines des questions subséquentes ne sont pas posées.

Le questionnaire de l'enquête n'était administré qu'en français ou en anglais, mais comme la population visée incluait les personnes ne parlant ni le français ni l'anglais, il importait de pouvoir obtenir la participation de ces personnes. Pour ce faire, nous avons eu recours, lorsque possible, à une tierce personne (personne qui répond pour la personne échantillonnée) afin de recueillir les informations concernant ces personnes dont l'incapacité à participer à l'enquête est due à une « barrière linguistique ». Pour remédier à cela, un questionnaire réservé aux tierces personnes répondantes, administré seulement au téléphone, a été conçu. Ce questionnaire plus court se limite à des questions factuelles auxquelles une tierce personne proche de la personne sélectionnée pourrait être en mesure de répondre.

La durée moyenne des entrevues téléphoniques est estimée à 17 minutes tandis que celle des questionnaires remplis sur le Web est évaluée à environ 13 minutes.

La durée moyenne pour les deux modes de collecte confondus est de 14 minutes et elle varie selon le nombre de langues que la personne répondante connaît suffisamment pour tenir une conversation simple, passant de 12 minutes pour une seule langue à 20 minutes pour 5 langues.

Modes de collecte

La collecte des données a été effectuée par interview téléphonique assistée par ordinateur (ITAO) et/ou par interview Web assistée par ordinateur (IWAQ). Ces modes de collecte présentent des avantages qui améliorent

la qualité des données. Premièrement, le libellé des questions, y compris les périodes de référence et les pronoms, est personnalisé automatiquement en fonction de facteurs comme l'âge et le genre de la personne, et des réponses aux questions précédentes. Deuxièmement, des mesures de contrôle isolent les réponses incohérentes ou hors norme, et des instructions apparaissent à l'écran lorsqu'une telle situation se présente. La personne remplissant le questionnaire reçoit une rétroaction immédiate, et une correction de l'incohérence peut être apportée. Troisièmement, le processus est prévu de sorte que les questions qui ne concernent pas la personne sont automatiquement omises.

Période de collecte

La collecte des données de l'ESLPQ 2024 s'est déroulée du 9 janvier au 26 août 2024. Comme mentionné à la section 1.3, l'échantillon de l'enquête a été séparé en trois vagues de collecte. La collecte de la première vague a débuté le 9 janvier, celle de la deuxième vague a été amorcée le 7 mars et celle de la troisième et dernière vague a commencé le 1^{er} mai. La collecte s'est terminée au même moment pour les trois vagues.

Stratégies de collecte

Une lettre de présentation a été envoyée aux personnes sélectionnées pour expliquer les objectifs de l'enquête et les inviter à remplir le questionnaire en ligne. Deux relances postales ont ensuite été envoyées à intervalles d'environ deux semaines aux personnes n'ayant toujours pas répondu. Si le questionnaire n'était pas encore rempli environ six semaines après l'envoi de la lettre initiale, des appels téléphoniques ont été passés pour encourager les réponses. Une entrevue téléphonique était proposée aux personnes préférant ne pas répondre en ligne ou n'ayant pas accès à Internet. Enfin, vers la fin de la collecte, une lettre supplémentaire a été envoyée à une partie des personnes n'ayant pas encore répondu, principalement celles injoignables par téléphone.

2.3 Résultats de la collecte

À l'instar des taux de réponse aux enquêtes généralement présentés par l'ISQ, ceux de l'ESLPQ sont pondérés⁸. Ils tiennent compte, entre autres, de la non-proportionnalité territoriale de l'échantillon de l'enquête. Au total, 45 280 personnes ont participé à l'ESLPQ 2024, ce qui correspond à un taux de réponse pondéré de 61 %. Le taux de réponse visé de 60 % a été atteint dans tous les territoires géographiques, sauf pour la RA de Montréal où le taux est très légèrement inférieur à la cible. Le nombre de personnes répondantes attendu a ainsi été atteint pour tous les territoires à l'exception de la RA de Montréal, pour laquelle il ne manquait que peu de personnes répondantes.

Parmi toutes les entrevues effectuées, 112 ont dû être réalisées auprès d'une tierce personne répondante, ce qui correspond à un faible taux de déclaration par procuration (0,2 %).

Comme prévu, le taux d'admissibilité pondéré à l'enquête est très élevé (98,8 %). Peu de personnes ont été jugées inadmissibles au moment de la collecte : celles qui avaient déménagé dans un logement collectif institutionnel ou à l'extérieur du Québec et celles qui étaient décédées. Finalement, 78,8 % des répondants et répondantes ont opté pour le questionnaire en ligne.

Tableau 2.1

Nombre de personnes répondantes et taux de réponse pondéré selon quelques caractéristiques¹, Québec, 2024

Caractéristiques	Nombre de personnes répondantes	Taux de réponse pondéré
	n	%
Ensemble du Québec	45 280	61,3
Territoire géographique		
RMR de Montréal	27 871	61,3
RA de Montréal	17 118	59,9
Reste de la RMR de Montréal	10 753	62,4
RA de la Capitale-Nationale	4 377	64,1
Municipalité de Gatineau	2 954	61,8
Ailleurs au Québec	10 078	60,7
Âge		
15-17 ans	1 718	61,0
18-24 ans	3 086	49,7
25-34 ans	5 897	51,6
35-44 ans	7 349	60,4
45-54 ans	6 898	60,2
55-64 ans	7 809	66,2
65 ans et plus	12 523	68,4
Sexe		
Femme	24 074	63,9
Homme	21 206	58,6
Langue de correspondance		
Français	38 839	62,2
Anglais	6 441	54,6

1. Les informations sur l'âge, le sexe et la langue de correspondance sont celles de la base de sondage (FIPA).

Source : Institut de la statistique du Québec, *Étude sur la situation des langues parlées au Québec 2024*.

8. Plus de renseignements sur la pondération sont fournis à la section 3.2.

3 Traitement des données

Cette section aborde les étapes de traitement des données recueillies. Elle traite de la validation de la qualité des données, de la pondération nécessaire à l'inférence à la population visée ainsi que de l'examen de la non-réponse totale et partielle et des traitements d'imputation effectués.

3.1 Validation et saisie

Une validation du statut de réponse des personnes sélectionnées (répondantes, non répondantes ou inadmissibles) a d'abord été effectuée. Certaines personnes dont les questionnaires étaient incomplets ont été considérées comme non répondantes. Des questions en début d'entrevue visaient à garantir que le questionnaire était bel et bien rempli par la personne sélectionnée. Des validations ont été effectuées en cours de collecte et a posteriori pour comparer les réponses fournies aux renseignements contenus dans la base de sondage, plus particulièrement en ce qui concerne la date de naissance et le code postal du lieu de résidence. La concordance étant exacte à quelques exceptions près, rien n'indiquait que certains questionnaires devaient être rejetés.

Étant entièrement informatisées, les collectes téléphoniques et sur le Web permettent d'effectuer plusieurs validations de base pendant l'entrevue, notamment une vérification du respect des choix de réponse ou de l'adéquation des sauts de section, pour les blocs de questions ne visant qu'un sous-groupe de répondants et répondantes. Une validation a posteriori a assuré la cohérence des réponses fournies par une même personne, ce qui a permis de corriger quelques erreurs. Les réponses ont été colligées de manière à distinguer les personnes ayant omis de répondre à une question de celles qui n'étaient pas concernées.

3.2 Pondération

La pondération est essentielle pour la production des résultats de l'enquête. Elle permet de faire des inférences adéquates à la population visée, bien que celle-ci n'ait pas été sondée dans sa totalité. Elle consiste à attribuer un poids statistique à chaque personne répondante. Ce poids correspond au nombre de personnes qu'un répondant ou une répondante représente au sein de la population visée. Il doit tenir compte, entre autres, de la probabilité de sélection de la personne, prédéterminée par le plan d'échantillonnage, et de la non-réponse à l'enquête. En effet, en raison des objectifs de diffusion de résultats à l'échelle des territoires, le plan de sondage a inévitablement entraîné des probabilités de sélection très variables. De plus, il est connu que dans les enquêtes, la probabilité de répondre varie selon plusieurs caractéristiques sociodémographiques. Il est donc nécessaire de tenir compte de ces éléments en les intégrant à la pondération.

De façon plus détaillée, le poids initial de chaque personne faisant partie de l'échantillon de l'ESLPQ 2024 a été déterminé en fonction de l'inverse de sa probabilité de sélection. Ensuite, un ajustement a été fait pour l'admissibilité à l'enquête, qui varie significativement selon les renseignements du FIPA suivants : l'âge, le nombre de personnes résidant à la même adresse et la langue de correspondance. Le statut d'admissibilité étant inconnu pour la plupart des personnes non rejointes, il fallait effectivement réduire leur poids afin de refléter le fait qu'une partie d'entre elles étaient probablement inadmissibles. Le poids de ces personnes a été multiplié par le taux

d'admissibilité pondéré observé à l'enquête parmi les personnes ayant des caractéristiques semblables pour lesquelles l'admissibilité a pu être déterminée.

L'étape suivante est l'ajustement pour la non-réponse totale. La méthode du score de propension à répondre a été employée (Haziza et Beaumont 2007 ; Eltinge et Yansaneh 1997). Elle consiste à modéliser le fait d'avoir répondu ou non à l'enquête selon les renseignements disponibles dans la base de sondage, soit l'âge, le sexe, le territoire de résidence, le nombre de personnes habitant à la même adresse, la langue de correspondance, la présence d'un conjoint ou d'une conjointe à l'adresse, la présence d'au moins un numéro de téléphone et l'indice de défavorisation matérielle et sociale (Gamache et autres 2017). Des classes composées de personnes ayant des caractéristiques et une propension à répondre semblables ont ainsi été formées. À l'intérieur de chaque classe, le poids des répondants et répondantes a été ajusté par l'inverse du taux de réponse observé à l'enquête.

Étant donné que les personnes qui ne connaissent ni l'anglais ni le français présentent des caractéristiques et une propension à répondre très différentes de celles des autres personnes sélectionnées, un ajustement spécifique a été fait pour les personnes questionnées par le biais de tierces personnes répondantes dans le cadre du traitement de la non-réponse totale. En raison de la faible taille de cette sous-population⁹ et du taux de réponse relativement faible dans celle-ci (26 %), une seule classe de pondération a été considérée et le poids des personnes questionnées via une tierce personne répondante a été ajusté par l'inverse du taux de réponse¹⁰.

Par la suite, on a vérifié qu'aucune personne n'avait de poids très élevé comparativement aux poids des personnes de la même strate¹¹, afin de s'assurer qu'une personne répondante n'ait pas une influence indue sur les statistiques produites. Pour ce faire, une méthode appelée « écart-sigma » a été utilisée (Bernier et Nobrega 1998). Lorsque jugé trop élevé, le poids d'une personne a été abaissé au poids inférieur le plus près dans la même strate. Moins de 0,1 % des poids ont ainsi été modifiés.

L'étape suivante, soit le calage aux marges, vise notamment à limiter les risques associés aux enjeux de couverture de la base de sondage¹². Cela consiste à ajuster la pondération afin que la somme des poids des personnes répondantes corresponde bien aux effectifs connus de la population visée par l'enquête¹³, et cela pour :

- le croisement du genre et de la catégorie d'âge (8 catégories : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans, 65-74 ans, 75-84 ans, 85 ans et plus) à l'échelle de l'ensemble du Québec ;
- le croisement du genre et de la catégorie d'âge (7 catégories : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans, 65-74 ans, 75 ans et plus) à l'échelle territoriale.

Pour finir, une nouvelle vérification a été faite pour s'assurer qu'aucune personne répondante n'avait un poids très élevé comparativement aux poids des personnes de la même strate définie par chacun des croisements considérés ci-dessus. La même méthode « écart-sigma » a été utilisée. Aucun ajustement n'a été nécessaire en lien avec cette vérification.

9. Les personnes non répondantes pour cette sous-population correspondent aux personnes qui ont été identifiées lors des entrevues téléphoniques comme incapables de répondre à l'enquête à cause de la langue (aucune ou très faible connaissance du français et de l'anglais). La taille de la population des personnes sélectionnées ne connaissant ni le français ni l'anglais est donc potentiellement sous-estimée. D'une part, il est possible que de tierces personnes aient rempli le questionnaire en ligne pour des personnes sélectionnées (seules les tierces personnes ayant passé une entrevue téléphonique ont été prises en compte.) D'autre part, parmi les personnes non répondantes, il y en a une partie qui n'a pas répondu à l'enquête à cause de la langue, mais que l'ISQ n'a pas pu identifier, car aucun contact n'a pu être établi avec elles.

10. Seules les personnes sélectionnées dont la raison de ne pas répondre était une incapacité à cause de la langue (raison de non-réponse offerte uniquement lors d'une entrevue téléphonique) ont été prises en compte. Dans l'ajustement des poids effectués à cette étape-ci de la pondération, les personnes ayant participé grâce à une tierce personne répondante ne « représentent » que ces personnes non répondantes.

11. La strate considérée est formée par le croisement du sexe, de la catégorie d'âge (7 catégories : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans, 65-74 ans et 75 ans et plus) et du territoire géographique.

12. Ainsi, en faisant l'hypothèse que les personnes absentes du FIPA ont des caractéristiques semblables à celles qui y figurent, il est possible d'inférer les résultats à l'ensemble de la population visée.

13. Ces effectifs sont dérivés des estimations de population de l'ISQ au 1^{er} juillet 2023.

3.3 Non-réponse totale

Définition

La non-réponse totale se produit lorsqu'une personne sélectionnée et admissible ne remplit pas le questionnaire de l'enquête pour une raison quelconque.

Cette non-réponse peut entraîner des biais dans les estimations si les personnes ayant répondu présentent des caractéristiques différentes de celles ne l'ayant pas fait, et que ces caractéristiques sont liées au sujet de l'enquête. La pondération est ajustée pour la non-réponse à l'enquête (section 3.2), ce qui permet de réduire le risque de biais dû à celle-ci. Toutefois, seules les informations contenues dans la base de sondage, donc connues pour l'ensemble de l'échantillon, peuvent être prises en considération pour cet ajustement. Ainsi, malgré l'utilisation de la pondération, des résultats de l'enquête peuvent quand même être biaisés si la non-réponse totale est liée à une ou plusieurs caractéristiques non disponibles dans la base de sondage, et que ces caractéristiques sont fortement corrélées à certains indicateurs.

3.4 Non-réponse partielle

Définition

La non-réponse partielle désigne l'absence de réponse à une question (ou la présence d'une réponse non valide) pour certaines personnes ayant rempli le questionnaire. Il est connu qu'une non-réponse partielle importante peut entraîner certains biais dans les estimations, au même titre que la non-réponse totale, s'il s'avère que les non-répondants et non-répondantes présentent des caractéristiques différentes de celles des personnes répondantes et que ces caractéristiques sont de surcroît liées au thème étudié. La pondération ne tient pas compte de la non-réponse partielle comme elle le fait pour la non-réponse totale.

Taux de non-réponse partielle

Le taux de non-réponse partielle pondéré à une question est défini comme le rapport entre le nombre pondéré de personnes n'ayant pas répondu à celle-ci et le nombre pondéré de personnes admissibles à y répondre. Plus ce taux est élevé, plus le risque de biais induits par la non-réponse partielle est grand. On fait généralement l'hypothèse qu'une non-réponse partielle inférieure à 5 % a une incidence négligeable sur les estimations.

La non-réponse partielle étant très faible pour la grande majorité des questions de l'ESLPQ 2024, son incidence sur le risque de biais dans les résultats est faible et aucune analyse approfondie de la non-réponse partielle n'a été nécessaire. Le seul indicateur qui affiche une non-réponse partielle supérieure à 5 % est le revenu du ménage. Ce dernier a toutefois été imputé pour tous les répondants et répondantes lorsqu'il était manquant, et ce, de manière à réduire les risques de biais (voir Annexe 2).

Par ailleurs, bien que les répondants et répondantes par procuration n'avaient pas à répondre à toutes les questions, ce qui a entraîné une non-réponse partielle pour les questions omises, le biais potentiel engendré par cette non-réponse partielle peut être considéré comme négligeable.

4

Analyse des données, précision des estimations et tests statistiques

Cette section porte sur certains aspects de l'analyse, dont la construction de certains indicateurs disponibles dans l'ESLPQ, l'estimation de la précision et les tests statistiques.

4.1 Source et construction des indicateurs sur les langues le plus fréquemment utilisées dans différentes situations

Dans l'*Étude sur la situation des langues parlées au Québec*, il est possible de calculer l'indicateur des langues le plus fréquemment utilisées, et ce, pour toute une série de situations (au travail, dans les commerces, etc.). Les indicateurs sur les langues le plus fréquemment utilisées dans différentes situations sont obtenus à partir de deux questions : celle sur la connaissance des langues et celle sur la fréquence d'utilisation des langues dans la situation. Voici une description des deux questions utilisées pour construire l'indicateur des langues le plus fréquemment utilisées.

4.1.1 Question sur la fréquence d'utilisation des langues dans différentes situations

Le questionnaire comporte une série de questions ayant la même structure visant à questionner les personnes répondantes sur la fréquence d'utilisation de différentes langues dans différentes situations. Toutes les questions ayant cette structure permettent de construire un indicateur des langues le plus fréquemment utilisées¹⁴ pour chacune des situations. À titre d'exemple, le questionnaire comporte une question sur la fréquence d'utilisation de différentes langues pour réaliser les tâches associées aux fonctions de l'emploi principal. La question est posée de la manière suivante : « Pouvez-vous me dire à quelle fréquence dans votre emploi principal, vous réalisez les tâches associées à vos fonctions en [CLX] ? ».

CLX fait référence à chacune des langues que la personne a déclaré connaître suffisamment pour tenir une conversation simple, c'est-à-dire celles nommées à la question sur la connaissance des langues. En plus des langues CLX, la personne répondante a la possibilité de déclarer une autre langue qu'elle utilise dans une situation donnée, mais qu'elle n'a pas déclarée à la question

14. Pour plus d'informations sur les différents indicateurs qu'il est possible de construire, veuillez consulter, sur le site de l'ISQ, les [fiches des indicateurs](#) ainsi que le [questionnaire de l'enquête](#).

sur la connaissance des langues¹⁵. Pour chacune de ces langues, les choix de réponses proposés à la question sur la fréquence d'utilisation des langues sont : toujours ou presque¹⁶, souvent¹⁷, parfois¹⁸, rarement¹⁹, jamais²⁰, ne s'applique pas²¹, ne sait pas, ne répond pas.

Avant d'aborder la construction de l'indicateur des langues le plus fréquemment utilisées en soi, une description de la question sur la connaissance des langues est proposée.

4.1.2 Question sur la connaissance des langues (intitulé, choix de réponses et consignes)

Au début du questionnaire, les personnes répondantes sont invitées à déclarer toutes les langues qu'elles connaissent suffisamment pour tenir une conversation simple. La question sur la connaissance des langues est posée de la manière suivante : « Quelle langue connaissez-vous suffisamment pour tenir une conversation simple ? Vous pouvez indiquer plusieurs langues. N'oubliez pas d'indiquer toutes les langues que vous connaissez, y compris celle dans laquelle vous répondez à l'étude ». Les options de réponse permettent de sélectionner jusqu'à cinq langues à partir de menus déroulants. Chaque menu déroulant propose une sélection

de 156 langues issues d'une classification utilisée au *Recensement de la population de 2021*^{22,23}. Les langues sont également présentées dans un ordre particulier, à savoir :

- Français ;
- Anglais ;
- Différentes langues autochtones²⁴, triées par ordre alphabétique ;
- Autres langues²⁵, triées par ordre alphabétique ;
- Langue absente de la liste²⁶.

La ou les langues sélectionnées par la personne répondante sont celles dans lesquelles elle déclare être en mesure de tenir une conversation simple. Ces langues sont reprises dans les questions sur la fréquence d'utilisation des langues dans différentes situations (au travail, dans les commerces, etc.).

4.1.3 Construction de l'indicateur des langues le plus fréquemment utilisées

Les catégories de fréquence d'utilisation (toujours ou presque, souvent, parfois, rarement) sont utilisées pour construire les indicateurs des langues le plus fréquemment utilisées dans différentes situations. Plus

-
15. Autrement dit, une langue dans laquelle la personne répondante n'a pas déclaré pouvoir tenir une conversation simple. Une instruction supplémentaire demande à la personne répondante d'indiquer, en cas d'utilisation de plusieurs langues, celle qu'elle utilise le plus souvent.
16. La définition fournie aux personnes répondantes est : vous utilisez uniquement ou presque exclusivement la langue indiquée pour réaliser l'action.
17. La définition fournie aux personnes répondantes est : vous utilisez la langue indiquée au moins la moitié des fois où vous réalisez l'action, mais pas toujours ou presque.
18. La définition fournie aux personnes répondantes est : vous utilisez la langue indiquée moins de la moitié des fois où vous réalisez l'action, tout en l'utilisant plus que de façon exceptionnelle.
19. La définition fournie aux personnes répondantes est : vous utilisez la langue indiquée de façon exceptionnelle pour réaliser l'action.
20. La définition fournie aux personnes répondantes est : vous n'utilisez pas la langue indiquée pour réaliser l'action.
21. La définition fournie aux personnes répondantes est : vous ne faites pas cette action.
22. Dictionnaire, Recensement de la population, 2021 – annexe 2.2 Langue maternelle, langue parlée à la maison et langue de travail, classifications de 2021, 2016 et 2011 (statcan.gc.ca).
23. Dans le menu déroulant se trouve aussi le choix « langue absente de la liste ». Si cette option est retenue, la personne répondante peut saisir sa langue dans une boîte ouverte. L'ISQ procède ensuite au recodage des langues saisies dans les boîtes ouvertes.
24. Pour figurer dans le menu déroulant, une langue autochtone doit avoir été déclarée par au moins une personne comme langue maternelle lors du Recensement de 2021.
25. Pour figurer dans le menu déroulant, une langue doit avoir été déclarée par au moins 100 personnes comme langue maternelle lors du Recensement de 2021.
26. Si cette option est retenue, la personne répondante peut saisir sa langue dans une boîte ouverte.

concrètement, pour chaque personne, la ou les langues ayant la fréquence d'utilisation la plus élevée sont conservées pour la construction de l'indicateur, et ce, quelle que soit l'intensité de la fréquence²⁷. Les trois exemples ci-dessous illustrent des situations fictives avec les langues retenues pour la construction de l'indicateur (indiquées en gras dans le tableau) des langues le plus fréquemment utilisées dans différentes situations.

Pour le cas 1, la langue retenue est le français, car c'est la langue pour laquelle la fréquence est la plus élevée. Pour le cas 2, l'anglais et le français sont les deux langues utilisées le plus fréquemment. Enfin, pour le cas 3, c'est le mandarin. À noter que dans la construction de l'indicateur, toutes les langues autres que le français et l'anglais sont regroupées dans une modalité « autre(s) langue(s) ».

Une fois que la ou les langues le plus fréquemment utilisées sont identifiées, nous pouvons construire trois variables intermédiaires : une première variable pour le français, une seconde variable pour l'anglais et une

dernière variable pour les autres langues que le français et l'anglais. Chaque variable intermédiaire peut avoir comme valeur :

- 1, pour la langue qui obtient la fréquence la plus élevée ;
- 2, pour les langues qui n'ont pas la fréquence d'utilisation la plus élevée (que ce soit parce que la langue n'est pas utilisée ou parce qu'elle n'est pas utilisée suffisamment fréquemment).

Autrement dit, la variable intermédiaire pour le français est égale à 1 si la personne déclare que le français est la langue ou une des langues utilisées ayant la fréquence d'utilisation la plus élevée, et la variable intermédiaire pour l'anglais est égale à 1 si la personne déclare que l'anglais est la langue ou une des langues utilisées ayant la fréquence d'utilisation la plus élevée. Pour les langues autres que le français et l'anglais, il suffit qu'une de ces langues soit déclarée comme celle ayant la fréquence la plus élevée pour que la variable intermédiaire des langues autres que le français et l'anglais soit égale à 1.

Langues connues (CLX)	Fréquence d'utilisation				
	Toujours ou presque	Souvent	Parfois	Rarement	Jamais
Cas 1 (fictif)					
Français	X				
Anglais			X		
Cas 2 (fictif)					
Français		X			
Anglais		X			
Espagnol				X	
Cas 3 (fictif)					
Français					X
Anglais					X
Mandarin		X			

27. La question posée ne tient pas compte de la fréquence à laquelle l'action est réalisée, que ce soit une fois par jour ou une fois par mois. Elle porte seulement sur la fréquence d'utilisation des langues.

Ensuite, les différentes combinaisons des trois variables intermédiaires permettent de construire l'indicateur de la langue le plus fréquemment utilisée. Par exemple, si la variable intermédiaire pour le français est égale à 1 et que les variables intermédiaires pour l'anglais et les autres langues sont égales à 2, alors on considérera que la personne utilise le plus fréquemment uniquement le français. Le tableau 4.1 ci-dessous décrit les différentes combinaisons possibles.

Finalement, les différentes combinaisons des trois variables intermédiaires permettent d'appliquer à l'indicateur de la langue le plus fréquemment utilisée les modalités suivantes :

- Français uniquement ;
- Anglais uniquement ;
- Autre(s) langue(s) uniquement ;
- Français et anglais ;
- Français et autre ;
- Anglais et autre ;
- Français, anglais et autre langue.

4.2 Indicateur de niveau de revenu du ménage

Un indicateur du niveau de revenu a été construit à partir des données du revenu total du ménage (déclarées par les répondants et répondantes ou imputées à la suite de la validation) en utilisant l'indicateur de faible revenu basé sur la Mesure de faible revenu (MFR) avant impôt. Rappelons que l'ESLPQ, contrairement aux enquêtes portant spécifiquement sur le revenu, ne comporte qu'une seule question sur ce sujet. L'indicateur construit permet de créer des groupes homogènes de personnes selon le niveau de revenu de leur famille. Le premier groupe contient les personnes de la population visée qui sont les moins bien nanties en termes relatifs. C'est donc dire que cet indicateur comme variable de croisement est tout indiqué pour examiner les liens entre l'utilisation des langues et le niveau de revenu. En revanche, comme c'est en général le cas dans les enquêtes menées auprès des individus et qui comportent une seule question globale sur le revenu total du ménage, on observe une surestimation de la défavorisation des ménages causée par une sous-déclaration des revenus. Pour une estimation plus fiable du taux de faible revenu, le lectorat est invité à consulter les publications de l'Institut sur le sujet (Institut de la statistique du Québec 2023b).

Tableau 4.1

Les différentes combinaisons possibles et la modalité de recodage associée

Variables intermédiaires			Modalité de recodage
Français	Anglais	Autre(s) langue(s)	
1	2	2	Français uniquement
2	1	2	Anglais uniquement
2	2	1	Autre(s) langue(s) uniquement
1	1	2	Français et anglais
1	2	1	Français et autre langue
2	1	1	Anglais et autre langue
1	1	1	Français, anglais et autre langue

4.3 Précision des estimations et tests statistiques

La plupart des enquêtes statistiques comportent des erreurs dites d'échantillonnage, dues au fait que seule une partie des unités de la population visée est sélectionnée pour y participer. Ces erreurs se répercutent sur les estimations produites, dont la précision est également influencée par la complexité du plan de sondage. Il est donc nécessaire de mesurer la précision de chaque estimation, et d'en tenir compte dans l'interprétation des résultats inférés à la population visée.

Dans l'ESLPQ, l'estimation de la variance et les tests statistiques sont effectués à l'aide de poids d'autoamorçage. Une série de 500 poids d'autoamorçage a été créée afin de tenir compte adéquatement, dans l'estimation de la variance et la production des tests statistiques, du plan de sondage complexe et de tous les ajustements de non-réponse et de calage apportés à la pondération. Pour ce faire, on a d'abord sélectionné 500 échantillons d'autoamorçage selon un plan de sondage avec remise à partir de l'échantillon initial. Ensuite, toutes les étapes

de la pondération ont été appliquées à chacun des échantillons, générant ainsi 500 poids d'autoamorçage (Rust et Rao 1996). Ces poids doivent être utilisés dans l'estimation de la variance et dans les tests statistiques à l'aide de logiciels tels que SAS et SUDAAN.

Le coefficient de variation²⁸ (CV) est utilisé comme indicateur de précision relative pour les diffusions de résultat de l'ISQ ; les estimations dont le CV est supérieur à 15 % sont annotées dans les tableaux, comme on l'indique à la section 5, ainsi que dans le texte si nécessaire.

Un test statistique d'indépendance du khi-deux²⁹ peut être utilisé pour faire une comparaison globale des proportions entre différents sous-groupes (p. ex. les territoires géographiques). En présence d'un écart significatif au seuil de signification fixé, et lorsque la variable d'analyse ou la variable de croisement compte plus de deux catégories, des tests de comparaison des proportions peuvent être menés afin d'identifier les écarts les plus importants. Ces tests reposent sur une statistique de Wald construite à partir de la différence de la transformation « logit » des proportions (Korn et Graubard 1999).

28. Le coefficient de variation est obtenu en divisant l'erreur type de l'estimation par l'estimation elle-même.

29. On utilise une version modifiée du test du khi-deux habituel qui tient compte du plan de sondage de l'enquête : il s'agit de l'ajustement de Satterthwaite du test du khi-deux. Plus précisément, c'est la statistique F correspondant à cette correction du test du khi-deux qui est utilisée dans les analyses.

5

Présentation des résultats

Les estimations de proportions présentées dans les produits de diffusion de l'ISQ sont arrondies à une décimale dans les tableaux et à l'unité dans le texte, à l'exception des estimations inférieures à 5 %, qui sont présentées avec une décimale. Les proportions dont la décimale est ,5 sont arrondies à l'unité inférieure ou supérieure selon la seconde décimale. En raison de l'arrondissement, la somme des proportions peut être légèrement supérieure ou inférieure à 100 %.

Dans les tableaux, les estimations dont le coefficient de variation est inférieur ou égal à 15 % sont suffisamment précises pour être présentées sans indication. Celles dont le coefficient de variation est supérieur à 15 %, mais inférieur ou égal à 25 %, sont marquées d'un astérisque (*) indiquant que leur précision est passable. Elles doivent donc être interprétées avec prudence. Les estimations dont le coefficient de variation est supérieur à 25 % sont marquées d'un double astérisque (**) pour indiquer leur faible précision. Elles sont fournies à titre indicatif seulement.

Une autre mesure de la précision d'une estimation présentée dans les tableaux est l'étendue de l'intervalle de confiance (IC) : plus l'intervalle est court, plus la valeur du paramètre est circonscrite. L'intervalle de confiance associé à une proportion représente la zone

d'incertitude liée à l'estimation. Cette zone d'incertitude est étroitement liée au niveau de confiance choisi. Ainsi, un intervalle de confiance pour une proportion à un niveau de confiance de 99 % signifie que si l'enquête était répétée 100 fois et qu'à chaque fois, on estimait la proportion et calculait son intervalle de confiance, 99 des 100 intervalles contiendraient la vraie valeur de la proportion dans la population. Dans cette enquête, les intervalles de confiance diffusés sont calculés selon la méthode « logit » (ou « logit-Wald »)³⁰.

Comme mentionné à la section 4.3, les associations entre deux variables (analyses bivariées) sont examinées à l'aide d'un test statistique d'indépendance du khi-deux. Si ce test global est significatif et qu'au moins une des deux variables compte plus de deux catégories, des tests de comparaison de proportions sont menés afin de déterminer quelles sont les proportions qui diffèrent significativement l'une de l'autre. Le seuil de signification est fixé à 1 % pour ces tests. Des lettres ajoutées en exposant aux statistiques présentées indiquent, pour chaque catégorie de la variable d'analyse, quelles sont les paires de catégories d'une variable de croisement pour lesquelles les résultats de la variable d'analyse diffèrent significativement. Une même lettre révèle un écart significatif au seuil de 1 % entre les résultats relatifs à deux catégories.

30. L'IC est calculé en ayant au préalable appliqué la transformation « logit » ($\text{logit}(p) = \log(p/1-p)$). Pour obtenir un IC pour la proportion p , les bornes de cet intervalle sont retransformées en appliquant la transformation inverse, ce qui génère un IC asymétrique pour p . Les bornes de l'IC ainsi généré sont toujours comprises dans l'intervalle [0,1]. Cependant, elles ne peuvent être calculées si la proportion estimée est de 0 ou de 1.

6

Portée et limites de l'enquête

Les données obtenues auprès des 45 280 personnes répondantes de l'ESLPQ 2024 offrent un grand potentiel analytique à l'échelle du Québec et à l'échelle de chacun des cinq territoires étudiés. Tout a été mis en place pour maximiser la qualité et la représentativité des résultats produits. Des efforts ont notamment été déployés pour obtenir des taux de réponse d'au moins 60 % dans chacun des territoires géographiques. Toutes les estimations produites dans le cadre de l'ESLPQ 2024 sont pondérées et tiennent compte non seulement du plan de sondage, mais aussi de la non-réponse totale, de manière à assurer la fiabilité de l'inférence à la population visée. De plus, toutes les mesures de précision et les tests statistiques ont été produits en considérant la complexité du plan de sondage de l'enquête. Les résultats de l'enquête sont donc représentatifs de l'ensemble de la population visée et peuvent être inférés non seulement à l'ensemble de la population visée, mais aux diverses sous-populations d'intérêt, comme les personnes d'un certain groupe d'âge, celles nées à l'extérieur du Canada ou celles détenant un diplôme de niveau secondaire, par exemple. Toutefois, il faut faire preuve de prudence lors de l'interprétation des estimations marquées d'un ou deux astérisques, pour lesquelles la précision est faible.

Il est important de mentionner que la collecte des données s'est déroulée dans un contexte où la situation linguistique québécoise a suscité une attention particulière sur la place publique dans les dernières années. Cette attention est liée, notamment, à des changements dans les programmes gouvernementaux et dans les lois visant à favoriser la langue française. Dans un tel contexte, il est possible que des personnes aient pu fournir, concernant leur usage des langues, des réponses influencées par une lecture donnée de la situation linguistique. On ne peut pas statuer sur l'existence ou non d'un tel phénomène. Soulignons toutefois que pour limiter le risque de biais dans les résultats, le questionnaire et la lettre d'invitation

ont été conçus pour porter sur l'ensemble des langues en usage au Québec plutôt que pour mettre l'accent sur l'utilisation du français ou de l'anglais.

Par ailleurs, notons que le taux de réponse obtenu auprès des personnes dont la langue de correspondance inscrite au FIPA est l'anglais a été un peu plus faible que celui obtenu auprès des personnes dont la langue de correspondance est le français. Ce phénomène est habituellement observé dans les différentes enquêtes de l'ISQ. Toutefois, pour l'ESLPQ en particulier (qui porte précisément sur les langues), il existe un risque que cette différence induise un biais. Afin de limiter ce risque de biais, la langue de correspondance a été prise en compte dans la pondération.

Le contexte dans lequel une enquête est réalisée peut avoir une incidence sur les comparaisons qu'il est possible de faire avec d'autres sources. D'ailleurs, la comparaison des résultats de l'ESLPQ avec ceux du recensement doit être évitée. Plusieurs différences rendent cette comparaison impossible. Le contexte d'enquête est différent, notamment parce que le recensement porte sur plusieurs sujets plutôt qu'exclusivement sur les langues, comme l'ESLPQ. Le fait que cette dernière se focalise sur la thématique des langues – et que les questions y soient posées dans un certain ordre – peut entraîner des réponses qui se distinguent de celles obtenues dans le cadre de questions similaires posées lors du recensement. Par ailleurs, la participation au recensement est obligatoire en vertu de la loi, alors que la participation à une enquête comme l'ESLPQ se fait sur une base volontaire. Le taux de réponse au recensement est très élevé en raison de son caractère obligatoire, alors que le taux de réponse à l'enquête est de 61 % et que les personnes ayant accepté de participer peuvent avoir des caractéristiques différentes de celles qui ne l'ont pas fait. Bien que des traitements statistiques aient été

appliqués pour corriger cette situation et obtenir des résultats représentatifs, il peut toujours y avoir un certain écart avec les résultats du recensement. De plus, la taille de l'échantillon de l'ESLPQ, bien que très importante, est loin du nombre de Québécois et Québécoises ayant répondu au recensement. Enfin, les questions posées dans le cadre de l'ESLPQ et l'ordre dans lequel elles sont posées diffèrent de ce que l'on retrouve dans le questionnaire du recensement, même si certaines questions se ressemblent.

Compte tenu des objectifs précis de l'enquête, les analyses présentées par l'ISQ s'appuient sur des méthodes bivariées. L'interprétation de certains résultats doit donc être faite avec prudence. Une analyse multivariée aurait été appropriée dans certains cas pour contrôler des facteurs confondants. L'approche retenue permet néanmoins d'explorer les données recueillies de façon utile et de fournir ainsi un portrait global de la situation des langues parlées au Québec.

Caractéristiques de la population visée

Tableau A1

Caractéristiques sociodémographiques de la population visée par l'*Étude sur la situation des langues parlées au Québec 2024*

	%
Genre	
Homme+	50,0
Femme+	50,0
Territoire géographique	
RMR de Montréal	50,7
RA de Montréal	24,3
Reste de la RMR de Montréal	26,4
RA de la Capitale-Nationale	9,1
Municipalité de Gatineau	3,3
Ailleurs au Québec	36,9
Âge	
15-17 ans	4,0
18-24 ans	8,9
25-34 ans	16,0
35-44 ans	16,0
45-54 ans	14,5
55-64 ans	16,2
65 ans et plus	24,4

Source : Institut de la statistique du Québec.

Imputation du revenu du ménage

L'imputation des données manquantes est une solution mise de l'avant pour minimiser l'effet de la non-réponse partielle dans les enquêtes. L'imputation consiste à remplacer la donnée manquante d'une personne par une valeur attribuée sur la base des renseignements disponibles au sujet de celle-ci. Une imputation valable requiert la présence de renseignements corrélés aux questions à imputer et qui ne présentent pas eux-mêmes une non-réponse partielle importante.

Avec un taux de non-réponse partielle pondéré de 18,5 %, la variable quantitative indiquant le revenu total du ménage a été imputée au moyen d'arbres de décision et de régression générés à partir de renseignements potentiellement liés au revenu, comme :

- l'âge de la personne répondante ;
- la composition de son ménage ;
- la perception de sa situation financière ;
- l'indice de défavorisation matérielle associé à son adresse.

En présence d'une non-réponse partielle aussi élevée, il est souhaitable de procéder à l'imputation de données, de manière à réduire les risques de biais. Toutefois, l'imputation a pour effet d'augmenter artificiellement le nombre de répondants et répondantes, ce qui fait que la variance des estimations produites est sous-estimée. Étant donné que la non-réponse partielle pour le revenu est plus élevée pour certaines sous-populations, notamment les jeunes de 15 à 24 ans, il est recommandé d'interpréter avec prudence les tests de différences dont le seuil observé est très près du seuil de signification fixé.

Références bibliographiques

- BERNIER, J., et K. NOBREGA (1998). "Outlier detection in asymmetric samples: A comparison of an inter-quartile range method and a variation of a sigma-gap method". *Congrès annuel de la société Statistique du Canada*, [En ligne], p. 137-141. [ssc.ca/sites/default/files/survey/documents/SSC1998_J_Bernier.pdf] (consulté le 27 septembre 2024).
- ELTINGE, J. L., et I. S. YANSANEH (1997). « Méthodes diagnostiques pour la construction de cellules de correction pour la non-réponse, avec application à la non-réponse aux questions sur le revenu de la U.S. Consumer Expenditure Survey », *Techniques d'enquête*, [En ligne], produit n° 12-001-X19970013103 au catalogue de Statistique Canada, vol. 23, n° 1, juin, p. 37-45. [www150.statcan.gc.ca/n1/fr/pub/12-001-x/1997001/article/3103-fra.pdf?st=YRiuivNK] (Consulté le 10 octobre 2024).
- AZEVEDO DA SILVA, M., N. GRAVEL, J. SYLVAIN-MORNEAU et al. (2024). *Indice de défavorisation matérielle et sociale 2021 : transfert des connaissances : guide d'utilisation*, [En ligne], [Québec], Institut national de santé publique du Québec, 9 p. [<https://www.inspq.qc.ca/sites/default/files/2024-04/3476-indice-defavorisation-materielle-sociale-2021.pdf>] (Consulté le 14 octobre 2022).
- HAZIZA, D., et J.-F. BEAUMONT (2007). "On the Construction of Imputation Classes in Surveys", *International Statistical Review*, [En ligne], vol. 75, n° 1, avril, p. 25-43. doi : [10.1111/j.1751-5823.2006.00002.x](https://doi.org/10.1111/j.1751-5823.2006.00002.x). (Consulté le 15 octobre 2024).
- INSTITUT DE LA STATISTIQUE DU QUÉBEC (2023b). *Revenu et faible revenu au Québec en 2019 : les plus récentes données et les tendances depuis 25 ans*, [En ligne], n° 2, 20 p. [statistique.quebec.ca/fr/fichier/revenu-faible-revenu-quebec-2019-donnees-tendances-depuis-25-ans.pdf] (Consulté le 27 février 2025).
- KORN, E. L., et B. I. GRAUBARD (1999). *Analysis of Health Surveys*, New York, John Wiley & Sons, 382 p.
- RUST, K. F., et J. N. K. RAO. (1996). "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research*, [En ligne], vol. 5, n° 3, septembre, p. 283-310. doi : [10.1177/096228029600500305](https://doi.org/10.1177/096228029600500305). (Consulté le 10 janvier 2025).

« Une organisation
statistique performante
au service d'une société
québécoise en évolution »

statistique.quebec.ca