

## Répertoire des corpus québécois de langue orale

Lionel Boisvert et Paul Laurendeau

Volume 17, numéro 2, 1988

Problèmes de lexicographie en Amérique

URI : <https://id.erudit.org/iderudit/602634ar>

DOI : <https://doi.org/10.7202/602634ar>

[Aller au sommaire du numéro](#)

### Résumé de l'article

Une brève présentation des sept principaux corpus de langue orale du français du Québec. Informations fournies : nom du corpus et sources d'information, coordonnées institutionnelles du corpus, informations sur les données et caractéristiques sociolinguistiques des informateurs, informations à propos de la cueillette et du traitement des données.

### Éditeur(s)

Université du Québec à Montréal

### ISSN

0710-0167 (imprimé)

1705-4591 (numérique)

[Découvrir la revue](#)

### Citer cette note

Boisvert, L. & Laurendeau, P. (1988). Répertoire des corpus québécois de langue orale. *Revue québécoise de linguistique*, 17(2), 241–261.  
<https://doi.org/10.7202/602634ar>

# RÉPERTOIRE DES CORPUS QUÉBÉCOIS DE LANGUE ORALE

Lionel Boisvert et Paul Laurendeau

En lexicographie, comme dans les autres branches de la linguistique, la description a été trop longtemps limitée à la seule langue écrite. Maintenant que la prise en compte de l'oral est une nécessité bien établie, il nous a semblé utile de présenter, de la façon la plus synthétique possible, différents corpus qui donnent immédiatement accès à la langue parlée des Québécois et notamment à son lexique. Constitués le plus souvent à des fins d'études sociolinguistiques, voire sociologiques ou anthropologiques, ces corpus de langue orale sont souvent fort mal connus de nombreux chercheurs qui s'intéressent aux études lexicales<sup>1</sup>.

Voici donc un bref aperçu de tous les corpus de français parlé québécois<sup>2</sup> dont l'existence nous a été signalée<sup>3</sup> et qui répondent aux deux critères suivants :

---

1. Nous sommes reconnaissants aux personnes qui nous ont aimablement transmis de l'information souvent difficile d'accès. Nous tenons particulièrement à remercier Gilles Bibeau, Denise Deshaies, André Dugas, Claire Lefebvre, Danielle Lepage, Pierre Maranda, Shana Poplack, David Sankoff et Diane Vincent. Ce répertoire a été constitué dans le cadre des travaux du TLFQ (Trésor de la langue française au Québec), projet de recherche subventionné par le Conseil de recherches en sciences humaines du Canada. Signalons finalement qu'au moment de la constitution de ce répertoire, nous n'avions pas encore pris connaissance de Blanche-Benvéniste et Jeanjean 1987, ouvrage sur la langue parlée à l'intérieur duquel on retrouve une brève description des corpus Sankoff-Cedergren et Centre-Sud (pp 50-53).

2. Les données du CORPUS DU FRANÇAIS PARLÉ À OTTAWA-HULL peuvent être considérées comme québécoises.

3. S'il y avait eu omission, ce serait ignorance et non malveillance. Il est cependant peu probable qu'un corpus important ait échappé à notre attention. En plus des contacts que nous avons nous même établis, nous avons diffusé dans tous les cégeps et universités du Québec une lettre circulaire dans laquelle nous avons défini ce que nous entendions par corpus au sens large : tout texte retenu sur un support

- a) être constitués d'énoncés oraux permettant la mise en contexte des unités;
- b) être dotés d'une méthode d'accès informatisée aux unités lexicales (nous avons ainsi exclu de notre inventaire les corpus bruts du type *Archives de folklore*).

Les corpus sont présentés ici dans l'ordre chronologique de leur constitution (la date des enquêtes en faisant foi), du plus ancien au plus récent. On trouvera pour chacun d'eux, et toujours dans cet ordre, les renseignements suivants :

- 1) Nom usuel et sources bibliographiques.
- 2) Coordonnées institutionnelles (localisation du corpus et renseignements sur les personnes à qui s'adresser pour de plus amples informations sur son utilisation éventuelle pour fins de recherche).
- 3) Renseignements sur les données et caractéristiques sociolinguistiques des informateurs.
- 4) Renseignements sur la méthode
  - a) cueillette des données;
  - b) traitement des données

---

(papier, magnétique, informatique) et pouvant être de quelque utilité dans l'étude des divers aspects de la langue parlée ou écrite (prononciation, vocabulaire, syntaxe etc.). Plus spécifiquement nous avons cherché à retracer des corpus en langue française suffisamment volumineux (60 pages ou deux heures d'enregistrement minimum), oraux (retranscrits ou uniquement enregistrés) ou écrits, ayant fait ou non l'objet d'un traitement (classifications, tables de concordances, etc.). À priori, tout type de relevé fut considéré comme valable : enquêtes ethnologiques, anthropologiques ou sociologiques, devoirs d'étudiants, listes d'inventaires, sondages d'opinions, archives juridiques, chansons, contes etc. Il était cependant essentiel que le corpus soit accessible au public d'une façon ou d'une autre (dans un dépôt d'archives ou un laboratoire de recherche) puisque son inclusion dans le présent répertoire vise à le rendre disponible pour des chercheurs qui en ignoreraient l'existence.

Aux fins de ce répertoire, nous nous sommes finalement limités aux corpus oraux informatisés.

**Corpus 1**1. *Nom usuel du corpus* : CORPUS BIBEAU-DUGAS*Source* : Dugas (1986).2. *Coordonnées institutionnelles du corpus*:

Corpus non publié, en dépôt à l'Université du Québec à Montréal. On peut y avoir accès à l'adresse suivante, avec l'autorisation des responsables :

Département de linguistique  
 Université du Québec à Montréal  
 Case postale 8888  
 Succursale A  
 Montréal (Québec)  
 H3C 3P8

Responsables : André Dugas, UQAM (514-282-3517)  
 Gilles Bibeau, U. de Montréal (514-343-7862)

3. *Renseignements sur les données*

Enquêtes réalisées en 1963-1964 auprès de 102 informateurs et informatrices de Montréal. 71 d'entre eux sont nés à Montréal ou y sont arrivés avant l'âge de 5 ans. Les informateurs sont ainsi répartis selon leur lieu de résidence :

— Saint-Henri, Lachine, Verdun, Lasalle	35
— Centre-Sud, Bas-Rosemont, Montréal-Est	28
— Ahuntsic, Montréal-Nord	16
— Notre-Dame-de-Grâce, Westmount	7
— Outremont, Hampstead, Côte-Saint-Luc	8
— Originaires de Montréal mais résidant ailleurs depuis peu	8

Répartition des 102 informateurs par sexe et tranche d'âge :

Âge	Sexe		Total
	M	F	
10-15 ans	10	1	11
16-20 ans	2	0	2
21-25 ans	4	2	6
26-30 ans	7	10	17
31-35 ans	4	10	14
36-40 ans	7	9	16
41-45 ans	0	4	4
46-50 ans	5	7	12
51-55 ans	1	4	5
56-60 ans	2	4	6
61-65 ans	1	3	4
66-70 ans	0	1	1
71-75 ans	0	1	1
76-80 ans	0	1	1
-----	-----	-----	-----
Total :	43	57	100

N.B. : Deux des informateurs n'ont pas donné leur âge.

Répartition des informateurs selon leur occupation :

- 48 des 57 femmes se disent «ménagères», 2 sont étudiantes et les autres sont des travailleuses semi-spécialisées ou spécialisées.
- 11 hommes sont étudiants; les autres pour la plupart sont des travailleurs semi-spécialisés ou spécialisés.

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

Les informations ont été recueillies sous forme d'interviews libres ou semi-dirigées. Les thèmes abordés concernent en général la biographie de l'informateur : histoire personnelle, travail, relations familiales, sports pratiqués (ou «regardés»), hobbies, activités culturelles. Les enregistrements sont d'une durée de 30 minutes chacun.

b) Traitement des données :

Les 102 enregistrements initiaux, toujours disponibles, ont été transcrits (dans une transcription approximative) par une «habile secrétaire de souche québécoise».

En 1979, avec en vue la création d'une base de données, le corpus initial a été réduit pour des raisons d'homogénéité aux seules entrevues réalisées auprès des 71 informateurs qui sont nés à Montréal ou y sont arrivés avant l'âge de 5 ans. C'est sur la base de ces 71 enregistrements qu'a été réalisée l'informatisation du corpus.

Les 71 enregistrements retenus ont fait l'objet d'une transcription normalisée qui respecte l'ordre syntaxique mais qui ne tient aucun compte des particularités de prononciation et qui restitue, par exemple, les pronoms, articles, conjonctions omis à l'oral. Ce n'est que dans la morphologie verbale qu'on a tenté, lors de la transcription, de tenir compte des formes effectivement produites en les restituant entre parenthèses après la forme normalisée. On dispose actuellement, à partir de cette transcription, d'un lexique trié dans l'ordre alphabétique avec des données sur la fréquence des entrées, et d'une mise en contexte de toutes les occurrences.

## Corpus 2

1. *Nom usuel du corpus* : CORPUS SANKOFF-CEDERGREN  
*Source* : Sankoff, Sankoff, Laberge et Topham (1976).

2. *Coordonnées institutionnelles du corpus*:

Corpus non publié, en dépôt à l'Université de Montréal. On peut y avoir accès à l'adresse suivante, avec l'autorisation du responsable :

Laboratoire d'ethnolinguistique  
Département d'anthropologie  
Université de Montréal  
C.P. 6128, Succursale A  
Montréal (Québec)  
H3C 3J7

Responsable : David Sankoff, Centre de recherches en mathématiques appliquées, U. de Montréal  
(514-343-7574)

### 3. Renseignements sur les données

Le corpus est constitué de 120 interviews effectuées en 1971 à Montréal, dans les secteurs à prédominance francophone (plus de 64 % de francophones), soit Outremont, Côte-des-Neiges, Saint-Léonard, Verdun, Ville-Saint-Pierre, Pointe-aux-Trembles, Rosemont, Lachine, Saint-Michel, Anjou, Rivière-des-Prairies, Plateau Mont-Royal, Saint-Henri, Centre, Centre-Sud et le port de Montréal. Tous les informateurs sont originaires de Montréal ou y ont vécu au moins depuis l'âge de six ans.

Les informateurs et informatrices, 60 hommes et 60 femmes, ont été sélectionnés de façon à fournir une image la plus fidèle possible des différentes strates socio-économiques de cette population : on a déterminé 6 catégories de revenus (\$5,100 et plus, \$4,100 à 5,099, \$3,600 à 4,099, \$3,100 à 3,599, \$2,600 à 3,099, \$2,200 à 2,599) et l'on a choisi 20 informateurs pour chacune de ces catégories.

#### Distribution selon l'âge :

55 ans et plus	30
35 - 54 ans	29
20 - 34 ans	32
15 - 19 ans	29
	<hr/>
Total	120

#### Distribution selon l'occupation du chef de famille :

professionnel-gérant	24
col blanc	26
col bleu	22
ouvrier	29
chômeur	14
ménagère	4
étudiant	1
	<hr/>
Total	120

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

Les entrevues ont pris la forme de conversations informelles où l'enquêteur tentait d'amener l'informateur à s'exprimer sur certains thèmes déterminés. Les thèmes couverts sont la vie et les coutumes au Québec dans le passé, la vie moderne à Montréal et les opinions de l'informateur sur la langue. À la fin de l'entrevue, on demandait à chacun de lire un petit texte intitulé «Une soirée de hockey».

Le corpus compte environ 100 à 120 heures d'enregistrement.

##### b) Traitement des données :

La transcription des entrevues a été faite directement sur cartes perforées IBM (il n'existe pas de version dactylographiée du corpus). Il s'agit d'une transcription selon l'orthographe du français standard, et non d'une transcription phonétique. On a cependant respecté la syntaxe des informateurs, l'ordre des mots. On rend même compte, dans une large mesure, des ratés, des pauses, des hésitations.

Les données de base se présentent sous la forme suivante : 120 bobines d'entrevues enregistrées et recopiées en deux exemplaires, 64 boîtes de cartes perforées (100,000 cartes), deux éditions imprimées par ordinateur et un fichier alphabétique de tous les mots avec contexte minimal et renvoi au texte suivi du corpus (tables de concordances éditées sur microfiches).

### Corpus 3

1. *Nom usuel du corpus* : CORPUS DE L'ESTRIE, appelé aussi CORPUS BEAUCHEMIN-MARTEL et CORPUS BEAUCHEMIN-MARTEL-THÉORET  
*Sources* : Beauchemin 1972, 1977 ; Beauchemin et Martel 1979 ; Beauchemin et Alii 1983, Beauchemin, Martel et Théoret 1983.



## 2. *Coordonnées institutionnelles du corpus:*

Ce corpus est publié; voir Beauchemin, Martel, Théoret (1973-1981). On peut en obtenir l'enregistrement sur cassettes à l'adresse suivante, avec l'autorisation des responsables :

Département d'études françaises  
 Faculté des Lettres et Sciences humaines  
 Université de Sherbrooke  
 Boulevard de l'Université  
 Sherbrooke (Québec)  
 J1K 2R1 (819-821-7266)

Responsables : Normand Beauchemin, U. de Sherbrooke  
 Pierre Martel, U. de Sherbrooke  
 Michel Théoret, U. de Sherbrooke

## 3. *Renseignements sur les données*

Relevés effectués en 1972 et 1973 dans vingt villes et villages des Cantons de l'Est (Cookshire, Bishopton, Waterville, Stoke, Saint-Elie-d'Orford, La Patrie, Compton, Coaticook, Bromptonville, Weedon, Saint-Denis-de-Brompton, Saint-Denis-d'Orford, Saint-Adolphe-de-Dudswell, Martinville, Rock-Forest, Lennoxville, East-Angus, Ascot-Corner, Ascot-Nord et Deauville) auprès de 78 hommes et de 65 femmes. Les informateurs sont tous francophones, âgés de 18 à 70 ans, nés dans une de ces municipalités ou paroisses, y ayant vécu au moins les 20 premières années de leur vie et ayant des parents (vivants ou morts) remplissant les mêmes conditions. Les 143 informateurs et informatrices ont été socialement répartis de la façon suivante :

— travailleurs et travailleuses non spécialisés (type : aide-déménageur, ménagère au presbytère, journalier, etc.)	18
— travailleurs et travailleuses semi-spécialisés (type : petit cultivateur, camionneur employé, chauffeur d'autobus scolaire, etc.) :	32
— travailleurs et travailleuses spécialisés (type : mécanicien, plombier, menuisier, etc.) :	53
— semi-professionnels et semi-professionnelles (type : enseignant, infirmière, gérant de petit magasin ou de rayon de grand magasin, etc.) :	28
— professionnels et professionnelles (type : directeur d'école, épicier-boucher propriétaire, entrepreneur-électricien ou plombier, etc.) :	12

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

Les informateurs et informatrices s'expriment en répondant aux questions d'un enquêteur dans une situation formelle d'entretien avec magnétophone visible. L'entrevue a été réalisée au domicile même de l'informateur. Les questions sont orientées sur des sujets précis (voir Beauchemin 1972). Les informateurs peuvent parfois savoir que l'objet de l'enquête est sociolinguistique. L'enquêteur interrompt fréquemment l'informateur pour passer à la question suivante, ce qui réduit une partie du contenu mais augmente les qualités interactionnelles de l'échange. Les entrevues durent entre 15 et 60 minutes chacune.

##### b) Traitement des données :

L'édition en translittération favorise la clarté de la lecture plutôt qu'une présentation exhaustive des faits phonétiques et morpho-syntaxiques (voir l'avant-propos du tome 4 des *Échantillons de textes libres*). Ce corpus est le seul des grands corpus de langue orale québécois à avoir été édité dans sa quasi-totalité (50 informateurs et 50 informatrices). Il est présenté en alphabet conventionnel et les répliques de l'enquêteur sont distinguées des répliques de l'informateur à l'aide de l'italique.

Pour trouver les coordonnées sociologiques de l'informateur, il faut consulter Beauchemin 1977 où un tableau associe à chaque numéro d'informateur une série d'indices numériques correspondant à des caractéristiques sociologiques présentées dans le corps de l'ouvrage. Pour localiser une unité lexicale, il faut consulter Beauchemin et Alii 1983 où la quasi-totalité des occurrences figurent avec un renvoi au texte du corpus.

**Corpus 4**1. *Nom usuel du corpus* : CORPUS CENTRE-SUD

Sources : Doran, Drapeau et Lefebvre (1982).

2. *Coordonnées institutionnelles du corpus* :

Corpus non publié, en dépôt à l'Université du Québec à Montréal. On peut y avoir accès à l'adresse suivante, avec l'autorisation de la responsable :

Département de linguistique  
 Université du Québec à Montréal  
 Case postale 8888  
 Succursale A  
 Montréal (Québec)  
 H3C 3P8

Responsable : Claire Lefebvre, UQAM (514-282-8464)

3. *Renseignements sur les données*

Enregistrements recueillis entre 1976 et 1978 à Montréal, dans le quartier Centre-Sud, à l'intérieur d'un quadrilatère délimité par les rues Sherbrooke, Sainte-Catherine, Amherst et Delorimier.

Le corpus est constitué essentiellement d'échantillons de discours d'adolescents et de pré-adolescents répartis ainsi, par sexe et groupe d'âges.

Âge	Sexe		total
	M	F	
6	0	1	1
7	0	1	1
8	0	0	0
9	0	4	4
10	0	2	2
11	1	1	2
12	3	1	4
13	1	5	6
14	4	7	11
15	4	1	5
16	3	0	3
17	1	1	2
18	1	1	2
-----	-----	-----	-----
Total :	18	25	43

De plus, en vue de constituer un échantillon-témoin, on a recueilli les interventions de 6 adultes, âgés de 42 à 52 ans, parents d'adolescents faisant partie du groupe ci-dessus.

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

Les informateurs adolescents ou pré-adolescents ont été enregistrés dans différentes situations de communication naturelle avec des amis, des parents, au cours d'activités variées : jeux, activités communes, etc. Les interviews menées spécifiquement auprès des parents sont les seules à avoir été dirigées. Le corpus compte une quarantaine d'heures d'enregistrement.

##### b) Traitement des données :

Le corpus a été transcrit dans sa quasi-totalité sur système «ordiné», selon l'orthographe standard. Il ne s'agit donc pas d'une transcription phonétique ou phonologique. On a toutefois respecté au mieux la syntaxe des informateurs. Certains morphèmes ont cependant été normalisés : ainsi [a:dzi] est retranscrit **elle a dit**.

Les données sont accessibles sous forme de transcription suivie des conversations enregistrées et de tables de concordances avec contextes minima et renvois au texte suivi.

### Corpus 5

#### 1. *Nom usuel du corpus* : LE FRANÇAIS PARLÉ DANS LA VILLE DE QUÉBEC, appelé aussi CORPUS DESHAIES

*Source* : Deshaies 1981.

## 2. Coordonnées institutionnelles du corpus:

Corpus non publié, en dépôt à l'Université Laval, Québec. On peut y avoir accès à l'adresse suivante, avec l'autorisation de la responsable :

Département de langues et linguistique  
Faculté des lettres  
Université Laval  
Cité universitaire  
Québec (Québec)  
G1K 7P4

Responsable : Denise Deshaies, U. Laval (418-656-5743)

## 3. Renseignements sur les données

Enquêtes effectuées en 1977 et 1979 auprès de 120 adolescents et parents francophones d'un quartier de la ville de Québec (quartier Saint-Sauveur) et d'un quartier de la ville voisine, Sainte-Foy (quartier Sainte-Foy). Amorcée auprès de groupes d'adolescents de ces deux quartiers en 1977, l'enquête s'est étendue en 1979 aux parents de ceux-ci qui étaient disposés à s'y prêter.

Répartition des informateurs selon le quartier, l'âge et le sexe (cette dernière donnée n'est présentée que pour les adolescents):

	quartier Sainte-Foy			quartier Saint-Sauveur			
	adolescents		parents	adolescents		parents	
	M	F		M	F		
69 ans et +	0	0	0	0	0	2	
50-59 ans	0	0	7	0	0	7	
40-49 ans	0	0	23	0	0	11	
35-39 ans	0	0	6	0	0	0	
31-34 ans	0	0	0	0	0	4	
15-16 ans	5	4	0	3	8	0	
12-14 ans	8	8	0	4	6	0	
10 ans	3	3	0	4	4	0	
Total :							120

Les parents des adolescents de Sainte-Foy (25 femmes et 11 hommes) sont pour la plupart des professionnels, des semi-professionnels et des travailleurs spécialisés. 10 femmes disent travailler à l'extérieur du foyer.

Les parents des adolescents de Saint-Sauveur (16 femmes et 8 hommes) sont pour la plupart des travailleurs spécialisés et semi-spécialisés. 4 femmes disent travailler à l'extérieur du foyer.

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

Les enquêteurs du corpus Deshaies ont réalisé trois types d'entrevues :

- L'entrevue informelle de groupe : trois ou quatre amis discutent de sujets qu'ils ont eux-mêmes choisis.
- L'entrevue formelle de groupe : l'enquêteur suscite des discussions portant sur les sujets suivants : sports, jeux, groupes de pairs (la structure du groupe, la façon dont il s'est formé, la stabilité des membres du groupe, les activités du groupe etc.), école, famille, langue, lectures, etc.
- L'entrevue individuelle : avec chaque membre du groupe ayant participé aux deux entrevues précédentes, un enquêteur réalise une nouvelle entrevue où l'on discute seul à seul de certains des sujets déjà abordés lors de l'entrevue formelle de groupe, ainsi que de sujets personnels.

Cette diversification des types d'entrevues confère au corpus une grande variété du point de vue des modes d'interaction. La durée des entrevues avec les adolescents est très variable. La durée moyenne des entrevues avec les parents est de trois heures. Le corpus en entier totalise 250 heures.

##### b) Traitement des données :

Toutes les entrevues des adolescents ainsi que celles de six parents sont disponibles sur support informatique (disquette, texte format .TXT accessible avec WordPerfect 4.1.). Le corpus est retranscrit en translittération avec restitution entre parenthèse des formes agglutinées ou élidées (exemple : [sesa:tab] est retranscrit

c'est sa (sur la) table. Des conventions de disposition ont permis de présenter toute l'information pertinente concernant les interruptions et les phrases énoncées par deux individus en même temps.

Des tables de concordance sont en cours d'élaboration. Il est actuellement possible d'accéder aux données lexicales à l'aide d'un logiciel lemmatiseur (fonctionnant en DBase III+). Il existe également une liste informatisée de 11,000 occurrences (le «dictionnaire») où figurent les unités lexicales et leur catégorie grammaticale.

Les bandes magnétiques et certains listages informatiques du corpus sont disponibles.

## Corpus 6

1. *Nom usuel du corpus* : CORPUS DU FRANÇAIS PARLÉ À OTTAWA-HULL  
*Sources* : Poplack (sous presse).
2. *Coordonnées institutionnelles du corpus* :

Corpus non publié, en dépôt à l'Université d'Ottawa. S'adresser à:

Shana Poplack  
Département de linguistique  
Université d'Ottawa  
Ottawa (Ontario)  
K1N 6N5  
Tél. (613) 564-6584

3. *Renseignements sur les données*

Le corpus consiste en 120 entrevues réalisées en 1982 dans 3 quartiers d'Ottawa à forte proportion francophones (Vanier, Basse-Ville et West-End) et deux quartiers de Hull, presque exclusivement francophones (Vieux-Hull et Mont-Bleu). Les informateurs ont été sélectionnés grosso modo d'après les principes

d'échantillonnage ayant présidé à la constitution du corpus Sankoff-Cedergren. Les informateurs sont tous francophones et sont natifs d'Ottawa ou de Hull ou y ont été élevés. Dans chaque quartier, on a interviewé 4 informateurs par tranche d'âge (2 de chaque sexe). Les tranches d'âge retenues sont : 15-24 ans, 25-34 ans, 35-44 ans, 45-54 ans, 55-64 ans et 65 ans et plus.

Répartition des informateurs d'après leur occupation :

	Ottawa			Hull	
	Vanier	Basse-Ville	West-End	Vieux-Hull	Mont-Bleu
professionnels	6	2	1	4	10
vente et services	0	2	3	1	1
employés de bureau	1	3	5	0	1
travailleurs spécialisés	6	5	5	3	5
travailleurs non spécialisés	4	6	5	8	4
chômeurs	6	3	3	6	0
ménagères	1	0	0	1	0
étudiants	0	3	2	0	2

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

Les entrevues sont d'une durée variant de 2 à 5 heures. Elles sont largement informelles et contiennent des récits, des chansons, des jeux, des discussions de groupe, etc. C'est le seul corpus québécois de langue orale où l'on retrouve d'importants échantillons de code-switching anglais/français. Avec 270 heures d'enregistrement, ce corpus est aussi le plus volumineux de tous.

##### b) Traitement des données :

La transcription des entrevues a été faite directement sur ordinateur dans l'orthographe standard, sans considération pour les particularités de prononciation. On a cependant tenu compte des variantes de prononciation qui impliquaient des changements morphologiques ou syntaxiques. On a respecté la syntaxe de



l'informateur. Les morphèmes élidés sont représentés par un symbole particulier. Tous les lexèmes analysés comme des anglicismes sont transcrits selon l'orthographe anglaise (ex: [file] «se sentir (bien ou mal)» est transcrit **feeler**).

Ce corpus est entièrement informatisé et apparaît, entre autres, sous forme de listes de fréquences, concordances, index, index inverse etc. On dispose aussi de rapports ayant été établis immédiatement après les entrevues, grâce auxquels il est possible de retrouver les parties les plus intéressantes du corpus.

## Corpus 7

### 1. *Nom usuel du corpus* : CORPUS MONTRÉAL 84

*Sources* : Thibault, Vincent et Rains 1986 ; Thibault et Vincent 1988 ; Vincent (communication personnelle).

### 2. *Coordonnées institutionnelles du corpus* :

Corpus non publié, en dépôt à l'Université de Montréal. On peut y avoir accès à l'adresse suivante, avec l'autorisation des responsables :

Laboratoire d'ethnolinguistique  
 Département d'anthropologie  
 Université de Montréal  
 C.P. 6128, Succursale A  
 Montréal (Québec)  
 H3C 3J7 Tél. 514-343-6572

Responsables : Pierrette Thibault, Département d'anthropologie, U. de Montréal  
 Diane Vincent, Département de langues et linguistique, U. Laval

### 3. *Renseignements sur les données*

Enregistrements effectués pendant l'été 1984 auprès de 72 informateurs originaires de Montréal. Plus de 90 des 120 informateurs et informatrices du corpus

Sankoff-Cedergren de 1971 ont été contactés. 60 d'entre eux ont accepté de se soumettre une seconde fois à l'entrevue, 13 ans plus tard. Parmi ces 60 informateurs, on compte 33 hommes et 27 femmes. 17 d'entre eux n'habitent plus à Montréal.

Répartition des informateurs selon l'âge :

28-38 ans	29
39-48 ans	8
49-58 ans	9
59-75 ans	14
	---
Total	60

Le corpus Montréal 1984 comprend aussi 12 entrevues faites auprès de jeunes âgés de 15 à 25 ans, ce groupe d'âge n'étant forcément plus représenté parmi les informateurs du Corpus Sankoff-Cedergren. Parmi ces 12 jeunes, on compte 6 hommes et 6 femmes, provenant de 3 quartiers résidentiels caractéristiques de Montréal : Outremont, Rosemont et Saint-Henri.

Les renseignements concernant les changements de statut socio-économique des informateurs depuis 1971 ne sont pas encore disponibles.

#### 4. Renseignements sur la méthode

##### a) Cueillette des données :

La méthode d'enquête du corpus Montréal 84 a été élaborée de façon à fournir des données pouvant être comparées à celles du Corpus Sankoff-Cedergren. Le protocole d'enquête est sensiblement le même qu'en 1971. Il s'agit d'entrevues semi-dirigées réalisées au domicile de l'informateur avec un magnétophone à cassettes visible. Les enquêteurs ont eu pour consigne de ne pas interrompre l'informateur et de l'inciter à parler le plus possible.

Le questionnaire est axé sur l'informateur, sur son mode de vie et l'évaluation qu'il en fait. L'entrevue proprement dite porte sur différents aspects de la trajectoire sociale de l'individu depuis 1971 : vie familiale, scolarité, vie

professionnelle, déplacements et déménagements, aspirations. À la fin de l'entrevue, les enquêteurs ont posé une série de questions portant sur les goûts et les activités des informateurs. Cette section comporte des questions plutôt fermées. Les répliques sont à peu près de la même longueur que celles que l'on retrouve dans le Corpus Sankoff-Cedergren. Les entrevues sont d'une durée moyenne de 90 minutes.

b) Traitement des données :

On dispose des enregistrements du corpus sur bobines et d'une transcription sur support informatique. Le corpus est transcrit en alphabet standard directement sur ordinateur. Des conventions orthographiques et diacritiques ont été développées pour indiquer certains phénomènes prosodiques, les tours de parole, les prononciations rares, etc. Des transcriptions complètes des entrevues et un dictionnaire de concordances seront disponibles à partir de 1988.

## Conclusion

Il existe bien sûr d'autres types de corpus oraux d'utilité directe pour le lexicologue et le lexicographe. Nous nous bornerons à citer ici, pour leur importance fondamentale, les travaux (du type «Atlas linguistique») de G. Dulong et G. Bergeron (1980), et de Th. Lavoie, G. Bergeron et M. Côté (1985), qui résultent d'enquêtes effectuées entre 1972 et 1978, auprès de témoins âgés de plus de 70 ans pour la plupart. Ces travaux recueillent presque exclusivement les mots reliés à la civilisation «traditionnelle» des Québécois, laissent de côté les mots grammaticaux, et finalement présentent des unités isolées, sans aucun contexte. C'est la raison pour laquelle, si intéressants qu'ils soient, nous ne les avons pas intégrés comme tels à notre répertoire des corpus québécois de langue orale.

Sans trop nous éloigner de notre propos, mentionnons pour conclure qu'il existe aussi un certain nombre de corpus de langue écrite où les informateurs ont été directement invités à produire du texte. Ces corpus n'ont pas encore été exploités par les lexicographes, qui fondent toujours leurs études presque uniquement sur des dépouillements d'imprimés (livres, revues, journaux, etc.). Citons, à titre

d'exemple, l'intéressant corpus constitué par l'anthropologue Pierre Maranda : LA MENTALITÉ QUÉBÉCOISE (MENTAQ), en dépôt au Département d'anthropologie de l'Université Laval. Il s'agit d'un corpus écrit obtenu à l'aide de courtes séries de mots stimuli, à partir desquelles l'informateur est invité à produire des suites lexicales associatives puis à rédiger un récit. La totalité des données de ce corpus est saisie sur support informatique, ce qui laisse entrevoir la possibilité de comparaisons entre corpus écrits et corpus oraux.

Constatons finalement que l'on dispose, sur le parler des Québécois, de beaucoup plus de données qu'on ne le croit généralement, données dont l'exploitation systématique est à peine amorcée (pour des remarques complémentaires sur l'utilisation de ces corpus en lexicographie, voir Poirier 1987, surtout annexe II).

*Lionel Boisvert*  
*Université Laval (Québec)*

*Paul Laurendeau*  
*Université York (Toronto)*

## Références

- BEAUCHEMIN, N. (1972) *Le questionnaire*, Document de travail n°1, 46 p.
- BEAUCHEMIN, N. (1977) *Données sociologiques*, Document de travail n°11, 54 p.
- BEAUCHEMIN, N. et P. Martel (1979) *Vocabulaire fondamental du québécois parlé*, document de travail n° 13, 269 p.
- BEAUCHEMIN, N. et Alii, (1983) *Concordance du corpus de l'Estrie: formes, fréquences et contextes*, document de travail n°19, environ 2,000 p. en 5 micro-fiches, avec introduction de 10 p. brochées et présentées en pochette de 15.5 sur 21.5 cm.
- BEAUCHEMIN, N., P. Martel et M. Théoret (1973-1981) *Échantillons de textes libres I à VI*, Université de Sherbrooke, Faculté des arts, 6 tomes.
- BEAUCHEMIN, N., P. Martel et M. Théoret (1983) *Vocabulaire du québécois parlé en Estrie : fréquence, dispersion, usage*, document de travail n° 20, 303 p.
- BLANCHE-BENVÉNISTE, Cl. et C. Jeanjean (1987) *Le français parlé — transcription et édition*, Didier érudition, Paris, 264p.
- DESHAIES, D. (1981) *Le français parlé dans la ville de Québec, une étude sociolinguistique*, CIRB, Université Laval, Publication G-1, 118 p.
- DORAN, M., L. Drapeau et C. Lefebvre (1982) «Le projet Centre-Sud: corpus et méthodologie» C. Lefebvre (dir.) *La syntaxe comparée du français standard et populaire : approches formelles et fonctionnelles*, tome 2, Office de la langue française, coll. Langues et sociétés, pp.401-420.
- DUGAS, A. (1986) *Le corpus Bibeau-Dugas*, document interne, Université du Québec à Montréal, 11 p.
- DULONG, G. et G. Bergeron (1980) *Le parler populaire du Québec et de ses régions voisines. Atlas linguistique de l'Est du Canada*, Gouvernement du Québec, Ministère des communications en coproduction avec l'Office de la langue française, 10 vol.
- LAVOIE, Th., G. Bergeron et M. Côté (1985) *Les parlers français de Charlevoix, du Saguenay, du Lac-saint-Jean et de la Côte-Nord*, Gouvernement du Québec, Office de la langue française, coll. Les publications du Québec, 5 vol.

- POIRIER, Cl. (1987) «Problèmes et méthodes d'un dictionnaire général du français québécois», *Revue québécoise de linguistique théorique et appliquée* (à paraître).
- POPLACK, S., (sous presse) «The care and handling of a mega-corpus: the Ottawa-Hull French Project», R. Fasold and D. Schiffrin (dir.), *Language variation and change*, Benjamins, Amsterdam.
- SANKOFF, D. et G. Sankoff (1973) «Sample Survey methods and computer assisted analysis in the study of grammatical variation», R. Darnell (dir.) *Canadian languages in their Social context*, Linguistic Research, Edmonton, pp.7-64.
- SANKOFF, D., G. Sankoff, S. Laberge et M. Topham (1976) «Méthode d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale», *Cahiers de linguistique de l'Université du Québec*, n° 6, *La sociolinguistique au Québec*, pp.85-125.
- THIBAUT, P., D. Vincent et C. Rains (1986) «Les locuteurs du Corpus Sankoff-Cedergren 13 ans plus tard», communication présentée à l'ACFAS, Montréal, (à paraître).
- THIBAUT, P. et D. Vincent (1988) «La transcription ou la standardisation des productions orales», *LINX — Analyse grammaticale de corpus oraux*, Centre de recherches linguistiques de Paris X, Nanterre, pp 20-33.