CANADIANA ROMANICA

publiés par
Hans-Josef Niederehe et Lothar Wolf

Volume 22

FRANÇAIS DU CANADA – FRANÇAIS DE FRANCE

Actes du septième Colloque international
de Lyon, du 16 au 18 juin 2003

Édités par Brigitte Horiot
avec la collaboration de Chiara Bignamini-Verhoeven

Max Niemeyer Verlag
Tübingen 2008

À la mémoire de M^gr Pierre Gardette (juin 1973–juin 2003)

Bibliografische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über *http://dnb.ddb.de* abrufbar.

ISBN 978-3-484-56022-2 ISSN 0933-2421

© Max Niemeyer Verlag, Tübingen 2008
Ein Imprint der Walter de Gruyter GmbH & Co. KG
http://www.niemeyer.de
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Gedruckt auf alterungsbeständigem Papier
Druck und Einband: AZ Druck und Datentechnik GmbH, Kempten

Table des matières

Brigitte Horiot, *Présentation* 1

I. La recherche lexicographique 3

Brigitte Horiot, *Le rôle de M^gr Gardette dans la naissance des études lexicologiques au Québec* 5

Hélène Cajolet-Laganière, Geneviève Labrecque, Pierre Martel, Chantal-Édith Masson, Louis Mercier et Michel Théoret, *Dictionnaires usuels du français et Banque de Données Textuelles de Sherbrooke (BDTS) : convergence et divergence des nomenclatures* 9

Annette Paquot, *Heurs et malheurs des dictionnaires au Québec : des débats révélateurs* 29

II. Langue et idéologie 37

Lionel Meney, *Langue et idéologie : aspect du champ linguistique québécois* 39

III. L'apport des atlas linguistiques 49

Jean-Claude Bouvier, *Quelques considérations sur le vocabulaire de la « terre cultivée » dans les parlers québécois et les parlers gallo-romans* 51

Liliane Rodriguez, *Le rôle des Atlas régionaux dans la description du français au Manitoba* 57

Naomi Statkewich-Maharaj, *La coccinelle et ses désignations lexicales dans le Sud de la France* 67

Catherine Bougy, *Variations graphiques et particularités dialectales dans les deux manuscrits du* Roman du Mont Saint-Michel *de Guillaume de Saint-Pair (vers 1155)*.. 81

Chiara Bignamini-Verhoeven, *Analyse linguistique et stylistique des franco-canadianismes dans* La Montagne Secrète *de Gabrielle Roy* 105

IV. La langue dans l'histoire et dans l'espace 135

Jean-Denis Gendron, *Aperçu sur la formation et l'évolution de la prononciation du français au Québec, des origines à nos jours* 137

Lothar Wolf, *Le mauvais usage dans le royaume de France au 17^e siècle et ses survivances au Canada* 151

V. Les enquêtes sociolinguistiques 159

Patrice Brasseur, *Les compétences linguistiques des jeunes locuteurs franco-terre-neuviens* 161

Ursula Reutner, *Aspects d'une comparaison sociolinguistique entre le Québec et les Antilles françaises* 183

Louise Péronnet et Sylvia Kasparian, *Le français standard acadien (à l'oral). Analyse des prépositions : procédés de variation*..... 199
Fabrice Jejcic, *Pratiques de l'écrit et perception de la norme : une enquête en France et au Canada en 2002*..... 209
Liselotte Biedermann-Pasques, *Bilan des rectifications de l'orthographe : une enquête dans la francophonie (France-Canada 2002-2003)*..... 233

VI. Chants folkloriques des deux côtés de l'Atlantique 257

Michèle Gardré-Valière et Michel Valière, *Chansons en regard*..... 259

VII. Conclusions du colloque..... 273

Jean-Claude Bouvier, *Conclusions*..... 275

Liste des participants..... 279

BRIGITTE HORIOT
Université Lyon III

Présentation

Tout comme les colloques antérieurs, cette 7ème rencontre internationale a eu pour premier objectif de réunir les principaux chercheurs qui s'intéressent à la variation géographique du français, et plus particulièrement à la comparaison de ses variétés canadiennes et françaises. Elle a rassemblé une trentaine de spécialistes de la dialectologie, de la lexicologie et de la lexicographie françaises, originaires de France, d'Allemagne et du Canada, ainsi que des étudiants du 3ème cycle.

Mis à part le premier colloque (Trèves, 1985), dont l'objectif était de faire un tour d'horizon de l'ensemble des problèmes qui touchent le Canada et la France, chaque colloque a été thématique et le 7ème colloque n'a pas échappé à cette tradition. Le 1er thème proposé (l'influence de Gardette dans les travaux lexicologiques du français du Canada) doit son origine au fait que, primitivement, le Centre d'Études Linguistiques avait l'intention de marquer le trentième anniversaire de la mort de Mgr Gardette (22 juillet 1973) par un colloque ou une journée d'étude qui aurait mis en relief le rôle joué par Gardette dans les études linguistiques. Les participants au colloque de Sherbrooke ont estimé qu'il était possible de regrouper dans un seul colloque la 7ème rencontre « français du Canada – français de France » et l'hommage à Mgr Gardette. Le comité scientifique, composé des organisateurs des six colloques précédents :
Hans-Josef Niederehe, Université de Trèves (1er colloque, 1985)
Lothar Wolf, Université d'Augsbourg (1er et 3ème colloques, 1985 et 1991)
Brigitte Horiot, Université Lyon III (2ème colloque, 1988)
Thomas Lavoie, Université de Chicoutimi (4ème colloque, 1994)
Marie-Rose Simoni-Aurembou, CNRS (5ème colloque, 1997)
Hélène Cajolet-Laganière, Université de Sherbrooke (6ème colloque, 2000)
Pierre Martel, Université de Sherbrooke (6ème colloque, 2000)
Louis Mercier, Université de Sherbrooke (6ème colloque, 2000)
a proposé que les communications se rattachent aux deux thèmes suivants :
L'influence de Gardette dans les travaux lexicologiques du français du Canada ;
L'exploitation des atlas linguistiques régionaux.

Placé sous la présidence du professeur Jean-Claude Bouvier, ancien président de l'Université de Provence (Aix-Marseille 2), le colloque de Lyon a donné lieu à dix-neuf communications qui se rattachent à six sous thèmes. Ces *Actes* regroupent dix-sept de ces communications ainsi que les conclusions du colloque présentées par Jean-Claude Bouvier. Par suite de circonstances douloureuses, Chantal-Edith Masson n'a pas pu nous donner sa communication (« La tâche dictionnairique : une tâche floue ») mais, à la demande de Pierre Martel, son nom a été joint aux auteurs de la communication « Dictionnaires usuels du français et Banque de Données Textuelles de Sherbrooke (BDTS) : convergence et divergence des nomenclatures ». Béatrice Bagola a jugé que sa communication - « La place

des études franco-canadiennes en Allemagne au seuil du 21ème siècle » - ne pouvait pas être publiée en l'état.

La tenue du colloque a été rendue possible grâce à l'appui financier de l'Université Lyon III, ce qui a permis au Centre d'Études Linguistiques de cette université d'organiser, en plus du banquet au château du Vivier à Ecully (Institut Paul Bocuse), une visite à la petite ville de Condrieu, située aux confins de trois provinces : Dauphiné, Forez, Lyonnais, à 45 km de Lyon.

Il m'est agréable d'exprimer tous mes remerciements à Mme Chantal Caranzan, la secrétaire et bibliothécaire du Centre d'Études Linguistiques, aux doctorantes du CEL, tout particulièrement Incarnacion Perrier et Chiara Bignamini. C'est à cette dernière, devenue Mme Verhoeven, que nous devons la préparation de ces *Actes*.

L'honnêteté aurait voulu que le nom de Chiara Bignamini-Verhoeven figure en premier, en qualité d'éditeur, sur la couverture du volume des Actes. Malheureusement, la tradition exige que le nom de l'éditeur renvoie à la personne chargée de la responsabilité scientifique et non à celle qui assure tout le travail de mise en pages des articles et de contact avec les divers auteurs. Que Chiara Bignamini-Verhoeven trouve dans ces quelques mots l'expression de ma profonde gratitude.

I. La recherche lexicographique

BRIGITTE HORIOT
Université Lyon III

Le rôle de M^gr^ Gardette dans la naissance des études lexicologiques au Québec

Il n'est pas dans les habitudes qu'un organisateur de colloque fasse une communication et je m'excuse de déroger à cette règle, d'ailleurs non écrite. Dans une lettre adressée le 11 janvier 2002 aux membres du comité scientifique, comité composé des organisateurs des six précédents colloques, je confirmais la tenue du 7ème colloque international « français du Canada – français de France » à Lyon, en juin 2003, et j'ajoutais :

> Nous serons alors à quelques semaines du 30ème anniversaire de la mort de M^gr^ Gardette ; j'avais l'intention de marquer cet anniversaire par un colloque mettant en relief le rôle joué par Gardette dans les études linguistiques. A Sherbrooke nous avons estimé qu'il était possible de regrouper dans un seul colloque la 7ème rencontre "français du Canada – français de France" et l'hommage à M^gr^ Gardette. Les thèmes qui se dégagent de ce regroupement sont clairs :
> 1. l'influence de Gardette dans les travaux lexicologiques du français du Canada ;
> 2. l'exploitation des atlas linguistiques régionaux.

Permettez-moi de vous décrire ou, pour certains, simplement de vous rappeler, le rôle tenu par M^gr^ Gardette dans la naissance des études lexicologiques au Québec. Pour cela il nous faut remonter à l'année 1952, à un coup de cœur du recteur des Facultés Catholiques de Lyon pour les Canadiens du Québec.

1. Laval 1952-1953

En 1952 M^gr^ Gardette, alors recteur des Facultés Catholiques de Lyon, avait assisté aux fêtes du centenaire de la fondation de l'Université Laval. Dans son discours de rentrée solennelle, le 12 novembre 1952, devant les autorités civiles, militaires et religieuses de la région lyonnaise, il consacrait plus de trois pages à dire son admiration pour Québec et son université.

> Sait-on assez en France tout ce que Laval, et sa fille l'Université Catholique de Montréal, ont fait pour maintenir et développer dans ce petit peuple de quatre millions de Canadiens du Québec la culture française ?[1]

L'année suivante, M^gr^ Gardette retournait à Québec, mais à titre de dialectologue cette fois-ci, pour prononcer quelques conférences et, surtout, faire des enquêtes en compagnie de M^gr^ Félix-Antoine Savard et de Luc Lacourcière. Rentré en France, il rédigeait et publiait dans la

[1] *Bulletin des Facultés Catholiques de Lyon*, 74ème année, nouvelle série n° 13, juillet-décembre 1952, p. 149.

Revue de Linguistique Romane qu'il dirigeait depuis un an, un article inspiré par ses enquêtes et intitulé : « Pour un dictionnaire de la langue canadienne » (Gardette 1954). C'était un véritable programme de recherche qu'il traçait à l'usage des chercheurs québécois, rappelant également aux dialectologues français qu'ils avaient eux aussi leur apport à fournir.

> Une enquête complète permettra de délimiter la place de chacun de ces mots paysans sur le sol canadien et cette localisation nous apprendra, sans doute, beaucoup de choses sur l'histoire de ces mots depuis leur arrivée au Canada. Mais le premier chapitre de leur histoire, leur départ de France, leur étymologie en somme, il faudra le chercher en France (*art. cit* : 88).

Pierre Gardette dégageait les caractères essentiels du québécois, à la fois conservateur par ses mots issus des patois de France et de l'ancienne langue et novateur par ses emprunts aux parlers indiens et à l'anglais. Il démontrait également, à l'aide d'exemples précis, recueillis au cours de ses enquêtes, que le français québécois n'était pas uniforme, mais qu'il existait, à l'intérieur du pays, des différences lexicales importantes, que seul un questionnaire approprié permettrait de dégager. Dans ce même article, Gardette traçait un programme complet des travaux à entreprendre en vue d'un grand dictionnaire historique du français canadien. Dans l'article nécrologique que lui ont consacré Kurt Baldinger et Georges Straka, ces derniers n'hésitent pas à écrire que « le *Trésor de la langue française au Québec* [...] lui devra [à ce programme] ses fondements méthodologiques et son orientation scientifique » (*Hommage :* XXIV), jugement repris par les rédacteurs du *Dictionnaire historique du français québécois*, dans l'avant-propos du volume (DHFQ *:* XII).

Responsable de la coordination et de l'administration des atlas linguistiques de la France par régions, de 1962 jusqu'à sa mort en juillet 1973, Pierre Gardette avait hâte de voir s'établir les cartes des atlas linguistiques de l'Ouest (Normandie, Bretagne romane, Poitou-Saintonge), car il attendait beaucoup de ces atlas pour la connaissance du français parlé au Canada. En 1969, il avait accepté que Marie-Rose Aurembou, chargée de l'atlas de l'Ile-de-France-Orléanais, aille passer quelque temps à l'Université Laval pour travailler à l'élaboration d'un questionnaire destiné aux enquêtes de l'atlas linguistique de l'Est du Canada. A mon tour, en 1973, j'allais passer deux mois à Laval pour élaborer les premières cartes de cet atlas, travail, faut-il le préciser, qui n'eut pas de suite puisque, en dépit de son sous titre : *Atlas linguistique de l'Est du Canada*, *Le parler populaire du Québec et de ses régions voisines* n'a pas de cartes mais des articles « qui présentent d'une façon structurée les données compilées pour chaque question de l'enquête » (ALEC 1 : 32).

2. Strasbourg, l'enseignement de Georges Straka

En 1957 Mgr Gardette rencontrait, tout à fait par hasard, un jeune Canadien, Jean-Denis Gendron, qui était venu poursuivre des études à Paris et qui s'y morfondait. Aussitôt Pierre Gardette l'envoyait à l'Université de Strasbourg, chez son ami Georges Straka, et commençait alors une étroite collaboration entre Québec et Strasbourg et, par l'intermédiaire de cette dernière ville, Québec et Lyon. A la suite de Jean-Denis Gendron, nombreux furent les jeunes chercheurs de l'Université Laval, ainsi que de celle de Montréal, qui vinrent se former dans le domaine de la linguistique et de la philologie au Centre de Philologie et de Littératures Romanes que dirigeait Straka tandis que de son côté, ce dernier, à partir de 1961,

allait enseigner chaque année quelques mois à Laval. De ces rapports constants, les universités québécoises ont tiré profit en accueillant, au titre d'enseignants, leurs anciens étudiants devenus docteurs de l'Université de Strasbourg. Rapidement, en effet, des thèses furent soutenues à Strasbourg avec, le plus souvent, M^gr Gardette parmi les membres du jury. Ce denier invitait souvent le candidat à venir passer quelques mois à Lyon pour préparer l'édition de sa thèse. A cet effet Straka avait créé, en 1966, à l'intérieur de la collection « Bibliothèque française et romane » fondée en 1960 par Paul Imbs, et publiée par le Centre de Philologie et de Littératures Romanes de l'Université de Strasbourg, une nouvelle série, la série E : Langue et Littérature françaises au Canada, destinée à accueillir les thèses, les travaux sur le français canadien et sur la littérature canadienne d'expression française. Voici ce qu'écrivait G. Straka dans la préface du premier volume :

> Le premier volume de la série *Langue et Littérature françaises au Canada* est une bibliographie, qui semble exhaustive, de tout ce qui a été écrit sur le français canadien. Avant de publier des travaux originaux, nous voulions d'abord offrir aux romanistes un manuel bibliographique précédé d'une mise au point des principaux résultats de recherches antérieures. L'ouvrage de M. Dulong, qui contient aussi l'ancienne bibliothèque de Geddes et Rivard et rendra, sans aucun doute, d'excellents services aux futurs chercheurs, pourra être ainsi un point de départ de nouvelles recherches sur le français parlé au Canada.
>
> Mais dès à présent, nous prévoyons la publication d'autres volumes, dont les manuscrits sont achevés : un ouvrage important de M. Jean-Denis Gendron, *Tendances phonétiques du français parlé au Canada* ; des thèses soutenues à l'Université de Strasbourg par M. l'abbé Marcel Boudreault sur le rythme et la mélodie de la phrase franco-canadienne, par M. Gilles Lavoie sur les aspects phono-stylistiques de l'œuvre littéraire de M^gr Savard, par M. l'abbé René Charbonneau sur les voyelles nasales dans la prononciation populaire de Montréal. Voilà, pour le moment, notre programme d'édition. Il est suffisamment riche et la qualité de ces travaux nous permet d'ouvrir la nouvelle série de notre collection avec confiance (Dulong 1966 : XVI).

Depuis, d'autres ouvrages importants ont été publiés, citons simplement les thèses de Normand Beauchemin (*Recherches sur l'accent d'après des poèmes d'Alain Grandbois : étude acoustique et statistique,* 1970), de Marcel Juneau (*Contribution à l'histoire de la prononciation française au Québec : étude des graphies des documents d'archives*, 1972) et de Micheline Massicotte (*Le parler rural de l'Île-aux-Grues (Québec) : documents lexicaux*, 1978).

3. Les colloques

Et les Français ? Pendant que les Québécois se succédaient à Strasbourg puis, pour quelques-uns, à Lyon auprès de Pierre Gardette, pour préparer l'édition de leurs thèses, que faisaient les dialectologues français ? Leurs atlas, bien sûr, mais ils avaient aussi l'occasion de rencontrer à Strasbourg, lors de colloques organisés par Georges Straka et présidés par Pierre Gardette, les chercheurs québécois et des amitiés se nouaient. Pour ma part j'ai eu la chance, durant la préparation de ma thèse 3^ème cycle, tout en travaillant au CNRS, dans l'équipe des atlas, et dans le cadre de l'institut de linguistique romane dirigé par Pierre Gardette dont j'avais été l'étudiante, de pouvoir passer chaque mois une semaine à Strasbourg.

4. L'après Gardette

Une fois Pierre Gardette décédé en juillet 1973, puis Georges Straka à la retraite en 1979, l'élan insufflé par Gardette allait-il être stoppé ? Vous connaissez la réponse, le bilan est éloquent tant au niveau des publications que des rencontres scientifiques.

- Des publications d'enquêtes sous forme d'articles : l'ALEC et le PFC ou de cartes : l'ALVMA *;*
- Un dictionnaire : le DHFQ ;
- Un ouvrage d'ensemble : *Français de France et français du Canada. Les parlers de l'Ouest de la France, du Québec et de l'Acadie ;*
- La création, en 1985, et l'institutionnalisation des colloques « français du Canada – français de France », sans parler des colloques « Dialectologie et littérature du domaine d'oïl occidental » pour lesquels, suivant le désir de leur créateur, le professeur René Lepelley de l'Université de Caen, en 1981, l'espace concerné, « c'était l'ouest de la France romane, avec son prolongement canadien. »[2]

Bibliographie

ALEC : Dulong, Gaston et Gaston Bergeron (1980) : *Le Parler populaire du Québec et de ses régions voisines. Atlas linguistique de l'Est du Canada*. Québec : Office de la langue française (Etudes et dossiers), 10 volumes.

ALVMA : Péronnet, Louise, Rose Mary Babitch [et alii] (1998) : *Atlas linguistique du vocabulaire maritime acadien*. Sainte-Foy : Presses de l'Université Laval.

DHFQ : Poirier, Claude [éditeur scientifique] (1998) : *Dictionnaire historique du français québécois*. Sainte-Foy : Presses de l'Université Laval (Trésor de la langue française au Québec).

Dulong, Gaston (1966) : *Bibliographie linguistique du Canada français* de James Geddes et Adjutor Rivard (1906) continuée par Gaston Dulong. Québec : Presses de l'Université Laval / Paris : Librairie C. Klincksieck (Bibliothèque française et romane, série E : 1).

Français de France et français du Canada. Les parlers de l'Ouest de la France, du Québec et de l'Acadie (1995), sous la direction de Pierre Gauthier et Thomas Lavoie. Université Lyon III Jean Moulin : Centre d'études linguistiques (série dialectologie 3).

Gardette, Pierre (1954) : « Pour un dictionnaire de la langue canadienne ». *Revue de Linguistique Romane* XVIII : 85-100. Article reproduit dans : Pierre Gardette, *Etudes de géographie linguistique*. Strasbourg : Société de Linguistique Romane, 1983, 787-802.

Hommage à la mémoire de Pierre Gardette. Mélanges de Linguistique et de Philologie romanes (1975). Lyon – Strasbourg : Société de Linguistique Romane.

PFC : Lavoie, Thomas, Gaston Bergeron et Michelle Côté (1985) : *Les parlers français de Charlevoix, du Saguenay, du Lac-Saint-Jean et de la Côte-Nord*. Québec : Gouvernement du Québec (Les Publications du Québec), 5 volumes.

[2] Avant-propos des *Actes* publiés en 1983, à Caen, par René Lepelley, et qui forment le n° 15 des *Cahiers des Annales de Normandie* sous le titre : *Dialectologie et littérature du domaine d'oïl occidental.*

HÉLÈNE CAJOLET-LAGANIÈRE, GENEVIÈVE LABRECQUE,
PIERRE MARTEL, CHANTAL-ÉDITH MASSON,
LOUIS MERCIER et MICHEL THÉORET
Université de Sherbrooke

Dictionnaires usuels du français et Banque de Données Textuelles de Sherbrooke (BDTS) : convergence et divergence des nomenclatures

Introduction

La présente recherche s'inscrit dans le cadre des travaux du Centre d'analyse et de traitement informatique du français québécois, et plus particulièrement du groupe de recherche FRANQUS de l'Université de Sherbrooke, responsable du projet « Nomenclatures, description et application dans les technologies de l'information et de la communication ».

Les responsables du projet prévoyaient, d'une part, la conception de tous les outils linguistiques et informatiques visant à l'élaboration d'une nomenclature originale du français en usage au Québec, essentiellement de son usage standard, et d'autre part, la rédaction des articles lexicographiques des mots retenus à la nomenclature.

La méthodologie employée pour élaborer cette nomenclature originale est basée sur la linguistique de corpus ; la sélection des quelque 60 000 mots à traiter dans le dictionnaire a été précédée de l'indexation et de la lemmatisation d'une banque de données textuelles représentative du français en usage au Québec. La liste des vocables tirés de cette banque a par la suite été validée à partir des nomenclatures de douze dictionnaires usuels français et québécois et enrichie grâce à un certain nombre de vocabulaires dans les domaines notamment de l'éducation, des relations du travail, de la faune et de la flore, etc., domaines susceptibles de contenir un certain nombre de particularités québécoises.

L'objectif de cette recherche est de livrer les résultats des différentes comparaisons entre les principaux dictionnaires usuels du français et la nomenclature extraite de la Banque de données textuelles de Sherbrooke (BDTS).

Dans la première partie de l'article, nous cernerons le vocabulaire commun à tous les dictionnaires et à la BDTS. Puis nous examinerons le vocabulaire commun à l'un ou l'autre dictionnaire et à la BDTS. Par la suite, nous analyserons le vocabulaire spécifique de l'un ou l'autre dictionnaire de même que celui spécifique de la BDTS, donc absent des dictionnaires sélectionnés aux fins de notre étude. Dans la deuxième partie de l'article, nous analyserons plus en détail deux tranches de vocables particulièrement intéressantes dans la perspective québécoise, soit un sous-ensemble du vocabulaire de la flore et de la faune et la série des emprunts critiqués à l'anglais.

I. Première partie

1. La composition de la BDTS

La BDTS est constituée de textes représentatifs des usages du français québécois. Il ne s'agit pas d'une représentativité au sens statistique du terme : la notion de *représentativité* de la BDTS doit plutôt être associée à la variété des textes qui la composent et qui reflètent la langue générale (orale et écrite, dans différentes situations de communication). La ventilation des pourcentages associés aux diverses catégories de textes correspond à l'objectif visé : donner accès aux différents registres de langue, notamment le niveau standard, actuellement en usage au Québec. Pour la notion de *français québécois*, nous faisons référence aux textes que nous stockons aux fins de diverses analyses et qui ont été rédigés dans l'espace géographique du Québec. Nous ne faisons nullement allusion à quelque caractère différentiel de ces textes par rapport à des textes rédigés en France, par exemple. Nous considérons ces textes pour eux-mêmes et dans leur totalité.

La BDTS compte actuellement plus de 52 millions de mots tirés de quelque 15 000 textes contemporains couvrant les années 1960 à 2004. En voici la composition :

Tableau I : Composition de la BDTS (2005)

	Occurrences	**Formes**	**Textes**	**% d'occurr.**
Langue spécialisée	22 562 548	---	2454	43,3 %
Langue littéraire	12 810 019	126 675	230	24,6 %
Langue journalistique	12 576 189	---	11 816	24,1 %
Langue didactique	2 245 341	49 918	65	4,3 %
Langue orale	1 906 787	34 889	590	3,7 %
Total	52 100 884	255 213	15 155	100 %

Au moment de la présentation de notre communication à Lyon, en juin 2003, la BDTS comptait 37 millions de mots tirés de quelque 10 000 textes représentatifs des cinq mêmes types de discours. Les données qui suivent sont tirées de ce noyau de base de la BDTS.

2. La présentation des différentes nomenclatures

2.1. La nomenclature de la BDTS

L'élaboration de la liste des vocables à partir des formes attestées dans le noyau de base de 37 millions d'occurrences de la BDTS s'est faite au moyen d'une lemmatisation par association, grâce au lemmatiseur de l'Association des bibliophiles universels (ABU). L'ABU a rassemblé 286 textes (romans, récits ou autres), en version intégrale, provenant de cent auteurs essentiellement français. Ces textes mis sur support informatique sont accessibles à partir du site de l'ABU. Cela nous a permis de lemmatiser une bonne partie des formes de la BDTS. Nous avons par ailleurs dû réviser cette lemmatisation automatique, car les textes de l'ABU ne permettent pas une lemmatisation fiable pour toutes les formes soumises.

Nous avons en outre lemmatisé une à une toutes les formes non reconnues par les textes de l'ABU et celles spécifiques de la BDTS. Ce travail s'est fait manuellement à l'aide des divers outils dictionnairiques de manière à identifier précisément le lemme ou vocable et la catégorie grammaticale de ce dernier.

Cette lemmatisation a donné précisément 57 323 vocables attestés dans la BDTS (corpus de 37 millions d'occurrences), excluant les noms propres, les sigles, les homonymes.

2.2. Les dictionnaires sélectionnés

Aux fins de cette étude, nous avons comparé la nomenclature de la BDTS avec celles des 12 dictionnaires sélectionnés suivants :
- Le *Petit Robert,* version électronique, 2001 (PRÉ)
- Le *Larousse illustré, version électronique*, 2002 (PLI)
- Le *Dictionnaire du français Plus*, version imprimée, 1988 (DFP)
- Le *Dictionnaire québécois d'aujourd'hui*, version imprimée, 1993 (DQA)
- Le *Multidictionnaire de la langue française, version électronique*, 2001 (Multi) ; notons qu'il s'agit d'un dictionnaire de difficultés,

et 7 dictionnaires pour enfants :
- Le *Dictionnaire HRW et thésaurus*, version imprimée, 2000 (HRW)
- Le *Dictionnaire CEC Intermédiaire*, version électronique, correspondant à la version imprimée de 1999 (CECI)
- Le *Dictionnaire CEC Jeunesse*, version imprimée, 1999 (CECJ)
- Le *Robert Junior illustré*, version nord-américaine, version imprimée, 1994 (Rjun)
- *Mon premier dictionnaire français illustré*, version imprimée, 1999 (MPD)
- Le *Larousse des débutants*, version imprimée, 2000 (Ldéb)
- Le *Robert Benjamin*, version imprimée, 1997 (Rbenj)

Ces dernières nomenclatures proviennent toutes de dictionnaires destinés à des enfants ou à des jeunes et sont donc de taille réduite. Ces nomenclatures sont utiles pour connaître le vocabulaire « fondamental » ou restreint du français. Comme nous le verrons ci-dessous, les nomenclatures de ces sept dictionnaires scolaires ont été groupées en une seule.

3. L'uniformisation des nomenclatures

Avant de procéder à la comparaison des nomenclatures, il nous a fallu les uniformiser pour ne pas doubler des entrées identiques. De fait, en examinant les diverses nomenclatures, nous avons observé une grande diversité dans l'adressage des entrées. Nous avons donc dû les « toiletter » de telle sorte que les mêmes unités lexicales constituant les entrées des dictionnaires soient identiques à des fins de comparaison et puissent ainsi être comptabilisées comme des unités lexicales uniques et non comme des unités différentes.

Nous présentons ci-après les principales sortes de variantes trouvées lors de l'examen des entrées des dictionnaires lorsque nous les avons comparées les unes aux autres. Nous reprenons ici en partie une description déjà présentée dans les *Mélanges* offerts à M. Lothar Wolf (p. 479 et suivantes). Mentionnons en outre que dans l'article des *Mélanges*, comme il s'agissait d'analyser l'accroissement des nomenclatures de dictionnaires usuels, nous avions

exclu le *Multidictionnaire des difficultés de la langue française*, ce qui peut faire varier les données présentées dans les tableaux ci-dessous.

Les homonymes et les entrées redoublées ou même triplées

Le traitement des homonymes varie beaucoup d'un dictionnaire à l'autre tant dans la version imprimée qu'électronique. Certains dictionnaires les multiplient, alors que d'autres, au contraire, les restreignent au minimum. Là où certains lexicographes font une entrée, d'autres doublent, et parfois même triplent leur nombre. De manière à uniformiser les nomenclatures, nous avons décidé, aux seules fins de comparaison de cette étude, de réduire tous les homonymes à une seule entrée. Ainsi, la liste suivante : *1. voile, 2. voile, 3. voile* (3 entrées) devient après uniformisation : *voile* (une seule entrée).

Certains verbes et leur rection

Certains dictionnaires, contrairement à d'autres, précisent la rection du verbe en entrée. Par exemple, le CECJ donne en entrée : *aboyer (à, contre, après)* ; on trouve également dans d'autres dictionnaires *accéder à, accoutumer à, convenir de, provenir de*, etc. De manière systématique, la préposition a été supprimée afin de permettre la comparaison uniforme de ces mots : *aboyer*, *accéder, accoutumer, convenir, provenir*, etc.

Les entrées qui renvoient à une autre entrée

La variation morphologique donnée dans une entrée qui renvoie à une autre entrée a été préservée. Ainsi, l'entrée *vice-roi, vice-reine* a été divisée en deux entrées distinctes : *vice-roi, vice-reine*. De même *s'il vous plaît (te)*, du verbe *plaire*, a été comptabilisé deux fois : *s'il vous plaît* et *s'il te plaît*. Cette norme a été suivie systématiquement pour tous les cas similaires.

Les variantes orthographiques de toute nature

Le DFP est le seul dictionnaire à présenter des mots avec les lettres *a, o, i* et *u* surmontées d'une barre horizontale : *shogun* (*o* surmonté d'un trait long), alors que le PRÉ écrit : *shogun* sans trait au-dessus du *o*. Nous avons supprimé le trait horizontal conformément à la graphie des autres dictionnaires.

Nous avons noté de plus diverses variantes de type typographique ; par exemple, les points de suspension présents dans certains dictionnaires ont tous été supprimés : *Made in…* dans le DFP et *archi…*, *extra…*, *hyper…*, etc. dans le Rjun. Des centaines de cas similaires ont été relevés et uniformisés pour éviter de les considérer comme des formes spécifiques de certains dictionnaires.

Également, le DFP contient plusieurs entrées se terminant par un point d'exclamation, contrairement aux autres dictionnaires : *aïe !, bah !, basta !, chut !, hop !, oh !, ouf !,* etc. ; c'est aussi le cas dans le Ldéb : *ah !, bravo !, chut !,* etc. De manière à uniformiser les nomenclatures, le point d'exclamation placé à la suite des entrées a été enlevé dans tous les cas.

Quelques dictionnaires présentent de plus d'autres particularités graphiques. Le PLI, par exemple, inscrit les noms déposés avec des majuscules : *Jardinerie, Bureautique, Martini*, etc., alors que les autres dictionnaires ont choisi de les écrire avec des minuscules. D'autres cas sont également à souligner pour l'écriture des noms de fêtes : *Pentecôte* ou *pentecôte* ; *Halloween* ou *halloween*, etc. Nous avons systématiquement uniformisé l'adressage de ces entrées avec des minuscules.

Par ailleurs, d'autres variantes orthographiques de type « lexical » ont été notées : *alcootest* ou *alcotest, maffia* ou *mafia*, *base-ball* ou *baseball*, *acupuncture* ou *acuponcture*, *cédrière* ou *cèdrière*, *fiord* ou *fjord*, etc. Pour des raisons pratiques de repérage, nous n'avons pas uniformisé ces variantes inscrites à la nomenclature de quelques dictionnaires. Nous avons fait de même pour certaines variantes en nombre : *haillons* ou *haillon*, *broussailles* ou *broussaille*, etc. Dans tous les cas, ces variantes ont été conservées dans la nomenclature.

Les entrées cachées

Enfin, la version imprimée du PRÉ renferme des milliers d'entrées cachées. Par exemple, les mots *indestructibilité, indestructible* et *indestructiblement* sont des entrées cachées placées à l'intérieur de l'article *indestructible* ; par contre, dans la version électronique du même dictionnaire, ces mots figurent à l'index. Nous avons reporté toutes les entrées cachées dans la nomenclature de ce dictionnaire de manière à les faire figurer à la nomenclature du PRÉ, conformément à la version électronique du même dictionnaire et à l'ensemble des autres nomenclatures.

4. Les résultats : comparaison des nomenclatures générales[1]

Une fois cette opération terminée, nous avons procédé à la comparaison des nomenclatures ainsi uniformisées des dictionnaires retenus. Le résultat des comparaisons nous a fourni les données suivantes (le nombre des entrées pour chacun des dictionnaires) :

Les dictionnaires pour enfants :

Rbenj	:	5 231
Ldéb	:	6 000
MPD	:	14 631
Rjun	:	16 730
CECJ	:	17 114
CECI	:	26 091
HRW	:	27 224

Ces nomenclatures ont été fusionnées aux fins de la comparaison et fournissent un total de 30 027 vocables différents.

Le dictionnaire de difficultés :

Multi	:	28 239

Les dictionnaires usuels français et québécois :

DQA	:	31 215
DFP	:	49 218
PLI	:	55 032
PRÉ	:	55 564

[1] Geneviève Labrecque, dans sa thèse de doctorat, a étudié les apports et les limites de la BDTS au regard de la nomenclature et de la description lexicographique du français en usage au Québec. Une partie des résultats et des exemples de cette partie sont tirés de sa recherche.

Nous avons fusionné dans une seule liste toutes ces nomenclatures et nous avons obtenu 69 251 vocables différents présents dans au moins un dictionnaire. Dans le but de comparer la nomenclature de la BDTS avec celles des dictionnaires, nous avons comparé la liste de ces vocables à la liste des 57 323 vocables (noms communs) de la BDTS (corpus de 37 millions de mots).

La comparaison de la nomenclature de la BDTS avec celles des dictionnaires retenus a révélé 40 978 mots présents à la fois dans la BDTS et dans l'une ou l'autre des nomenclatures des dictionnaires. Ces quelque 41 000 mots constituent le noyau lexical commun aux dictionnaires et aux textes de la BDTS et représentent le vocabulaire partagé par l'ensemble des francophones, en tout état de cause entre Québécois et Français (selon le reflet de l'usage que donnent les dictionnaires usuels du français de France).

En revanche, quelque 16 345 mots sont spécifiques de la BDTS.

Tableau II : Comparaison des nomenclatures des dictionnaires avec celle de la BDTS

	Nombre de vocables	**% dans la BDTS**	**% dans les dictionnaires**
Nombre de vocables présents dans la BDTS et dans au moins un dictionnaire	40 978	71,5 %	59 %
Nombre de vocables propres à la BDTS	16 345	28,5 %	
Nombre de vocables propres aux dictionnaires	28 273		41 %

Il ressort que 71,5 % des vocables de la BDTS sont présents dans au moins un dictionnaire. Le reste, soit 28,5 % des vocables de la BDTS, est spécifique de la BDTS. Par ailleurs, 59 % des vocables des dictionnaires sont attestés dans la BDTS ; 41 % des vocables des dictionnaires sont donc absents de la BDTS.

4.1. Les vocables communs à la BDTS et aux dictionnaires

Parmi les 40 978 vocables communs à la BDTS et à au moins un dictionnaire sélectionné, 20 196 sont communs à la BDTS et aux six dictionnaires ; ils sont donc présents dans toutes les nomenclatures. Par contre, de ce nombre, quelque 3098 sont communs à la BDTS et à un seul dictionnaire.

Tableau III : Répartition des vocables communs à la BDTS et à au moins un dictionnaire

	Nombre de vocables
Vocables présents dans 6 dict. + BDTS	20 196
Vocables présents dans 5 dict. + BDTS	4819
Vocables présents dans 4 dict. +BDTS	4188
Vocables présents dans 3 dict. + BDTS	5610
Vocables présents dans 2 dict. + BDTS	3067
Vocables présents dans 1 dict. + BDTS	3098
Vocables communs PRÉ + BDTS	974
Vocables communs PLI + BDTS	1156
Vocables communs DFP + BDTS	336
Vocables communs DQA + BDTS	471
Vocables communs Multi + BDTS	72
Vocables communs ENF + BDTS	89

Nous observons que sur les 40 978 vocables communs à la BDTS et à au moins un dictionnaire, plus de la moitié (20 196) sont en outre communs à tous les dictionnaires et à la BDTS. Cette portion importante du vocabulaire commun à tous les dictionnaires et à la BDTS représente sans doute l'essentiel du vocabulaire de base commun à tous les francophones.

Les résultats montrent qu'environ 3100 vocables sont communs à la BDTS et à un seul dictionnaire ; la répartition de ces vocables se distribue *grosso modo* selon la taille des dictionnaires. Nous notons par ailleurs que les vocables communs à la BDTS et au PLI sont plus nombreux que ceux partagés entre le PRÉ uniquement et la BDTS, malgré la taille légèrement plus grande de ce dernier. Le vocabulaire spécialisé (notamment technique), commun au PLI et à la BDTS, explique sans doute, du moins en partie, ce constat.

Nous présentons ci-dessous des exemples de vocables communs à la BDTS et à un seul dictionnaire.

Vocables communs entre PRÉ et BDTS : *algicide*, *antiparasitaire*, *connotatif*, *ethnicité*, *harder*, *microgramme*, *polymérase*, *transdisciplinaire*, *tuyauteur*, *vermiculite*, etc.

Vocables communs entre PLI et BDTS : *agissement*, *baudelairien*, *biotechnologique*, *cognitiviste*, *contextualiser*, *évaluatif*, *flamboyance*, *métacognition*, *motomarine*, *réfutabilité*, etc.

Vocables communs entre DFP et BDTS : *concréter*, *criminalisation*, *débusquage*, *francisme*, *gestalt*, *hydraulicité*, *intrusif*, *nébulaire*, *perceptuel*, *phtalate*, etc.

Vocables communs entre DQA et BDTS : *amancher*, *avironneur*, *baboune*, *barachois*, *défoliateur*, *labradorien*, *marchette*, *renchausser*, *subventionnaire*, *vitement*, etc.

Vocables communs entre Multi et BDTS : *antibuée*, *contre-offre*, *compétitionner*, *déréglage*, *mégajoule*, *odonyme*, *porte-panier*, *stylicien*, *ultimement*, *vidéoscope*, etc.

Vocables communs entre ENF et BDTS : *cacasser*, *cupidon*, *folacine*, *hyperlien*, *incomestible*, *indispensablement*, *lessiveur*, *niacine*, *partagiciel*, *vélociraptor*, etc.

Si quelques termes communs à un dictionnaire et à la BDTS sont carrément spécialisés ou techniques (*algicide, polymérase, cognitiviste*, etc.), la plupart sont des mots de la langue générale qu'il peut être surprenant de ne pas voir partager par l'ensemble des dictionnaires (*connotatif, agissement, évaluatif*, etc.). Il n'est par ailleurs pas étonnant de constater que les dictionnaires québécois ajoutent des mots typiquement québécois, de registres différents (*hydraulicité, amancher, barachois, marchette*, etc.).

4.2. Les vocables propres aux dictionnaires

Parmi les 28 273 vocables spécifiques des dictionnaires, mais absents de la BDTS, 588 sont décrits dans les six dictionnaires et 11 714 ne sont décrits que dans un seul dictionnaire.

Tableau IV : Répartition des vocables propres aux dictionnaires et absents de la BDTS

	Nombre de vocables
Vocables présents dans 6 dictionnaires	588
Vocables présents dans 5 dictionnaires	847
Vocables présents dans 4 dictionnaires	1985
Vocables présents dans 3 dictionnaires	6362
Vocables présents dans 2 dictionnaires	6777
Vocables présents dans 1 dictionnaire	11 714
Vocables spécifiques du PRÉ	3747
Vocables spécifiques du PLI	3754
Vocables spécifiques du DFP	2316
Vocables spécifiques du DQA	921
Vocables spécifiques du Multi	520
Vocables spécifiques des ENF	454

Voici quelques exemples de vocables présents dans six dictionnaires et absents de la BDTS : *daurade, garde-boue, laverie, pentathlon, pince-nez, rhumatologue, unijambiste, vibromasseur*, etc. Ces exemples montrent que toute banque de textes, même de grande taille, ne saurait jamais comprendre l'ensemble du vocabulaire d'une langue. D'où la nécessité de valider la nomenclature de la BDTS grâce aux dictionnaires existants de manière à combler ses lacunes vocabulairiques.

Il faut toutefois noter que, parmi les 588 mots considérés comme étant présents dans les six dictionnaires mais absents de la BDTS, certains sont bel et bien attestés dans le corpus. Leur présence dans la liste de mots absents de la BDTS résulte de la méthodologie employée lors du processus de lemmatisation automatique des formes de la BDTS. En effet, la lemmatisation automatique n'a pas distingué toutes les formes homographes. Par exemple, les formes *panique* et *plainte* ont été lemmatisées uniquement sous le verbe (*paniquer* et *plaindre*) et non sous le nom (*panique* et *plainte*). Également, les formes *douillette* et *nouvelle* ont été lemmatisées uniquement sous l'adjectif et non sous le nom. Il importe donc de tenir compte des limites de la lemmatisation automatique des formes de la BDTS lors de l'analyse des résultats statistiques du tableau ci-dessus. Ce chiffre relativement élevé de 588 mots absents de la BDTS et présents dans tous les dictionnaires doit donc être pris avec réserve.

Voici cette fois quelques exemples de vocables propres à chacun des dictionnaires et absents de la BDTS.

Vocables spécifiques du PRÉ : *adultisme*, *échographique*, *mercatique*, *précognition*, *repose-bras*, *rétrochargeuse*, *stop-over*, *tortellini*, *téléopérateur*, etc.

Vocables spécifiques du PLI : *aérogel*, *astigmatique*, *bruxisme*, *chouchoutage*, *discompte*, *labéliser*, *néorural*, *sky-surfing*, *surinformer*, *zoothèque*, etc.

Vocables spécifiques du DFP : *abréviativement*, *actomyosine*, *blennorhagique*, *causaliser*, *coriacité*, *jardiniste*, *nativement*, *pigeonnage*, *tocographie*, *transaméricain*, etc.

Vocables spécifiques du DQA : *atocatier*, *autoneigiste*, *badmintonneur*, *carnavaler*, *couvre-matelas*, *débossage*, *épinettière*, *gogosse*, *richelain*, *work-out*, etc.

Vocables spécifiques du Multi : *antiéblouissant*, *archimillionnaire*, *colophon*, *disgression*, *émondement*, *oncologiste*, *pelable*, *porte-nom*, *stylique*, *trentièmement*, etc.

Vocables spécifiques des ENF : *audiocassette*, *butterscotch*, *clipvidéo*, *congédiable*, *cyberspatial*, *déneigeuse*, *gratuiciel*, *houspilleur*, *ornithoptère*, etc.

Nous constatons la présence de termes techniques et rares dans la langue générale (*bruxisme, actomyosine, tocographie*, etc.). D'autres appartiennent véritablement à la langue générale et devraient faire partie, croyons-nous, de la nomenclature d'un dictionnaire général et normatif du français au Québec (*astigmatique, jardiniste, zoothèque,* etc.). Il est naturel que certains soient caractéristiques du français au Québec : *transaméricain, autoneigiste, épinettière, déneigeuse*, etc.

Il peut être étonnant à première vue de constater que la nomenclature des dictionnaires pour enfants, de taille réduite comme nous l'avons vu, apporte quand même des vocables laissés pour compte à la fois dans les autres dictionnaires de grande nomenclature et dans la BDTS : *audiocassette, cyberspatial, ornithoptère*, etc. Il y a donc une certaine liberté exercée par les lexicographes, laquelle fait varier une part de la nomenclature (voir p. 496 des *Mélanges*).

4.3. Les vocables propres à la BDTS

Enfin, parmi les 16 345 vocables spécifiques de la BDTS, mais absents des dictionnaires, seuls 130 ont une fréquence supérieure à 100.

Tableau V : Répartition des vocables spécifiques de la BDTS et absents des dictionnaires

	Nombre de vocables
Fréquence 100 et +	130
Fréquence 10 à 99	1308
Fréquence 2 à 9	6074
Fréquence 1	8833

Nous présentons ci-dessous des exemples de mots spécifiques de la BDTS classés par fréquence.

Fréquence 100 et + : *contaminant*, *diplomation*, *ethnoculturel*, *fraie*, *homogénat*, *itinérance*, *pessière*, *racinaire*, *radiodiffuseur*, *supplantation*, *térawattheure*, *unilinguisme*, etc.

Fréquence 10 à 99 : *bascophone*, *écotourisme*, *laquaiche*, *mentorat*, *méthicilline*, *museur*, *particulaire*, *plaza*, *résinique*, *topolectal*, *visionnement*, etc.

Fréquence 2 à 9 : *acétaminophène*, *allomorphe*, *aquaforme*, *climatogramme*, *désoxycholique*, *gaminet*, *hétérosexiste*, *podiatre*, *praxéologie*, etc.

Fréquence 1 : *achrétien*, *électroactif*, *filmathèque*, *filmeuse*, *francitude*, *monocyclette*, *radiothon*, *télédélestage*, *zooarchéologie*, etc.

Notons au passage que ces derniers mots, à l'exception de *contaminant, fraie* et *supplantation*, sont soulignés en rouge dans le logiciel de traitement de textes *Word*, ce qui signifie que le dictionnaire machine que comprend ce logiciel ne prend pas en compte ces vocables, dont certains sont propres au français en usage au Québec.

On peut être étonné de voir que quelques mots, fréquents dans la BDTS, sont absents des dictionnaires existants du français : *diplomation, racinaire, radiodiffuseur*, etc. L'apport de la BDTS est essentiel pour une description adéquate et complète du français en usage au Québec. D'autres apports, de mots moins fréquents, ne sont pas moins intéressants : *écotourisme, allomorphe, télédélestage*, etc.

Il faut ici encore apporter certaines réserves quant à l'interprétation du nombre de mots propres à la BDTS. De fait, en examinant la liste des quelque 16 000 vocables propres à la BDTS, nous avons constaté que certains mots sont en réalité consignés dans les dictionnaires, mais en tant que sous-entrées. Seules les nomenclatures des dictionnaires (et non leur sous-nomenclature) ont été comparées avec la nomenclature de la BDTS. Par exemple, le mot composé *auteur-compositeur-interprète* ne devrait pas être considéré comme étant propre à la BDTS, puisqu'il se trouve en sous-entrée dans le *Petit Robert* sous *auteur*. Il nous faut donc tenir compte de ces observations afin d'évaluer le nombre de vocables réellement spécifiques de la BDTS. Ici encore, ce nombre doit être pris avec prudence.

Pour conclure cette partie, mentionnons que la BDTS montre l'utilisation, en contextes de près de 41 000 mots déjà consignés dans un ou plusieurs dictionnaires. La BDTS sert également à dégager le noyau de base du vocabulaire. En revanche, parmi toutes les nomenclatures des dictionnaires à l'étude, plus de 28 000 mots répertoriés dans un ou plusieurs dictionnaires sont absents de la BDTS, d'où l'importance de valider la nomenclature de cette banque avec les vocables retenus par les lexicographes, afin de combler les lacunes inévitables de toutes banques de textes. Enfin, plus de 16 000 noms communs sont présents dans la BDTS, mais absents des nomenclatures des dictionnaires. Ces dernières données, nous l'avons précisé précédemment, doivent par ailleurs être interprétées avec prudence et nuancées en fonction de la méthodologie utilisée. La fréquence des vocables dans la BDTS de même que leur dispersion dans les types de discours sont en outre très variées et doivent par conséquent être prises en compte pour déterminer l'importance à accorder à chacun de ces mots.

II. Deuxième partie

Dans cette deuxième partie de l'article, nous commentons les résultats obtenus à la suite de l'étude plus en détail de deux tranches de vocables particulièrement intéressantes dans la perspective d'une description globale du français québécois, soit un sous-ensemble du vocabulaire de la faune et de la flore, et la série des emprunts critiqués à l'anglais.

1. Examen d'un sous-ensemble du vocabulaire de la faune et de la flore : les noms d'arbres[2]

Nous porterons d'abord notre attention sur le sous-ensemble thématique des noms d'arbres et d'arbustes,[3] par lequel débutera le traitement lexicographique du vocabulaire de la faune et de la flore. Il faut préalablement préciser que les chiffres présentés ici sont encore provisoires puisque certaines vérifications mineures restent à faire, notamment en ce qui a trait au statut et à la vitalité de quelques dénominations spécifiquement québécoises. De plus, ces chiffres ne concernent que les dénominations simples (*amélanchier*, *bouleau*, *érable*, *épinette*, etc.) pouvant correspondre aux entrées-vedettes directement accessibles à la nomenclature des dictionnaires retenus pour la comparaison. Nous remettons à une étape ultérieure l'examen des dénominations complexes (du type *bouleau à papier*, *érable à sucre*, *épinette bleue*, etc.) dont il faudra aussi tenir compte pour évaluer globalement l'originalité et la richesse de la BDTS en ce qui a trait à ce sous-ensemble lexical.

Le sous-ensemble en cause est relativement restreint par rapport à l'ensemble global présenté précédemment : 475 unités (excluant les taxons botaniques de niveaux supérieurs, comme *acéracées*, *bétulacées*, etc, qui seront traités avec le reste des taxons techniques de familles, d'ordres et de classes). Une partie seulement de ces 475 unités sont bien intégrées à la langue générale et affichent une fréquence moyenne ; la plus large part d'entre elles, de fréquence basse, relèvent du vocabulaire plus spécialisé de la botanique.

Étant donné la diversité des espèces végétales qui croissent sur les cinq continents et la richesse de la taxinomie botanique française servant à désigner ces multiples espèces, nous savions au départ que les dictionnaires comparés ne retiendraient qu'une partie seulement des noms français servant à désigner des arbres, et en priorité ceux de la zone tempérée de l'hémisphère nord ; nous savions également que la BDTS renseignerait davantage sur les noms d'arbres du continent nord-américain que sur ceux des autres parties du monde.

1.1. Répartition générale des vocables entre dictionnaires et BDTS

Avant de comparer le contenu des dictionnaires avec celui de la BDTS, nous jetterons un coup d'œil à la répartition générale des vocables de notre sous-ensemble dans les nomenclatures de ces dictionnaires. Comme le montrent les données du tableau VI, 95 % des noms d'arbres en question sont attestés dans au moins une de ces nomenclatures, mais moins du tiers (30 %) est commun aux six nomenclatures. Par ailleurs, on peut observer une chute

---

2 L'examen de ce sous-ensemble a été réalisé par Louis Mercier.

3 Par souci de simplification, nous utiliserons désormais le mot *arbre* comme générique englobant *arbre* et *arbuste*.

sensible de pourcentage (– 13 % ou 61 éléments en moins) entre les séries de dénominations communes à deux (88 %) et à trois (75 %) nomenclatures, ainsi qu'une chute particulièrement importante (– 29 % ou 136 éléments en moins) entre les dénominations communes à trois (75 %) et à quatre nomenclatures (46 %).

Tableau VI : Comparaison du contenu des dictionnaires

Nombre de dénominations présentes	**Bilan détaillé**	**Bilan cumulatif (% sur 475)**
dans 6 dict.	142	142 (30 %)
dans 5 dict.	+ 35	177 (37 %)
dans 4 dict.	+ 42	219 (46 %)
dans 3 dict.	+ 136	355 (75 %)
dans 2 dict.	+ 61	416 (88 %)
dans 1 dict.	+ 35	451 (95 %)

Le tableau VII intègre la BDTS à la comparaison. On constate que, dans le sous-ensemble examiné, la proportion de données communes aux dictionnaires et à la BDTS (56 % de l'ensemble ou 59 % du contenu des dictionnaires) et celle des données spécifiques aux nomenclatures de dictionnaires (39 % de l'ensemble ou 41 % du contenu des dictionnaires) correspondent aux proportions globales présentées précédemment (v. le tableau II). Les données spécifiques à la BDTS sont relativement limitées (5 %), ce qui tient au fait que seules les dénominations simples sont prises en compte ici et que, de façon générale, les textes versés à la BDTS font relativement peu de place aux dénominations « populaires » ou régionales québécoises.

Tableau VII : Comparaison du contenu des dictionnaires et de la BDTS

Données communes		**Données spécifiques**		
267 (56 %)		**aux dictionnaires** 184 (39 %)		**à la BDTS** 24 (5 %)
BDTS + 6 dict.	136 (29 %)	dans 6 dict.	6 (1 %)	–
BDTS + 5 dict.	163 (34 %)	dans 5 dict.	14 (3 %)	–
BDTS + 4 dict.	185 (39 %)	dans 4 dict.	33 (7 %)	–
BDTS + 3 dict.	244 (51 %)	dans 3 dict.	111 (23 %)	–
BDTS + 2 dict.	256 (54 %)	dans 2 dict.	160 (37 %)	–
BDTS + 1 dict.	267 (56 %)	dans 1 dict.	184 (39 %)	–

1.2. Comparaison des sept nomenclatures comparées

1.2.1. Aspects quantitatifs

Nous nous intéresserons maintenant au contenu de chacune des nomenclatures comparées pour évaluer leur étendue ou richesse relative. Dans le tableau suivant, les sources sont classées de gauche à droite par ordre décroissant.

Tableau VIII : Richesse relative des nomenclatures comparées (1)

Liste globale	**Dict. à nomenclature plus riche**				**Dict. à nomenclature moins riche**		
	PRÉ	**PLI**	**DFP**	**BDTS**	**ENF**	**DQA**	**MUL**
475 noms	423	418	364	291	206	179	170
	89 %	88 %	77 %	61 %	43 %	38 %	36 %

Les nomenclatures du PRÉ et du PLI s'avèrent nettement les plus riches. La différence observée entre les deux est très faible (5 vocables, soit à peine 1 %), et il est intéressant de noter que c'est le dictionnaire de langue et non le dictionnaire encyclopédique qui donne la représentation la plus riche des noms d'arbres disponibles en français, un sous-ensemble thématique relativement spécialisé. Une démarcation sensible apparaît entre le DFP et les deux premiers dictionnaires : le DFP répertorie environ 10 % moins de dénominations que le PRÉ et le PLI. Dans le cas des dictionnaires suivants, c'est moins de la moitié par rapport aux deux répertoires les plus riches, d'où la chute importante observée plus haut entre les emplois communs à trois et à quatre dictionnaires.[4] Avec ses 61 %, la BDTS se situe clairement entre les deux blocs de dictionnaires.

Les données du tableau VIII ne tiennent pas compte des parties communes de ces différentes nomenclatures. C'est le cas des données du tableau suivant qui permettent de nuancer la comparaison.

Tableau IX : Richesse relative des nomenclatures comparées (2)

Liste globale	PRÉ	PLI	DFP	ENF	DQA	MUL	BDTS	
475	423	418	363	206	179	170	291	
	Entrées communes							
	aux dictionnaires (du dictionnaire le plus riche au moins riche)						**aux dictionnaires et à la BDTS**	
							BDTS / Dict.	
	PRÉ 423						258 / 423	(61 %)
	PRÉ + PLI = 394						251 / 394	(64 %)
	PRÉ + PLI + DFP = 344						237 / 344	(69 %)
	PRÉ + PLI + DFP + ENF = 196						170 / 196	(87 %)
	PRÉ + PLI + DFP + ENF + DQA = 158						151 / 158	(96 %)
	PRÉ + PLI + DFP + ENF + DQA + Multi = 142						136 / 142	(96 %)

À la lecture de ce tableau, on constate que la BDTS atteste 96 % des dénominations communes aux 6 nomenclatures de dictionnaires. Elle atteste également les deux tiers (64 %) des données communes aux PRÉ et PLI, et presque 70 % des données communes aux trois dictionnaires les plus riches. Ces pourcentages s'avèrent relativement élevés, compte tenu du contenu essentiellement québécois des textes relatifs aux espèces naturelles et à l'environnement qui ont été versés à la BDTS.

[4] Étant donné la nature différente du DQA (dictionnaire général à nomenclature réduite) et du Multi (dictionnaire de difficultés), on peut s'étonner de constater que ces deux ouvrages répertorient sensiblement le même nombre d'emplois.

1.2.2. Spécificités des dictionnaires et de la BDTS

Le tableau X passe en revue les 184 dénominations d'arbres qui sont répertoriées dans les dictionnaires mais ne sont pas attestées dans la BDTS. On observe effectivement que la plupart de ces dénominations réfèrent à des espèces "exotiques" (du point de vue québécois), c'est-à-dire qui n'appartiennent pas à la flore indigène et qui ne sont pas cultivées dans l'Est de l'Amérique du Nord : comme l'arganier d'Afrique du Nord, l'aleurite d'Asie, l'anone et l'anacardier d'Amérique tropicale.

Tableau X : Dénominations spécifiques des dictionnaires retenus

184 dénominations absentes de la BDTS mais présentes dans au moins 1 dictionnaire			
Présentes dans			
6 dict.	6	*cacaotier*, *caféier*, *pamplemoussier*, *papayer*, *sensitive*, *sycomore*	
5 dict.	8	*anacardier*, *avelinier*, *calebassier*, *goyavier*, *jujubier*, *kolatier*, *limettier*, *mirabellier*	
4 dict.	20	*balsamier*, *bergamotier*, *bigaradier*, *cédratier*, *ébénier*, *germandrée*, *kapokier*, *latanier*, *niaouli*, *paliure*, *palmiste*, *passerine*, *pitchpin*, *pomelo*, *prunellier*, *rotang*, *sapotier*, *sesbani*, *sophora*, *yeuse*	
3 dict.	77	*ailante*, *ampélopsis*, *anone*, *aucuba*, *bancoulier*, *boule-de-neige*, *brugnonier*, *buisson-ardent*, *busserole*, *campêche*, *chamaerops*, *cirier*, *copayer*, *coronille*, *cycas*, *doum*, *éléis*, *érythrine*, *fragon*, *frangipanier*, *gainier*, *gattilier*, *grisard*, *guignier*, *hautin*, *hièble*, *ilang-ilang*, *indigotier*, *jaborandi*, *laurier-tin*, *lavandin*, *lépidodendron*, *liquidambar*, *malpighie*, *manglier*, *mangoustan*, *mangoustanier*, *marsault*, *mélampyre*, *myrobolan*, *myroxylon*, *pandanus*, *phytéléphas*, *pilocarpe*, *plaqueminier*, *quassia*, *quebracho*, *quercitron*, *rauwolfia*, *ravenala*, *rocouyer*, *rondier*, *rônier*, *rouvre*, *sainbois*, *séné*, *sidéroxylon*, *sigillaire*, *simaruba*, *souchong*, *spondias*, *storax*, *strophantus*, *strychnos*, *styrax*, *suber*, *taxodium*, *térébinithe*, *upas*, *uva-ursi*, *vélani*, *vomiquier*, *wellingtonia*, *yèble*, *yohimbehe*, *zamier*	
2 dict.	49	*agnus-castus*, *aleurite*, *aliboufier*, *arabica*, *arole*, *azerolier*, *borasse*, *cajeput*, *calamite*, *canéficier*, *carambolier*, *cassier*, *colatier*, *courbaril*, *cubèbe*, *demi-tige*, *douçain*, *éphédra*, *farigoule*, *filao*, *gnète*, *grape-fruit*, *gratiole*, *griottier*, *hippophaé*, *icaque*, *icaquier*, *jojoba*, *jussiée*, *khat*, *lantanier*, *limonier*, *maurelle*, *médicinier*, *muscadine*, *myroxyle*, *orvale*, *passe-velours*, *piassava*, *putier*, *robusta*, *sagoutier*, *sipo*, *tallipot*, *tauzin*, *tchitola*, *washingtonia*, *ximénie*, *ypréau*	
1 dict.	24	PRÉ (12) : *alaterne*, *bauhinie*, *cascara*, *durian*, *iroko*, *mûron*, *pinastre*, *rambour*, *rhizophore*, *rosage*, *toute-bonne*, *turbith*	PLI (12) : *actinidia*, *arganier*, *avodiré*, *badamier*, *cocaïer*, *corossolier*, *ipé*, *lantana*, *margousier*, *moringa*, *néré*, *oléastre*

La dernière ligne du tableau montre que le PRÉ et le PLI sont les deux seuls dictionnaires à enregistrer des éléments exclusifs, une douzaine chacun. Ces ouvrages répertorient en outre quelques dénominations que la BDTS est la seule autre source à attester, à savoir quatre pour le PRÉ et cinq pour le PLI. Cette situation se présente également une fois en ce qui a trait au DFP[5] et au DQA. Les onze dénominations en cause sont présentées dans le tableau XI. Quatre de ces dénominations (*andromède*, *aralia*, *caryer* et *pommetier*) se démarquent par une fréquence supérieure à dix dans la BDTS. Les deux éléments spécifiques aux dictionnaires québécois (*andromède* et *pommetier*) sont particulièrement bien attestés. La fréquence 5 de *brimbelle*, nom régional de l'airelle en France, appelle une remarque : les 5

[5] Au moment de sa parution en 1988, le DFP présentait davantage d'éléments spécifiques par rapport au *Petit Robert* et au *Petit Larousse* ; mais la plupart de ces éléments ont été intégrés aux nomenclatures postérieures, notamment celle du DQA, paru quatre ans plus tard.

occurrences viennent en fait d'un texte métalinguistique québécois relatif au traitement lexicographique des régionalismes de la francophonie.

Tableau XI : Dénominations répertoriées dans 1 seul dictionnaire et attestées dans la BDTS

Dict.		**Dénominations et nombre d'attestations dans la BDTS**
PRÉ	4	*berbéris* (2 attestations), *brimbelle* (5 att.), *dracéna* (7 att.), *lambruche* (1 att.)
PLI	5	*aralia* (14 att.), *caryer* (88 att.), *cryptomeria* (2 att.), *pyracantha* (2 att.), *sal* (2 att.)
DFP	1	*andromède* (16 att.)
DQA	1	*pommetier* (38 att.)
ENF	0	
MUL	0	

Le tableau XII est consacré aux vingt-quatre dénominations qui sont exclusives à la BDTS. La moitié d'entre elles affichent une fréquence supérieure à dix et sont attestées dans trois sources différentes et plus.

Tableau XII : Dénominations spécifiques de la BDTS

Fréquence (F) et dispersion (D) dans la BDTS								
Dans plus de 5 sources			**Dans 3 sources**			**Dans 1 ou 2 sources**		
6 dénominations	F	D	7 dénominations	F	D	11 dénominations	F	D
kalmia	12	8	*némopanthe*	15	3	*clèthre*	7	2
ostryer	31	7	*sorbaria*	10	3	*parthénocisse*	2	2
weigelia	14	7	*arctostaphyle*	5	3	*diervillée*	4	1
physocarpe	13	7	*chalef*	5	3	*rhodora*	4	1
myrique	12	7	*lédon*	5	3	*céphalanthe*	3	1
caragana	23	6	*aronia*	4	3	*quinquefeuille*	3	1
			shépherdie	3	3	*célastre*	1	1
						coumarine	1	1
						dirca	1	1
						phellodendron	1	1
						ptéléa	1	1

Pour une large part, ces dénominations font référence à des arbres indigènes du Québec (*kalmia*, *lédon*, *myrique*, *ostryer* et *physocarpe*, par exemple) ou encore à des essences qui sont largement utilisées au Québec à des fins ornementales (*caragana*, *sorbaria* et *weigelia*, par exemple). Les contextes d'emploi stockés dans la BDTS, dont nous donnons ici quelques exemples, témoignent avec éloquence de cette valeur référentielle nord-américaine.

caragana — *Plusieurs arbres et arbustes ont aussi souffert du froid, dont les* ***caraganas*** *pleureurs, un des arbres-fétiches de nos banlieues.* (Gingra 1994 : C14)

myrique — *Un autre arbuste indigène à l'arôme balsamique, le* ***myrique*** *baumier (Myrica gale) était appelé par nos ancêtres « bois sent bon » ! Il offre un joli port semi-globulaire et des feuilles dentées vert grisâtre. Ses fruits et fleurs sont peu remarquables. Il tolère diverses conditions, mais il est surtout utile dans les coins humides.* (Hodgson 2001)

ostryer — *L'****Ostryer*** *de Virginie est un arbre commun des différentes érablières du Québec. Son écorce se détache en fines lanières de 8 à 12 mm de largeur et de 8 à 15 cm de longueur. Cette caractéristique permet de l'identifier aisément.* (Fortin et Famelart 1990 : 287)