PIERRE MARTEL et HÉLÈNE CAJOLET-LAGANIÈRE
Centre d'analyse et de traitement informatique du français québécois
Université de Sherbrooke

L'apport de la Banque de données textuelles de Sherbrooke : des nomenclatures enrichies

Depuis une quinzaine d'années, la description du français québécois s'est considérablement enrichie en raison de la parution de plusieurs dictionnaires et répertoires consacrés à cette variété de français. Mais cette description peut encore être améliorée, notamment grâce à l'exploitation de la Banque de données textuelles de Sherbrooke (BDTS). Dans cette contribution, nous montrerons comment la BDTS peut, de diverses façons, servir à compléter les différentes nomenclatures existantes du français de France et du français québécois. Nous nous intéresserons principalement aux unités lexicales appartenant au niveau soigné du français québécois (ou français standard en usage au Québec). Le mot « standard » s'applique ici à une catégorie de textes dont le niveau de langue est généralement soigné, par opposition principalement à des textes oraux de niveaux familier, vulgaire ou autre. Il est entendu que tous les emplois lexicaux ou sémantiques rencontrés dans les textes de la BDTS ne relèvent pas de la langue dite « standard », c'est-à-dire du bon usage du français québécois.[1]

1. La Banque de données textuelles de Sherbrooke

La Banque de données textuelles de Sherbrooke est un corpus informatisé et indexé de vingt millions de mots (septembre 2000), et elle est constituée de textes représentatifs des différents usages du français québécois. Il ne s'agit pas d'une représentativité au sens statistique du terme : s'agissant de la BDTS, la notion de « représentativité » doit plutôt être associée à la variété des textes qui la composent et qui reflètent la langue générale (orale et écrite) de même que la langue littéraire, journalistique et relativement plus spécialisée en usage au Québec dans différentes situations de communication.

Les textes que nous stockons aux fins de diverses analyses ont été rédigés dans l'espace géographique du Québec, d'où leur association à la notion de « français québécois ». 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 dans leur totalité. La BDTS est constituée de cinq grandes catégories de textes répartis comme suit :

- Dix pour cent de transcriptions de langue orale (10 %), comprenant des échantillons de langue parlée spontanée, de langue parlée non spontanée (des monologues, par exemple) et de transcriptions diverses (émissions radiophoniques, tribunes téléphoniques, témoignages devant les tribunaux, etc.). Cette série de textes oraux a été constituée de

[1] Sur la question des niveaux de langue du français québécois, v. Langlois, 1999 ; pour la notion de « français standard », v. Martel, 2001.

manière à représenter ce type de discours et à hiérarchiser les usages à l'aide d'un système approprié de marques.

- Vingt pour cent de textes journalistiques (20 %), comprenant des échantillons extraits de quotidiens (*La Presse*, *Le Devoir*, *Le Soleil*, etc.) et de périodiques spécialisés (*Québec Sciences*, *Interface*, *Franc-vert*, etc.). Ce sous-ensemble est tiré de textes qui ont une certaine influence sur l'usage linguistique de la population en général : ces textes reflètent en grande partie la langue commune ; ils permettent en outre de noter les mots, sens et expressions dont les gens ont besoin pour prendre connaissance des informations diffusées par les médias.
- Vingt pour cent de textes didactiques (20 %), comprenant des échantillons de textes tirés de manuels scolaires (de niveaux secondaire, collégial, universitaire) ; de logiciels informatiques (textes de présentation et guides de l'utilisateur) ; de manuels pour une formation technique, etc. Ces textes jouent souvent un rôle d'exemplarité ; il s'agit de documents généralement révisés, acceptés et valorisés comme outils de formation ; ils permettent de cerner une langue générale et une langue spécialisée de base dont les élèves ont besoin pour consulter et exploiter leurs manuels.
- Vingt pour cent de textes littéraires (20 %), comprenant des échantillons tirés de romans, d'essais, de pièces de théâtre, de chansons, de textes de poésie, de contes, etc. Ce sont des écrits qui reproduisent une langue parfois soignée, soutenue, mais parfois aussi une langue très familière (littérature populaire, « joualisante »).
- Trente pour cent de textes spécialisés (30 %), comprenant des échantillons de textes techniques, scientifiques, sociopolitiques, administratifs, environnementaux, juridiques, médicaux et autres. Ces écrits sont nombreux et très diversifiés ; la plupart d'entre eux reflètent une langue relativement soignée et valorisée ; ils fournissent des mots de langue générale et le vocabulaire de base de très nombreux domaines spécialisés.

La ventilation des pourcentages associés aux diverses catégories de textes correspond à l'objectif visé : avoir accès aux différents registres de langue, notamment le niveau standard, actuellement en usage au Québec. Mentionnons par ailleurs que la BDTS s'enrichit d'une manière constante et progressive selon une représentativité et une variété de textes en fonction des besoins identifiés et des tests de validation effectués. Par exemple, nous achevons la constitution d'un sous-ensemble de trois millions de mots représentatifs du vocabulaire des relations de travail (textes de loi, de règlement, de convention collective, etc.). Parallèlement, nous élaborons un autre sous-ensemble dans le domaine de la langue et de l'aménagement linguistique au Québec. Un troisième sous-ensemble porte sur des textes relatifs à la faune et à la flore du Québec, constitué en collaboration avec Louis Mercier, spécialiste de cette portion du vocabulaire québécois. Enfin, à la BDTS s'ajoutent les cédéroms *Actualités/Québec*, *Repère*, *Internet*, ainsi que le *Réseau des corpus lexicaux québécois* qui est diffusé sur le Web par le Secrétariat à la politique linguistique du Québec ;[2] c'est la documentation disponible en périphérie du corpus.

Une fois les 20 millions d'occurrences de la BDTS indexées, nous avons dressé la liste des 170 000 formes présentes. Par « forme », nous entendons l'une des variantes graphiques sous lesquelles peut se présenter un vocable. En langage informatique, les formes peuvent également correspondre à ce qui est compris entre deux espaces. Comme cette liste de formes n'est pas entièrement lemmatisée, nous ne pouvions la comparer à la nomenclature des

[2] Pour une description de ce réseau de corpus, qui intègre une partie des textes de la BDTS, voir la contribution de Normand Maillet.

dictionnaires papier ou électroniques existants. Il était cependant possible de comparer la liste de ces 170 000 formes de la BDTS à celle du dictionnaire-machine du logiciel de traitement de texte WordPerfect.

Cette comparaison faite, nous avons trouvé 100 000 formes communes aux deux ensembles et 70 000 formes spécifiques à la BDTS. À première vue, ce dernier chiffre peut sembler élevé par rapport aux 100 000 formes communes. Nous allons justement nous intéresser à la composition de ce sous-ensemble de formes spécifiques.

Parmi les 70 000 formes spécifiques indexées, on note, dans des proportions à peu près équivalentes, des noms communs et des noms propres. Nous sommes à traiter ces deux listes pour répartir toutes les formes en différentes sous-listes homogènes, prenant soin de classer à part les formes avec coquilles (*adésion*, *acroitre*, par exemple) et autres formes non régulières (*internalisée*, *surperposait*, par exemple). Ce travail est en cours d'analyse ; nous pouvons néanmoins présenter dès maintenant les quelques résultats suivants.

Nous précisons tout de suite que les exemples donnés ci-après correspondent aux formes extraites de la BDTS. Il est certain que la nomenclature d'un dictionnaire du français québécois inclura des éléments originaux de la BDTS, mais les choix devront être faits en fonction de la politique éditoriale. De même, un système uniforme et approprié de marques permettra une hiérarchisation des usages.

2. L'apport de la BDTS

2.1. Premier type d'apport : les noms propres

Une première catégorie de spécificités linguistiques québécoises présentes dans la BDTS s'impose d'emblée. Il s'agit des noms propres : noms de personnes, de lieux, gentilés, raisons sociales et marques de commerce, sigles, acronymes et autres dénominations. Cette grande catégorie comporte quelque 35 000 formes.

À titre d'exemples, voici quelques noms et prénoms qui ne semblent pas d'emploi commun chez tous les francophones puisqu'ils ne sont pas intégrés au dictionnaire-machine de WordPerfect : *Bouchard*, *Laporte*, *Larose*, *Tremblay*, *Sansfaçon*, etc. ; *Anthony*, *Kevin*, *Lynda*, *Mario*, *Ginette*, *Lysiane*, etc.

À ces exemples s'ajoutent des toponymes et gentilés de toutes sortes : *Boucherville*, *Drummondville*, *Estrie*, *Gaspésie*, *Granby*, *Lennoxville*, *Longueuil*, *Magog*, *Mauricie*, *Missisquoi*, *Montréal*, *Rimouski*, *Saint-Hyacinthe*, *Sherbrooke*, *Valcourt*, *Yamachiche* ; *Abitibien*, *Adélois*, *Chicoutimien*, *Estrien*, *Forestois*, *Joliettain*, *Louperivois*, *Montérégien*, *Percéen*, *Saguenéen*, *Sorelois*, etc.

Pour ce qui est des raisons sociales et des marques de commerce, citons : *Alcan*, *Avon*, *Bell Canada*, *Campbell*, *Clairol*, *Domtar*, *Donahue*, *Sears*, *Hydro-Québec*, *Labatt*, *Kino-Québec*, *Molson*, *Québécor*, *Seagram*, *Sears*, *Weston*, *Gatorade*, *Skidoo*, *Jell-O*, *Jos-Louis*, etc.

Enfin, nous avons relevé environ 1700 sigles et acronymes différents, ce qui montre le foisonnement de ces formes en français contemporain et notamment en français québécois. Leur repérage s'est effectué à l'aide d'un dictionnaire des sigles que nous avions constitué à

partir de différentes sources québécoises déjà existantes. Quelque 1100 sigles ne figurent que dans la BDTS. En voici quelques exemples, correspondant à autant de domaines généraux (identifiés grossièrement) :

Tableau 1

Domaines	Sigles	Appellations
Santé	RAMQ	Régie de l'assurance maladie du Québec
Éducation	DEC	Diplôme d'études collégiales
Affaires municipales	MRC	Municipalité régionale de comté
Culture	BNQ	Bibliothèque nationale du Québec
Ministères, conseils et commissions	SAAQ	Société de l'assurance automobile du Québec
Langue	OLF	Office de la langue française
Partis politiques	BQ	Bloc québécois
Professions	FIIQ	Fédération des infirmières et infirmiers du Québec
Syndicats	CSD	Centrale des syndicats démocratiques

2.2 Deuxième type d'apport : les noms communs

Les noms communs (simples, dérivés ou composés) spécifiques à la BDTS, c'est-à-dire non attestés dans le dictionnaire-machine de WordPerfect, sont au nombre d'environ 35 000 formes.

— Des mots de langue orale et des termes techniques

Une partie de ces formes sont des graphies reflétant la langue orale. Celles-ci, en nombre limité, font l'objet de sous-listes distinctes et seront traitées ultérieurement ; néanmoins, chacune des formes relevées est rattachée au lemme approprié.

Quelques centaines de mots présents dans la BDTS présentent un caractère spécialisé, voire très spécialisé. Puisque cette banque contient de nombreux textes techniques et scientifiques, il n'est pas surprenant de rencontrer une grande quantité de termes relevant de domaines spécialisés, comme le démontrent les quelques exemples suivants : *fluorimètre*, *géobiologie*, *hydromassage*, *isomètre*, *isomorphie*, *nanogramme*, *macro-écologie*, *nanoseconde*, *neuro-moteur*, *nivo-glacique*, *télédélestage*, *climatogramme*, *argilite*, *génotoxique*, etc. Ces mots de langue spécialisée méritent d'être examinés en fonction de leur fréquence et de leur dispersion dans divers types de textes, notamment les textes didactiques, journalistiques (tirés de magazines spécialisés, par exemple) et autres documents usuels provenant d'entreprises et d'organismes de recherche. De la même manière que pour la langue orale, ils font l'objet de sous-listes distinctes et seront traités ultérieurement.

— **Des mots de langue générale**

Aux fins de la présente contribution, nous nous sommes davantage intéressés aux mots standard de langue plus générale que nous avons groupés temporairement selon diverses catégories : mots simples, mots composés ou dérivés.

Quelques exemples de mots spécifiques à la BDTS. — De manière à montrer l'apport de la BDTS comparativement aux différentes nomenclatures déjà existantes du français de France et du français du Québec, les formes non reconnues par le dictionnaire-machine de WordPerfect ont été vérifiées dans les nomenclatures des dictionnaires suivants : le *Robert électronique* (RÉ, 1996), le *Dictionnaire du français Plus* (DFP, 1988) et le *Dictionnaire québécois d'aujourd'hui* (DQA,1993). Nous illustrerons notre propos à l'aide d'une série de petits échantillons de mots, choisis parmi les dizaines de milliers de formes spécifiques relevées.

Le tableau 2 présente un premier groupe d'exemples illustrant, d'une part, des vocables absents du RÉ mais présents dans l'un ou l'autre des dictionnaires québécois, et, d'autre part, quelques mots nouveaux par rapport aux trois nomenclatures de référence. Ces mots sont tous répertoriés dans différents types de discours de type « standard » (journalistique, spécialisé ou didactique). Dans ces listes d'exemples et toutes celles qui suivront, le point noir (•) est symbole de présence, et le tiret (–), d'absence ; les chiffres donnés dans la dernière colonne correspondent à la fréquence observée dans la BDTS.

Tableau 2

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Assermentation	–	•	•	11	*Ruralité*	–	–	–	14
Caucus	–	•	•	43	*Impartition*	–	–	–	11
Dézonage	–	–	•	6	*Pulperie*	–	–	–	16
Aluminerie	–	•	•	78	*Tomaté*	–	–	–	5
Blanchon	–	•	•	12	*Infocentre*	–	–	–	3
Rouli-roulant	–	–	•	5	*Sociodémographique*	–	–	–	29
Mésadapté	–	–	•	5	*Subpolaire*	–	–	–	6
Perceptuel	–	•	–	126	*Surpêche*	–	–	–	14
Désencrage	–	–	•	7	*Téléthéâtre*	–	–	–	15
					Verbo-moteur	–	–	–	5

Quelques exemples illustrant le développement de certaines familles de mots. — Les échantillons du tableau 3 illustrent l'apport de la BDTS quant à la description de la dérivation touchant certaines familles lexicales. Ici, nous faisons d'abord état des mots déjà mentionnés dans le RÉ, le DFP et le DQA, puis de ceux qui ne sont attestés que dans la BDTS. Le tableau 4, quant à lui, est consacré à deux séries de dérivés apparentés aux mots *américain* et *français*.

Tableau 3

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Aviron	•	•	•	58	*Acériculture*	–	•	•	16
Avironner	–	•	•	11	*Acériculteur, trice*	–	•	•	12
Avironneur, euse	–	–	•	–	*Acéricole*	–	–	•	6
Avironnage	–	–	–	1	*Acérico-forestier*	–	–	–	3
Audiologie	•	•	–	–	*Dysfonctionnement*	•	•	•	14
Audiologiste	–	–	–	4	*Dysfonction*	•	–	•	8
					Dysfonctionnel	–	–	–	10

Tableau 4

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Américanisation	•	•	•	18	*Franciser*	•	•	•	122
Américaniser	•	•	•	12	*Francisation*	•	•	•	902
Américanisme	•	•	•	40	*Francisable*	•	–	–	1
Américaniste	•	•	–	1	*Francisme*	–	•	–	33
Américanophile	–	•	–	1	*Francité*	•	–	–	11
Américanité	–	–	–	13	*Francophile*	•	•	•	12
Américano-canadien	–	–	–	2	*Francophobe*	•	•	•	3
Américanolâtrie	–	–	–	1	*Francophobie*	•	•	–	5
Américanomanie	–	–	–	1	*Francophone*	•	•	•	4124
					Francophonie	•	•	•	323
					Francophonisation	–	–	•	35
					Francophoniser	–	–	•	4
					Francisant	–	–	–	10
					Franctitude	–	–	–	1
					Francofête	–	–	–	1
					Francofolie	–	–	–	12
					Antifrancophone	–	–	–	1
					Francophonisme	–	–	–	6

Les familles lexicales représentées comportent en outre un bon nombre de formes composées. À l'article *Américain*, le DQA donne en exemples de nombreux composés : *anglo-américain* (présent dans le RÉ), *anti-américain*, *centre-américain*, *franco-américain* (donné en exemple sous *franco-* dans le RÉ et présent dans le DFP), *hispano-américain* (présent dans le RÉ et le DFP), *italo-américain*, de même que *latino-américain*, *nord-américain* et *sud-américain* (tous trois présents dans le RÉ et le DFP). Pour sa part, la BDTS comporte toute une série de forme spécifiques composées à partir de l'élément franco- : *franco-africain*, *franco-albanais*, *franco-algérien*, *franco-allemand*, *franco-anglais*, *franco-catholique*, *franco-espagnol*, *franco-européen*, *franco-gréco-romain*, *franco-latin*, *franco-laurentien*, *franco-normand*, *franco-protestant*, *franco-wallon*, etc.

Nous pourrions multiplier à loisir les exemples de cette sorte. Comme on a pu le constater, la BDTS permet, dans plusieurs cas, de compléter des séries ou familles de mots déjà représentées dans les dictionnaires français et québécois de référence (RÉ 1996, DFP et DQA).

Quelques exemples de la productivité de certains éléments affixaux. — Parmi les diverses listes élaborées, nous avons classé les formes dérivées en fonction de l'affixe utilisé. Nous avons ainsi constitué deux sous-listes : l'une tenant compte des éléments préfixaux, l'autre des éléments suffixaux.

La formation de mots nouveaux à l'aide d'affixes est un phénomène ouvert, et il est impossible pour un lexicographe de les énumérer tous dans la liste des entrées d'un dictionnaire. Par exemple, le préfixe *re-* peut facilement s'accoler à un très grand nombre de verbes, et les recenser tous poserait un problème d'espace. Les lexicographes prennent généralement soin de noter ces éléments préfixaux, dont ils font une entrée indépendante et précisent les différents sens : *re-* « indique la répétition » (comme dans *redire*), et *anti-* signifie « qui s'oppose à » (comme dans *antiétatiste*), etc.

Par ailleurs, tous les dictionnaires enregistrent un nombre variable de ces formes dérivées. On peut croire que la sélection de ces dérivés est faite en fonction de divers critères, comme leur fréquence, leur degré de lexicalisation ou leur attestation dans certains textes, etc.

De manière, ici encore, à évaluer l'apport de la BDTS quant à l'enrichissement de nomenclatures existantes, nous avons comparé les formes dérivées commençant par les préfixes *anti-*, *agro-*, *kilo-* et *bi-* présentes tant dans la BDTS que dans les dictionnaires français et québécois de référence. Les échantillons qui suivent illustrent les formes communes aux dictionnaires dépouillés et celles spécifiques à la BDTS.

À titre simplement indicatif, notons que le préfixe *anti-* atteint 1398 occurrences et donne lieu à 385 unités lexicales différentes dans la BDTS ; 185 d'entre elles ont plus d'une occurrence. Le RÉ compte 214 entrées, dont 107 sont communes à celles de la BDTS. Le DFP[3] en compte 110, dont 63 sont communes à celles de la BDTS. Quant au DQA, il en présente 82, dont 67 sont communes à celles de la BDTS. Dès le départ, nous nous attendions à ce que les unités lexicales comportant le préfixe *anti-* soient plus nombreuses dans la BDTS que dans les dictionnaires consultés ; les auteurs de ces derniers effectuent une forte sélection parmi celles attestées dans les textes (un examen de la banque de textes FRANTEXT nous le confirmerait sans doute). Par ailleurs, on note un faible consensus dans la sélection des entrées retenues par les lexicographes, comme le montre le constat suivant : seulement 45 unités lexicales sont communes aux RÉ, DQA, DFP (toutes représentées dans la BDTS). Le tableau 5 donne quelques exemples de formes spécifiques à la BDTS ; ces mots sont tirés d'au moins deux discours de langue publique (journalistique, didactique, spécialisé ou autre).

Tableau 5
Anti-

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Anti-acide	–	–	–	2	*Antiquébec*	–	–	–	5
Antibuée	–	–	–	2	*Antisalissure*	–	–	–	10
Anticorrosif	–	–	–	6	*Antiscab*	–	–	–	3
Anti-émeute	–	–	–	16	*Antistress*	–	–	–	3
Antifumeur	–	–	–	3	*Antisyndical*	–	–	–	3
Antiparasitaire	–	–	–	28					

[3] Nous n'avons pas pris en compte les nombreuses formes simplement listées comme exemples sous l'entrée *Anti-*.

Les tableaux 6 à 9 offrent divers exemples de formes illustrant l'emploi des préfixes *agro-*, *auto-*, *bi-* et *kilo-*, tout en soulignant l'apport de la BDTS.

Tableau 6
Agro-

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Agroalimentaire[4]	•	•	•	148	*Agro-environnement*	–	–	–	2
Agro-alimentation	–	–	–	1	*Agro-environnemental*	–	–	–	6
Agrobiologie	•	–	–	10	*Agroforesterie*	–	–	–	3
Agrobiologique	–	–	–	2	*Agroforestier*	–	–	–	22
Agrochimie	•	–	–	–	*Agrotextile*	–	–	–	9
Agrochimique	–	–	–	1					
Agro-industrie	•	–	–	–	*Agrotourisme*	–	–	–	8
Agro-industriel	–	–	–	3	*Agrotouristique*	–	–	–	2

Tableau 7
Auto-

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Autopatrouille	–	•	•	4	*Autoparc*	–	–	–	1
Autocaravanier	–	–	–	2	*Autobloquant*	–	–	–	8
Autocueillette	–	–	–	5	*Autofocalisation*	–	–	–	1
Autodéterminer	–	–	–	9					

Tableau 8
Bi-

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Bisannuel	•	•	•	15	*Bidirectionnel*	–	•	–	3
Bi-annuel	–	–	–	2	*Bidirectionnalité*	–	–	–	3
Bi-athlon	•	•	–	19	*Bi-énergie*	–	–	–	28
Bi-athlonien	–	–	–	2	*Bi-ethnique*	–	–	–	2
Biculturalisme	•	•	•	25	*Binaire*	•	•	•	80
Biculturel	•	•	•	5	*Binarité*	–	–	–	1
Biculture	–	–	–	2	*Biparental*	–	–	–	5
Bidimensionnel	•	•	–	44	*Bipartisan*	–	–	–	3
Bidimension	–	–	–	1					
Bidimensionnalité	–	–	–	3					

[4] À cette étape-ci de notre recherche, nous ne tenons pas compte de la variation graphique relative à la soudure, à la présence ou à l'absence d'un trait d'union. Cette information sera prise en compte ultérieurement.

Tableau 9
Kilo-

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Kilogrammètre	–	•		–	*Kilo-ampère*	–	–	–	1
Kilohertz	•	•		4	*Kilobar*	–	–	–	2
Kilocalorie	•	•		11	*Kilohm*	–	–	–	4
Kilofranc	•	–		1	*Kilonewton*	–	–	–	9
Kilojoule	•	–		13	*Kilobit*	–	–		3
Kilométrage	•	•	•	34	*Kilodollar*	–	–		1
Kilométrique	•	•	•	5	*Kilo-électron*	–	–		2
Kilomètre	•	•	•	1492	*Kilolitre*	–	–		3
Kilométrer	–	•		–	*Kilovolt*	–	–		17
Kilotonne	•	•		1					
Kilowatt (heure)	•	•	•	65					
Kilopascal	–	–	•	21					
Kilocycle	–	•		1					
Kilo-octet	–	–	•	12					

De la même façon, la BDTS vient compléter des séries de mots pour lesquelles le français n'avait pas encore enregistré les formes lexicales dérivées. On note également l'intérêt de certaines d'entre elles qui présentent une fréquence relativement élevée et qui sont attestées dans plusieurs types de discours. Le tableau 10 donne quelques exemples de formes construites à partir des suffixes *-isme* et *-ment* :

Tableau 10
-isme* et *-ment

Vocable	RÉ	DFP	DQA	BDTS	Vocable	RÉ	DFP	DQA	BDTS
Amérindianisme	–	•	•	9	*Dépendamment*	–	–	•	18
Duplessisme	–	–	•	14	*Supposément*	–	–	•	43
Francisme	–	•	–	33	*Démographiquement*	–	–	–	5
Créditisme	–	–	–	2	*Émotivement*	–	–	–	13
Étapisme	–	–	–	6	*Minoritairement*	–	–	–	1
Normativisme	–	–	–	1	*Nationalement*	–	–	–	2
Rangisme	–	–	–	2	*Paritairement*	–	–	–	1
					Presquement	–	–	–	38
					Publicitairement	–	–	–	1
					Syndicalement	–	–	–	3

2.3. Troisième type d'apport de la BDTS : les noms communs à caractère « normatif »

La BDTS fournit des informations nouvelles pour des séries de mots liés à l'aspect normatif du français québécois, notamment les emprunts critiqués à l'anglais, les emprunts aux autres langues, les mots faisant l'objet d'un avis de recommandation ou de normalisation de la

Commission de terminologie de l'Office de la langue française (CTOLF) et la féminisation des titres de fonction.

— Les emprunts critiqués à l'anglais

La BDTS se révèle une source intéressante pour renouveler et justifier la nomenclature des emprunts critiqués à l'anglais devant figurer dans les dictionnaires usuels et divers répertoires d'anglicismes ou autres difficultés du français québécois. Nous avons déjà traité cet aspect dans divers articles scientifiques (v. notamment Cajolet-Laganière, Martel et Théoret, 2000 et à paraître).

Dans le cadre de cette contribution, nous illustrerons l'apport de la BDTS à partir de données récentes recueillies sur les emprunts de forme anglaise. Ceux-ci totalisent 4876 formes sur 10 733, soit presque 50 % des anglicismes colligés dans le *Dictionnaire électronique d'anglicismes* mis au point par Michel Théoret et nous-mêmes[5].

Une fois que la BDTS a été balayée avec la liste des 4876 formes anglaises répertoriées dans ce dictionnaire électronique, on constate que 2029 formes sont carrément absentes du corpus, soit 50 % de la liste de départ. Après analyse de tous les contextes présents, il appert que la très grande majorité des formes attestées affichent des fréquences basses, voire très basses. De plus, la plupart des formes répertoriées peuvent être associées à certains discours spécifiques. Ces données nous permettent donc d'évaluer la présence des emprunts dans les textes et leur dispersion dans divers types de discours ; elles nous fournissent également les emprunts actualisés dans des contextes diversifiés. Les tableaux 11 à 13 reprennent quelques exemples, dont certains ne sont répertoriés que dans les discours oral et littéraire (littérature populaire, dialogues) ; d'autres, dans les discours oral, littéraire et journalistique ; d'autres enfin, dans les discours spécialisés. Notons également plusieurs groupes complexes relevés en contextes.

Nous avons par ailleurs relevé, pour les lettres A et B, quelque 234 emprunts à l'anglais non répertoriés dans le *Dictionnaire électronique des anglicismes*. C'est pour nous un excellent moyen de compléter et d'actualiser ce dictionnaire, et d'enrichir la nomenclature relative aux emprunts.

---

[5] Rappelons brièvement que la nomenclature de ce dictionnaire comprend les emprunts critiqués recensés dans *Le Nouveau Petit Robert* (1993), dans les trois premières éditions du *Dictionnaire des anglicismes* de Colpron (1970, 1982 et 1994), dans le *Multidictionnaire des difficultés de la langue* de Marie-Éva de Villers (1992), dans diverses chroniques linguistiques de l'Office de la langue française et dans le *Dictionnaire des canadianismes* de Gaston Dulong (1989).

Tableau 11
Exemples d'emprunts de forme anglaise attestés uniquement dans les discours oral et littéraire (littérature populaire, dialogues)

Emprunt critiqué	Type de texte et sous-fréquence					Expression
	oral	litt.	journ.	didact.	spéc.	
Batch	3	–	–	–	–	
Boileur	3	–	–	–	–	
Bumpeur	4	–	–	–	–	
Crinque	4	–	–	–	–	
Drum	7	–	–	–	–	
Jumper	27	–	–	–	–	*jolly jumper*
Crinquer (verbe)	7	7	–	–	–	
Flyer (verbe)	7	19	–	–	–	

Tableau 12
Exemples d'emprunts de forme anglaise attestés dans les discours oral, littéraire et journalistique

Emprunt critiqué	Type de texte et sous-fréquence					Expression
	oral	litt.	journ.	didact.	spéc.	
Gyproc	3	–	3	–	–	
Body	1	8	2	–	–	*body check* ; *body piercing*
Break	17	17	4	–	–	*tie-break*
Cash	6	11	4	–	–	*cash-flow* ; *passer au cash* ; *prendre pour du cash*
Lousse	5	5	3	–	–	

Tableau 13
Exemples d'emprunts de forme anglaise attestés également dans les discours spécialisés

Emprunt critiqué	Type de texte et sous-fréquence					Expression
	oral	litt.	journ.	didact.	spéc.	
Bag	7	4	1	–	1	*punching bag* ; *sleeping bag*
Flash	4	3	6	2	–	*flash-back*
Hobby	5	3	1	1	3	
Kit	12	4	10	1	5	

Outre les emprunts à l'anglais, nous avons relevé une cinquantaine de mots d'origine latine ou grecque. La majorité d'entre eux sont des noms génériques et spécifiques faisant référence à des plantes, des animaux et d'autres espèces naturelles ; on trouve aussi quelques expressions usuelles. Enfin, nous avons relevé quelque 40 emprunts à l'espagnol, à l'italien et à l'allemand. Aux termes de l'analyse de la totalité des listes recensant les diverses formes d'emprunts actualisés dans la BDTS, nous serons mieux en mesure d'évaluer la pertinence

d'intégrer ces emprunts à la nomenclature du dictionnaire dont nous planifions la rédaction[6] et de procéder, selon le cas, à leur description.

— Les avis de normalisation et de recommandation de la CTOLF

La quatrième édition du *Répertoire des avis terminologiques et linguistiques* de la CTOLF compte plus de 1630 avis de normalisation ou de recommandation. Il importe de noter que, dans tous ces cas, il s'agit de l'emploi d'un mot ou d'un sens spécifique à l'usage québécois. Dans la presque totalité des cas, ces mots ou ces sens ne sont pas encore explicités dans les dictionnaires français ou québécois.

Ici encore, la BDTS offre un moyen intéressant de préciser l'usage, dans divers types de discours, des mots et sens ayant fait l'objet d'un avis. Elle nous permet d'isoler ceux qui sont d'usage courant en langue générale et ceux qui sont peu ou non présents dans les différents types de discours.

Nos données touchent pour l'instant environ 10 % des avis du dernier répertoire, soit tous les termes de la première tranche alphabétique (lettre A) (v. Labrecque, 2001). Ce premier travail d'analyse nous permet de constater qu'un peu moins de la moitié des termes recommandés ou normalisés sont absents de la BDTS, soit 70 termes sur 157, ou 45 % des avis analysés. Parmi les 87 termes attestés, soit 55 % du corpus, nous avons tiré les quelques exemples qui suivent, dont nous précisons la fréquence et la dispersion selon le type de discours (v. tableau 14).

Tableau 14

Objet de l'avis		Spéc.	Journ.	Littér.	Didact.	Oral	Total
Accotement	n. m.	57	1	14		3	75
Aiguillat	n. m.	14	4				18
Alose	n. f.	21	1	2		3	27
Aluminerie	n. f.	52	23	1		2	78
Analphabète	n.	17	12	3	6		38
Andragogie	n. f.	3			3		6
Anguille	n. f.	37	51	28		6	122
Aréna	n. m.	3	54	8	1	14	80
Arrêt	n. m.	100	6	9			115

— La féminisation des titres de fonction

Les avis de la CTOLF concernent également la féminisation du discours et des titres de fonction. Il existe peu d'études quantitatives sur la féminisation des titres de fonction dans les textes. Dans le même but, à savoir d'évaluer la fréquence et la dispersion des titres de fonction féminisés préconisés par l'Office de la langue française, nous avons balayé la BDTS

[6] Le projet planifié d'un dictionnaire du français standard en usage au Québec a effectivement démarré en septembre 2001.

avec la liste des quelque 950 formes féminisées diffusées dans *Le français au bureau* (Guilloton et Cajolet-Laganière, 2000).

À ce sujet, la principale particularité québécoise est la suivante : la coexistence des formes féminines régulières en *-euse*, *-teuse* et *–trice*, et celles moins régulières en *-eure* et *-teure*. Voici ce que nous révèle la BDTS. Des 330 formes féminisées présentes dans la BDTS, on en trouve 91 se terminant en *-eur*, dont 69 sont féminisées en *-euse* et 22 en *-eure*. Certaines parmi ces dernières sont préconisées par l'OLF : par exemple, *auteure*, *professeure*, *assureure*, *metteure en scène*, *réviseure*, *sculpteure*, *superviseure*, *ingénieure*, etc. D'autres, spécifiques à la BDTS, représentent des créations libres des rédacteurs et rédactrices : *chercheure*, *compositeure*, *entraîneure*, *entrepreneure*, *acquéreure*, *agresseure*, *amateure*, *arpenteure*, *arrangeure*, *défenseure*, *inventeure*, *professeure-chercheure*, *recenseure* et *relieure*.

En plus des données quantitatives, la BDTS fournit la dispersion des formes selon les types de discours ainsi que les contextes d'emploi des formes relevées. Cela permet sans aucun doute une description et une hiérarchisation meilleures des usages, notamment en ce qui concerne les formes en concurrence.

3. Conclusion

L'examen préliminaire des diverses composantes de la BDTS démontre d'ores et déjà plusieurs apports importants de cette banque au regard des nomenclatures des trois dictionnaires français et québécois sélectionnés aux fins de cette étude. Tout d'abord, mentionnons l'attestation, la fréquence et la dispersion des unités lexicales (noms propres et noms communs) figurant dans les nomenclatures actuelles des dictionnaires papier et électronique, français et québécois. Les différentes informations à caractère lexical contenues dans la BDTS permettent aux lexicographes d'effectuer des choix basés sur des données réelles et objectives. De fait, les exemples puisés dans les nomenclatures des trois dictionnaires de référence dénotent une grande disparité entre les choix effectués par les lexicographes. De surcroît, les quelques exemples présentés ci-dessus montrent que de nombreux vocables de la BDTS sont carrément absents des nomenclatures des dictionnaires français et québécois consultés. Leur intégration dans les nomenclatures d'éventuels dictionnaires du français québécois pourra se faire en fonction de certains critères linguistiques et sociolinguistiques au regard de la politique éditoriale, des contraintes d'espace, etc.

Notons aussi que, compte tenu du nombre, de la variété et de la représentativité des textes composant la BDTS, nous pouvons tirer de celle-ci des informations de nature à faciliter la hiérarchisation des usages, notamment du français standard en usage au Québec (ou bon usage au Québec) et des différentes formes en concurrence (critiquées, recommandées).

De plus, pour ajouter à ces informations essentielles à toute description lexicographique, nous pouvons extraire de la BDTS une grande quantité d'exemples et de contextes d'emploi permettant une description plus complète des unités lexicales, incluant les collocations (comme *société distincte*), les cooccurrents (comme *chanteur francophone*) et les phraséologismes (comme *donner l'heure juste*). Mentionnons également que la BDTS se

prête à l'explicitation des références culturelles et encyclopédiques spécifiques au Québec, grâce notamment aux exemples, signés ou non, et aux citations (littéraires, journalistiques, didactiques et autres).

La BDTS permet finalement aux concepteurs d'outils langagiers informatiques d'y puiser tous les éléments nécessaires à l'enrichissement et à la prise en compte de la variété linguistique et culturelle québécoise dans les dictionnaires usuels papier et électroniques, les dictionnaires-machines voués à la correction orthographique des textes, les logiciels de dictée vocale et de reconnaissance de la parole, notamment ceux en usage au Québec et en Amérique du Nord, les moteurs de recherche et autres logiciels, ainsi que les vocabulaires spécialisés et la veille terminologique.

En dernier lieu, nous sommes conscients des limites de la présente recherche, car il s'agit ici d'une étude fragmentaire. Il importe de terminer l'analyse en cours, de raffiner les outils de recherche, de procéder à certaines vérifications et de compléter nos données dans des corpus plus grands.

Références

Cajolet-Laganière, Hélène, et Noëlle Guilloton (2000). *Le français au bureau*, 5e édition, revue, augmentée et mise à jour, Québec, Les Publications du Québec, 503 p.

Cajolet-Laganière, Hélène, et Pierre Martel (1996). *Le français québécois : usages, standard et aménagement*, Québec, Institut québécois de recherche sur la culture (« Diagnostic, 22 »), 141 p.

Cajolet-Laganière, Hélène, et Pierre Martel (éd.) (1994). *Actes du Colloque sur les anglicismes et leur traitement lexicographique* (septembre 1991) suivi de *Bibliographie sur les emprunts à l'anglais et les anglicismes en français 1945-1993*, Québec, Office de la langue française, 382 p.

Cajolet-Laganière, Hélène, Pierre Martel et Marie-France Langlois (2001). « La pertinence culturelle et linguistique dans les outils linguistiques informatisés couramment utilisés au Québec », *Revue belge de philologie et d'histoire*, n° 79, p. 861-881.

Cajolet-Laganière, Hélène, Pierre Martel et Michel Théoret (2000). « Des emprunts à l'anglais des deux côtés de l'Atlantique », dans Marie-Rose Simoni-Aurembou (éd.), *Actes du cinquième Colloque international de Bellême du 5 au 7 juin 1997*, Tübingen, Niemeyer (« Canadiana Romanica, 13 »), p. 207-215.

— (à paraître). « Les emprunts critiqués à l'anglais : un fait historique à évaluer », communication présentée au colloque *La lexicographie canadienne à l'aube de l'an 2000* (congrès de l'ACFAS, Université d'Ottawa, mai 1999).

Colpron (1970 ; 1982 ; 1994) : Colpron, Gilles. *Les anglicismes au Québec. Répertoire classifié*, Montréal, Libraire Beauchemin limitée, 1970. – *Dictionnaire des anglicismes*, Beauchemin, 1982. – *Le Colpron. Le nouveau dictionnaire des anglicismes*, mise à jour de Constance et Louis Forest, Laval, Éditions Beauchemin ltée, 1994.

De Villers, Marie-Éva (1992). *Multidictionnaire des difficultés de la langue française*, 2e édition, Montréal, Éditions Québec/Amérique.

Dulong, Gaston (1989). *Dictionnaire des canadianismes*, Boucherville (Québec), Larousse Canada.

FRANTEXT : Base de données textuelles FRANTEXT, Nancy, Institut national de la langue française (Centre national de la recherche scientifique).

Labrecque, Geneviève (2001). *La présence des avis de recommandation et de normalisation de la Commission de terminologie de l'Office de la langue française dans un corpus de langue générale et spécialisée*, mémoire de maîtrise, Sherbrooke, Université de Sherbrooke, 246 p. et annexes.

Langlois, Marie-France (1999). *Une nouvelle grille de marques de registres de langue : perception et usage des rédacteurs et communicateurs québécois*, mémoire de maîtrise, Sherbrooke, Université de Sherbrooke, 257 p.

Martel, Pierre (2001). « Le français de référence et l'aménagement linguistique », *Cahiers de l'Institut de linguistique de Louvain*, vol. 27, nos 1-2 (*Actes du colloque international Le français de référence : constructions et appropriations d'un concept*), Louvain-la-Neuve, p. 123-139.

Nouveau Petit Robert (Le), sous la direction de Josette Rey-Debove et Alain Rey, Paris, Dictionnaires Le Robert, 1993.

Poirier, Claude (dir.) (1998). *Dictionnaire historique du français québécois. Monographies lexicographiques de québécismes*, Sainte-Foy (Québec), Les Presses de l'Université Laval, xlviii-641 p.

Répertoire des avis terminologiques et linguistiques : mai 1979 – novembre 1996, 4e édition revue et augmentée, Gouvernement du Québec, Office de la langue française, 1998, 361 p.

Site Web du Secrétariat à la politique linguistique (Gouvernement du Québec), Réseau des corpus lexicaux québécois (http://www.spl.gouv.qc.ca/banque/banque.html).