

INSTITUT
DE LA STATISTIQUE
DU QUÉBEC

www.stat.gouv.qc.ca

MÉTHODOLOGIE

Enquête sur les pratiques linguistiques des ministères et organismes publics du gouvernement du Québec 2018

Méthodologie de l'enquête



Pour tout renseignement concernant le contenu de ce document :

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
Télécopieur : 418 643-4129

ou

Téléphone : 1 800 463-4090
(aucuns frais d'appel au Canada et aux États-Unis)

Site Web : www.stat.gouv.qc.ca

Dépôt légal
Bibliothèque et Archives nationales du Québec
4^e trimestre 2019
ISBN 978-2-550-85447-0 (en ligne)

© Gouvernement du Québec, Institut de la statistique du Québec, 2019

Toute reproduction autre qu'à des fins de consultation personnelle
est interdite sans l'autorisation du gouvernement du Québec.
www.stat.gouv.qc.ca/droits_auteur.htm.

Novembre 2019

Ce rapport a été réalisé à l'Institut de la statistique du Québec par :

Luc Belleau, statisticien

L'Enquête sur les pratiques linguistiques des ministères et organismes publics du gouvernement du Québec 2018 a été menée à bien grâce à la collaboration de plusieurs personnes.

Chargé de projet et travaux méthodologiques :

Luc Belleau, statisticien

Sous la direction de :

Éric Gagnon, directeur par intérim, Direction de la méthodologie

Les travaux de la Direction de la gestion de la collecte (DGC) de cette enquête ont été réalisés par :

Charles Alleyn, chargé d'enquête

Guillaume Rousseau, chargé d'enquête

Marcel Godbout, coordination

Avec la collaboration de :

Danny Dion, technicienne

Intervieweurs principaux :

Denis Bourgelas et Jonathan Lagacé

Autres intervieweurs :

Michelle Cloutier, Martine Labrecque, Clémence Mercille, France Pépin, Marcel Poirier et José Taschereau.

Sous la direction de :

Nathalie Loignon

Édition

Julie Boudreault, réviseure linguistique

Isabelle Jacques, mise en page

Direction de la diffusion et des communications

Notice bibliographique suggérée :

BELLEAU, Luc (2019). *Enquête sur les pratiques linguistiques des ministères et organismes publics du gouvernement du Québec 2018. Méthodologie de l'enquête*, Québec, Institut de la statistique du Québec, 24 p.

TABLE DES MATIÈRES

INTRODUCTION	7
LA MÉTHODOLOGIE	7
1 LE PLAN DE SONDAGE RETENU	9
1.1 La population visée par l'enquête et la base de sondage	9
1.2 Le plan d'échantillonnage, la taille et la répartition de l'échantillon.	10
2 L'INSTRUMENTATION ET LE MODE DE COLLECTE.	11
2.1 L'instrumentation.	11
2.2 Le mode de collecte.	11
3 LA COLLECTE DE DONNÉES	13
3.1 Le prétest	13
3.2 Le déroulement de la collecte.	13
3.2.1 Recherche de coordonnées	13
3.3 Les taux de réponse.	14
4 LA PONDÉRATION	15
5 LES MÉTHODES D'ESTIMATION ET D'ANALYSE DES RÉSULTATS	17
5.1 Les types d'estimation.	17
5.2 Les tests d'association	17
6 L'ÉVALUATION DE LA QUALITÉ STATISTIQUE DES RÉSULTATS	19
6.1 Les erreurs dues à l'échantillonnage	19
6.1.1 Intervalle de confiance	19
6.1.2 Coefficient de variation	19

6.2	Les erreurs non dues à l'échantillonnage	20
6.2.1	Le taux de réponse global et les taux de non-réponse partielle	20
6.2.2	Les réponses aux questions	21
6.2.3	Couverture de la base de sondage	21
7	LA CONFIDENTIALITÉ	23
8	L'APPRÉCIATION GLOBALE	23

INTRODUCTION

L'Institut de la statistique du Québec (Institut) a été mandaté par le Conseil supérieur de la langue française (CSLF) pour réaliser l'*Enquête sur les pratiques linguistiques des ministères et organismes publics du gouvernement du Québec 2018*. L'objectif de cette enquête est de déterminer les meilleures pratiques parmi celles repérées et de suggérer, le cas échéant, des pistes d'amélioration pour l'adoption de pratiques conformes aux objectifs de la Politique gouvernementale relative à l'emploi et à la qualité de la langue française dans les services gouvernementaux. Le présent rapport expose la méthodologie utilisée ainsi qu'une évaluation de la qualité de l'enquête.

LA MÉTHODOLOGIE

La méthodologie de l'*Enquête sur les pratiques linguistiques des ministères et organismes publics du gouvernement du Québec* est présentée dans les sections qui suivent. On y traite du plan de sondage retenu, de l'instrumentation et du mode de collecte, du prétest, de la collecte des données, de l'estimation, de l'évaluation statistique de l'enquête et de la confidentialité. On y trouve également une section faisant état de l'appréciation globale de l'enquête.

1 LE PLAN DE SONDAGE RETENU

1.1 LA POPULATION VISÉE PAR L'ENQUÊTE ET LA BASE DE SONDAGE

La population visée par l'enquête est constituée des employés qui ont des interactions avec des personnes physiques ou morales, et ce, parmi les ministères et organismes du gouvernement du Québec qui interagissent le plus couramment avec les citoyens.

Pour créer la base de sondage, le CSLF a d'abord sélectionné les ministères et organismes qui, selon lui, interagissent le plus couramment avec les citoyens. Dix-neuf ministères et organismes ont été sélectionnés :

1. Agence du revenu du Québec (Revenu Québec)
2. Autorité des marchés financiers (AMF)
3. Commission des normes, de l'équité, de la santé et de la sécurité du travail (CNESST)
4. Hydro-Québec
5. Ministère de la Justice
6. Ministère de la Santé et des Services sociaux (MSSS)
7. Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ)
8. Ministère de l'Éducation et de l'Enseignement supérieur (MEES)
9. Ministère de l'Immigration, de la Diversité et de l'Inclusion (MIDI)
10. Ministère des Affaires municipales et de l'Habitation (MAMH)
11. Ministère des Transports (MTQ)
12. Ministère de l'Environnement et de la Lutte contre les changements climatiques (MELCC)
13. Ministère du Travail, de l'Emploi et de la Solidarité sociale (MTESS)
14. Office de la protection du consommateur (OPC)
15. Régie de l'assurance maladie du Québec (RAMQ)
16. Régie du bâtiment du Québec (RBQ)
17. Régie du logement (RDL)
18. Retraite Québec
19. Société de l'assurance automobile du Québec (SAAQ)

Chacun de ces ministères ou organismes a fourni la liste de ses employés susceptibles d'avoir des interactions avec des personnes physiques ou morales. La base de sondage correspond à l'ensemble de ces listes. Une question filtre lors de la collecte nous permet de cibler les employés de ces listes qui ont effectivement des interactions avec les personnes physiques ou morales.

1.2 LE PLAN D'ÉCHANTILLONNAGE, LA TAILLE ET LA RÉPARTITION DE L'ÉCHANTILLON

Le CSLF désirait obtenir des renseignements fiables selon les régions administratives (14 catégories)¹.

Pour chaque région, un échantillon a été tiré de façon aléatoire. La taille de l'échantillon a été fixée à 6 086 employés, avec un taux d'admissibilité estimé de 82 % et un taux de réponse visé de 80 %, pour un objectif de 4 088 répondants. Le tableau 1 présente la taille de la population, la taille de l'échantillon et le nombre de répondants attendu, selon les régions administratives.

Tableau 1
Nombre d'employés dans la population, nombre d'employés dans l'échantillon et nombre de répondants attendu, selon les régions administratives

Région	Taille de la population	Taille de l'échantillon	Nombre de répondants attendu
Bas-Saint-Laurent	732	387	260
Saguenay–Lac-Saint-Jean	675	378	254
Capitale-Nationale	10 506	595	400
Mauricie et Centre-du-Québec	1 329	460	309
Estrie	655	374	251
Montréal	7 877	573	385
Outaouais	670	377	253
Abitibi-Témiscamingue, Côte-Nord et Nord-du-Québec	888	412	277
Gaspésie–Îles-de-la-Madeleine	559	349	234
Chaudière-Appalaches	754	392	263
Laval	1 697	484	325
Lanaudière	539	350	235
Laurentides	1 002	430	289
Montréal	2 929	525	353
Total	30 812	6 086	4 088

1. En raison du faible nombre d'employés dans la base de sondage de certaines régions, des regroupements de régions ont dû être effectués.

2 L'INSTRUMENTATION ET LE MODE DE COLLECTE

2.1 L'INSTRUMENTATION

Le questionnaire a été élaboré par le CSLF en collaboration avec l'Institut. La programmation du questionnaire Web a été réalisée par l'Institut. Le questionnaire a fait l'objet d'un prétest (voir la section 3.1).

2.2 LE MODE DE COLLECTE

La collecte de données a été réalisée au moyen de questionnaires Web et de relances par courriel et téléphone.

Cette collecte a été effectuée au moyen d'un système d'interview sur le Web assisté par ordinateur (IWAO).

3 LA COLLECTE DE DONNÉES

3.1 LE PRÉTEST

Le questionnaire Web a fait l'objet d'un prétest auprès de 300 employés du 1^{er} au 15 octobre 2018. Notons que le prétest a été effectué uniquement auprès des organismes dont la liste était finalisée au moment de débiter le prétest.

Au total, 226 questionnaires ont été remplis. Essentiellement, ce prétest a permis la validation de la stratégie de collecte ainsi que la correction ou même la suppression de questions problématiques.

3.2 LE DÉROULEMENT DE LA COLLECTE

La collecte de données s'est déroulée à l'automne 2018. Les répondants ont rempli les questionnaires Web du 12 novembre 2018 au 4 janvier 2019.

Une équipe de cinq intervieweurs ainsi que deux intervieweurs principaux ont été formés pour procéder à la relance téléphonique. Leur formation comportait une partie théorique sur le contexte de l'étude, une présentation du questionnaire informatisé et une période de mise en pratique dans le central téléphonique pour qu'ils se familiarisent avec les outils de collecte.

Le 12 novembre, on a envoyé un premier courriel à chaque participant l'invitant à répondre au questionnaire en ligne. Ce courriel comportait une description de l'enquête, un identifiant unique ainsi que l'adresse Internet pour accéder au questionnaire. Les répondants avaient jusqu'au 3 décembre pour répondre au questionnaire. Un courriel de rappel a ensuite été envoyé le 22 novembre à tous les employés qui n'avaient pas encore rempli le questionnaire, puis un dernier, le 20 décembre. En plus des courriels de relance, des appels téléphoniques ont été effectués du 3 au 21 décembre afin d'inciter les employés à répondre au questionnaire.

Au fur et à mesure que la collecte progressait, le taux de réponse par organisme était analysé. De cette manière, il a été possible de détecter rapidement qu'une intervention de la part du CSLF était nécessaire auprès des responsables de huit organismes dont le taux de réponse était particulièrement bas. Finalement, six de ces huit organismes ont eu un taux de réponse supérieur à 80 %.

3.2.1 Recherche de coordonnées

À la suite de l'envoi du courriel d'invitation, on a procédé à la recherche des numéros de téléphone associés aux courriels qui n'ont pas été acheminés correctement. Au cours de la collecte, les intervieweurs du central téléphonique ont effectué d'autres recherches pour les dossiers dont les coordonnées n'étaient pas bonnes. Le répertoire téléphonique du personnel de la fonction publique du gouvernement du Québec leur a permis de trouver les coordonnées exactes de plusieurs répondants. En tout, la recherche de numéros de téléphone a été infructueuse pour 228 des 6 086 dossiers que contenait l'échantillon (3,7 %).

Pour plusieurs dossiers, seul un numéro commun ne nous permettant pas de joindre la personne sélectionnée était disponible. De plus, même avec un numéro de poste, le service téléphonique ne nous permettait pas de joindre certains des employés sélectionnés.

3.3 LES TAUX DE RÉPONSE

Au total, 4 475 employés ont rempli le questionnaire, alors que l'on attendait 4 088 répondants (tableau 2).

Comme les employés n'ont pas tous la même probabilité d'être choisis, il est intéressant de considérer un taux de réponse établi sur la base des données pondérées¹ par l'inverse de cette probabilité. Le taux de réponse

pondéré final a été de 78,1 % (tableau 2), ce qui est légèrement inférieur que celui attendu, fixé à 80 %. Le taux d'admissibilité pondéré a été, quant à lui, plus élevé que celui attendu (92,4 % c. 84,0 %). Les résultats par région administrative sont affichés dans le tableau 2.

Tableau 2

Nombre d'employés répondants, nombre d'employés répondants attendu, taux de réponse et taux d'admissibilité pondérés, selon les régions administratives

Région	Nombre de répondants	Nombre de répondants attendus	Taux de réponse pondéré	Taux d'admissibilité pondéré
	n		%	
Bas-Saint-Laurent	274	260	77,2	93,5
Saguenay–Lac-Saint-Jean	285	254	78,8	95,6
Capitale-Nationale	442	400	79,7	88,0
Mauricie et Centre-du-Québec	350	309	80,4	94,1
Estrie	304	251	84,9	93,3
Montréal	441	385	76,6	95,0
Outaouais	269	253	75,6	95,5
Abitibi-Témiscamingue, Côte-Nord et Nord-du-Québec	372	277	81,5	95,1
Gaspésie–Îles-de-la-Madeleine	256	234	80,1	93,7
Chaudière-Appalaches	286	263	80,0	93,3
Laval	352	325	79,6	93,6
Lanaudière	253	235	76,4	96,4
Laurentides	270	289	66,5	93,4
Montérégie	375	353	75,9	95,5
Total	4 475	4 088	78,1	92,4

1. La pondération est abordée à la section 4.

4 LA PONDÉRATION

La pondération a pour objectif fondamental l'inférence des résultats observés à la population étudiée conformément au plan de sondage ayant permis de recueillir les données. En fait, le poids accordé à un employé reflète le nombre d'employés que celui-ci « représente » dans la population. Ce poids est donc fonction de deux réalités. D'abord, il constitue l'inverse de la probabilité de sélection d'un employé dans sa strate. Ensuite, il tient compte de la structure de la réponse obtenue dans la classe de pondération à laquelle l'employé appartient, en supposant que le profil des non-répondants de cette classe de pondération est semblable à celui des répondants. Cette dernière étape permet la constitution du poids final.

Pour effectuer une pondération adéquate, il faut être en mesure d'estimer le nombre total d'employés inadmissibles de la population. Dans la présente enquête, on a obtenu 4 475 répondants admissibles, 317 établissements inadmissibles et 1 294 non-répondants. Seuls ces derniers possèdent un statut d'admissibilité inconnu. Pour cette raison, l'hypothèse suivante a été émise : parmi les 1 294 non-répondants à l'enquête, certains peuvent être considérés comme admissibles, d'autres, comme inadmissibles. On peut déterminer un nombre (i_{nr}) de non-répondants inadmissibles notamment en examinant la nature de la non-réponse par strate. Une fois déterminé, ce nombre additionné à celui des répondants inadmissibles équivaut au nombre total d'établissements inadmissibles de l'échantillon. Pour ce faire, on suppose que le profil des non-répondants est similaire au profil des répondants par classe de pondération. Une classe de pondération est un regroupement d'employés ayant des caractéristiques similaires qui ont une influence sur le statut de réponse (répondant ou non-répondant) des employés.

Pour construire ces classes de pondération, on utilise la technique connue sous le nom de *chi-square automatic interaction detection* (CHAID). Cette technique consiste à sélectionner tout d'abord des variables ou

des caractéristiques, connues autant pour les employés répondants que ceux non répondants, selon lesquelles la non-réponse pourrait varier. On construit ensuite un tableau de fréquences à deux dimensions croisant le statut de réponse de l'employé avec chacune des variables choisies. Pour tous les tableaux ainsi créés, l'association ou la dépendance entre le statut de réponse et la variable est évaluée. Cela consiste à effectuer plusieurs tests statistiques du khi-deux. On calcule donc un poids final en tenant compte de la répartition des répondants à l'intérieur de chacune des strates de pondération à l'aide des équations suivantes :

- soit N_h = nombre total d'employés de la classe h ;
- soit n_h = nombre d'employés sélectionnés pour faire partie de l'échantillon de la classe h ;
- soit r_h = nombre d'employés répondants de la classe h ;
- soit nr_h = nombre d'employés non répondants de la classe h ;
- soit i_{nr_h} = nombre d'employés non répondants considérés inadmissibles de la classe h ;
- soit i_h = nombre d'établissements inadmissibles de la classe h .

Alors le poids $P_{ih}^{rép}$, tenant compte de la probabilité de sélection et de la non-réponse, accordé au **répondant** i de la classe de pondération h est :

$$P_{ih}^{rép} = \frac{N_h}{n_h} \times \frac{r_h + nr_h - i_{nr_h}}{r_h}$$

Et le poids P_{ih}^{inad} , tenant compte de la probabilité de sélection et de la non-réponse, accordé à l'employé **inadmissible** i de la classe de pondération h se calcule par :

$$P_{ih}^{inad} = \frac{N_h}{n_h} \times \frac{i_h + i_{nr_h}}{i_h}$$

5 LES MÉTHODES D'ESTIMATION ET D'ANALYSE DES RÉSULTATS

Les estimations produites par l'Institut dans le cadre de la présente enquête respectent les particularités du plan de sondage et sont exposées sous forme de tableaux de fréquences et de tableaux croisés.

5.1 LES TYPES D'ESTIMATION

Les estimations sont de deux types : l'estimation des paramètres (proportion, moyenne ou quartiles) et l'estimation de la précision associée à chacun des paramètres. Elles sont produites avec les données pondérées de sorte qu'elles puissent être inférées à la population visée. Les marges d'erreur, l'intervalle de confiance¹ et le coefficient de variation associé (voir la section 6.1) sont calculés selon le plan de sondage, à l'aide des logiciels SAS et SUDAAN.

5.2 LES TESTS D'ASSOCIATION

Un test global du khi-deux² a été effectué pour chacun des tableaux croisés produits par l'Institut et remis au CSLF, selon le plan d'analyse. Ce test nous permet de voir si une association existe entre la variable analysée et la variable de croisement. Le seuil de signification statistique a été fixé à 5 % pour tous les tests.

Les analyses sont uniquement bivariées, c'est-à-dire que chaque question est mise en relation avec une variable de croisement à la fois. Sauf exception, on a analysé chaque question selon la région administrative.

1. Consultez la section 6.1.1 afin d'obtenir plus de détails sur la méthode retenue pour construire les intervalles de confiance.

2. Plus précisément, c'est la statistique F (version log-linéaire) correspondant à un ajustement de Satterthwaite du test du khi-deux qui est utilisée.

6 L'ÉVALUATION DE LA QUALITÉ STATISTIQUE DES RÉSULTATS

Tout résultat tiré d'enquêtes par échantillon ou par recensement est potentiellement entaché d'erreurs dues ou non à l'échantillonnage. Les sections suivantes décrivent les efforts investis afin de limiter ces deux types de sources d'erreurs, de les identifier et de les quantifier.

6.1 LES ERREURS DUES À L'ÉCHANTILLONNAGE

Les erreurs dues à l'échantillonnage proviennent du fait que seule une partie de la population visée est échantillonnée aléatoirement. Ces erreurs sont contrôlables a priori à l'étape de la définition du plan de sondage et de la détermination de la taille d'échantillon. De plus, elles sont quantifiables a posteriori et permettent la prise en considération de la variabilité des estimations produites lors de leur analyse.

La marge d'erreur, les intervalles de confiance et le coefficient de variation sont des mesures utilisées afin de juger de la précision d'une estimation.

6.1.1 Intervalle de confiance

L'étendue de l'intervalle de confiance (IC) est l'une des mesures de la précision d'une estimation. Plus l'IC est court, plus la valeur du paramètre est circonscrite. L'IC 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, pour une proportion, un IC à un niveau de confiance de 95 % signifie que si l'on répétait l'enquête 100 fois et qu'à chaque reprise on estimait la proportion et calculait son IC, 95 des 100 intervalles ainsi créés contiendraient la vraie valeur de la proportion dans la population.

Il existe plusieurs types d'IC. Ceux-ci se distinguent par le choix de la loi servant à l'approximation de la distribution de l'estimateur. Les méthodes standard font habituellement

appel à la loi normale ou à la loi binomiale. Pour la présente enquête, les IC produits sont basés sur la méthode *logit*-Wald, qui fait appel à une transformation *logit* de la proportion et génère un intervalle asymétrique (Korn, 1998). Ces intervalles sont cohérents avec le test d'égalité de deux proportions, qui fait également appel à une transformation *logit*. L'IC de la fonction *logit* (\hat{p}) est délimité par les bornes inférieure et supérieure suivantes :

$$I_{logit} = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) - z_{1-\alpha/2} \left(\frac{1}{\hat{p}(1-\hat{p})}\right) \sqrt{\hat{v}(\hat{p})}$$
$$S_{logit} = \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) + z_{1-\alpha/2} \left(\frac{1}{\hat{p}(1-\hat{p})}\right) \sqrt{\hat{v}(\hat{p})}$$

Par transformation inverse, les bornes de l'IC pour sont définies comme suit :

$$I_{\hat{p}} = \frac{\exp(I_{logit})}{1 + \exp(I_{logit})}$$
$$S_{\hat{p}} = \frac{\exp(S_{logit})}{1 + \exp(S_{logit})}$$

Le niveau de confiance utilisé pour les intervalles produits à partir des données de cette enquête est de 95 %.

6.1.2 Coefficient de variation

Le coefficient de variation (CV) est une mesure qui permet de quantifier la précision relative d'une estimation et est également utilisé pour quantifier l'erreur d'échantillonnage. Il se définit comme suit :

$$CV = \frac{\sqrt{\text{variance de l'estimation}}}{\text{estimation}}$$

Cette mesure contribue à faciliter l'interprétation quant à la précision d'une estimation. Plus le CV est élevé, moins l'estimation est précise, et vice versa. Notamment, il faut être très prudent lors de l'utilisation des estimations dont le CV est très élevé (supérieur à 25 %), puisque cela indique une très grande variabilité. Ces estimations ne devraient être utilisées qu'à titre indicatif. Le CV a également l'avantage de permettre la comparaison de la précision de différentes estimations. Le tableau 3 présente la relation entre la valeur du CV et la précision de l'estimation utilisée par l'Institut.

Tableau 3
Relation entre la valeur d'un coefficient de variation et la précision de l'estimation associée

Bornes	Cote	Mention
$CV \leq 5\%$	A	Excellente précision
$5\% < CV \leq 10\%$	B	Très bonne précision
$10\% < CV \leq 15\%$	C	Bonne précision
$15\% < CV \leq 25\%$	D	Précision passable
$CV > 25\%$	E	Faible précision, à utiliser avec circonspection, fournie à titre indicatif seulement

Dans les tableaux produits pour le CSLF, la précision est exprimée par le CV de façon numérique et avec une cote telle que présentée au tableau 3.

La grande majorité des estimations de proportions et de moyennes effectuées pour l'ensemble des unités ont un $CV < 15\%$, ce qui signifie qu'elles sont assez précises et qu'elles peuvent être utilisées sans crainte. Toutefois, lorsque l'estimation d'une proportion est très petite, la précision de celle-ci tend à être moins satisfaisante ($CV > 15\%$). C'est aussi le cas pour les estimations de proportions sur des domaines particuliers (ex. : « Si oui, passez à la question X. ») qui sont parfois moins précises ($CV > 15\%$) en raison du nombre moindre de répondants admissibles. Enfin, le croisement d'une variable d'analyse avec une variable contenant également plusieurs catégories (ex. : la variable de la région administrative) engendra régulièrement de très grands CV dans les tableaux à cause du petit nombre de répondants sur lesquels sont basées les estimations. Il faut alors user de prudence dans l'interprétation des résultats. Ces estimations devraient être toujours accompagnées de leur CV exprimé en signe ou de leur IC.

6.2 LES ERREURS NON DUES À L'ÉCHANTILLONNAGE

Les erreurs qui ne sont pas dues à l'échantillonnage incluent les différents types de biais pouvant influencer sur les résultats : biais attribuables à la non-réponse, biais liés au mode de collecte, biais de réponse, surcouverture ou sous-couverture de la population par la base de sondage, etc. Elles ne sont pas formellement quantifiables, contrairement aux erreurs qui sont dues à l'échantillonnage. Il est donc important de tenter de détecter le plus possible leur présence et d'évaluer, le cas échéant, leur ampleur et leur effet.

6.2.1 Le taux de réponse global et les taux de non-réponse partielle

Le taux de réponse global à une enquête est un indicateur de la qualité de celle-ci, notamment en ce qui a trait aux biais dus à la non-réponse qui pourraient être introduits dans les estimations. En effet, plus le taux de non-réponse est élevé, plus il y a de risque que des biais soient introduits dans les estimations lors de l'inférence des données recueillies auprès des répondants à l'ensemble de la population, car les non-répondants peuvent présenter des caractéristiques différentes de celles des répondants.

L'Institut, comme tout organisme statistique, veille à limiter les risques de biais dans les enquêtes. C'est pourquoi il cherche à s'assurer du meilleur taux de réponse possible, allant même jusqu'à s'abstenir de procéder à l'inférence statistique lorsque ce taux n'atteint pas un niveau satisfaisant. Dans un tel cas, des analyses de biais sont faites afin de mesurer les risques d'erreur, et des correctifs sont apportés avant de produire les estimations.

Dans le cas présent, le taux de réponse global à l'enquête se situe à 78,1 %. Ce taux est un peu plus faible que l'hypothèse initiale de 80 %. Ce taux, combiné aux ajustements pour la non-réponse, assure quand même un faible risque de biais.

Pour ce qui est de la non-réponse partielle, c'est-à-dire de la non-réponse à l'une ou l'autre des questions, il est habituel de considérer qu'un taux de 5 % ou moins ne suscite pas d'inquiétude. Pour la présente enquête, toutes les questions ont un taux de non-réponse partielle inférieur à 5 %.

6.2.2 Les réponses aux questions

En général, les questions ont été bien comprises par la majorité des répondants; du moins, il n'y a pas eu d'indications contraires. Il n'existe pas de moyen absolu de valider la véracité de toutes les réponses reçues dans le cadre d'une enquête. Cependant, dans le cas présent, rien ne laisse supposer que des répondants ont délibérément fourni des réponses erronées.

6.2.3 Couverture de la base de sondage

Même si l'on peut supposer que les listes formant la base de sondage contiennent la presque totalité des employés ayant des interactions avec des personnes physiques ou morales, il est aussi certain qu'elle compte des employés qui n'en ont pas. Ce phénomène est connu comme la surcouverture de la population étudiée par la base de sondage.

Cette surcouverture fait en sorte qu'il est difficile de connaître la taille exacte de la population étudiée. Bien entendu, on peut faire des hypothèses: dans certaines enquêtes, il est possible de déceler toutes les unités échantillonnées qui ne sont pas admissibles et d'extrapoler ce résultat à l'ensemble de la population étudiée.

Dans la présente enquête, il n'a pas été possible de déterminer le statut de tous les employés. Dans bien des cas de non-réponse, il n'était pas facile d'attribuer le statut d'admissibilité ou d'inadmissibilité à l'enquête. Voici quelques exemples de non-réponse où le statut était ambigu: pas de réponse au numéro composé, mauvais numéro de téléphone, ligne occupée. Il fallait donc faire

l'hypothèse suivante: parmi ces cas se trouvent un certain nombre de cas admissibles et un certain nombre de cas inadmissibles. Ces nombres ont pu être estimés, par strate, à l'aide de la répartition des employés dont le statut était connu. En déterminant un nombre de non-répondants admissibles, il a été possible de donner un poids aux répondants pour que ceux-ci représentent tous les cas admissibles de l'échantillon. De cette façon, et avec la somme des poids sur l'ensemble des employés répondants, une estimation de la population admissible à l'enquête a pu être produite.

7 LA CONFIDENTIALITÉ

L'Institut est tenu de protéger la confidentialité des renseignements fournis par les répondants. Les résultats produits sous forme de tableaux sont donc vérifiés. Lors de la vérification, il est possible de déceler deux types de divulgation, soit une cellule d'un tableau contenant peu de répondants, soit un petit nombre de répondants contribuant à un grand pourcentage d'une estimation présentée dans le tableau.

Les tableaux présentant des problèmes de confidentialité ont été masqués, soit par un regroupement de catégories ou par un masquage primaire et secondaire des cellules problématiques.

8 L'APPRÉCIATION GLOBALE

Dans le contexte général de l'enquête, et en tenant compte des explications des sections précédentes, on peut conclure qu'il n'existe pas de problème majeur à inférer les résultats obtenus à la population étudiée.

Aussi, comme c'est le cas pour toute enquête, il y a lieu de faire preuve de prudence dans l'usage des résultats qui sont accompagnés d'un coefficient de variation supérieur à 15 %.

RÉFÉRENCES BIBLIOGRAPHIQUES

KORN, E. L., et B. I. GRAUBARD (1998). "Confidence intervals for proportions with small expected number of positive counts estimated from survey data", *Survey Methodology*, [En ligne], vol. 24, n° 2, décembre, p. 193-201. [www150.statcan.gc.ca/n1/en/pub/12-001-x/1998002/article/4356-eng.pdf?st=Vz25RgV6].

Des statistiques sur le Québec d'hier et d'aujourd'hui
pour le Québec de demain