QUELQUES ÉLÉMENTS VISANT À RENOUVELER LA PRATIQUE LEXICOGRAPHIQUE AU QUÉBEC

Pierre Martel
Professeur
Université de Sherbrooke

Hélène Cajolet-Laganière
Professeure
Université de Sherbrooke

La présente communication s'inscrit dans le cadre plus large des travaux menés au Centre d'analyse et de traitement informatique du français québécois (CATIFQ) par un groupe de chercheuses et de chercheurs de l'Université de Sherbrooke. Ces travaux visent notamment à la description du français québécois dans une perspective synchronique, prenant en considération la hiérarchisation des usages internes.

Nous nous intéressons principalement aux unités lexicales appartenant au niveau standard du français québécois. Le mot *standard*, employé dans le contexte du présent article, désigne une catégorie de textes dont le niveau de langue est généralement soigné; il s'oppose principalement à des textes oraux de niveau familier, vulgaire, etc. L'objet de cette communication est de présenter les principales caractéristiques de notre pratique lexicographique. Nous avons groupé ces caractéristiques autour de cinq grands axes soit : répondre aux besoins exprimés par la population; élaborer une nomenclature et une description originales du français standard en usage au Québec; établir une nouvelle grille de marques d'usage adaptée au contexte sociolinguistique du Québec; développer une méthode de travail informatisée pour toutes les étapes d'élaboration du dictionnaire du français au Québec et finalement assurer un transfert technologique.

1. Répondre aux besoins exprimés par la population

Nous avons mené, au cours des dernières années, plusieurs enquêtes dans le but de connaître les attentes du public québécois en matière de dictionnaire usuel. Nous ne mentionnerons que la plus importante, soit une enquête sous forme de questionnaire écrit, menée en 1997 dans six régions du Québec (Montréal, Estrie, Saguenay—Lac-Saint-Jean, Outaouais, Québec et Mauricie) auprès de quelque huit cents personnes

de différents âges et fonctions. À la suite de l'analyse des résultats de cette enquête, il ressort que les Québécois et Québécoises désirent essentiellement un dictionnaire de type normatif. Ils veulent être informés sur le « bon usage » ou l'usage « dominant » au Québec, de telle sorte que leur expression soit juste et correcte selon le type de discours. Ils veulent savoir, d'une part, ce qui est « reçu » ou « accepté » selon un certain « standard québécois », et d'autre part, les emplois « critiqués » de même que les emplois qui peuvent varier en fonction de certaines situations de communication. Ils souhaitent enfin un dictionnaire général à l'intérieur duquel les mots et les sens employés au Québec seront distingués des mots et des sens employés dans les autres pays francophones à l'aide d'une marque quelconque. Un tel dictionnaire servirait de « guide » en matière de « bon usage » ou d'usage « dominant » au Québec. L'analyse des informations recueillies nous a en outre fourni les attentes des répondants et répondantes concernant chaque élément de la macro et de la microstucture du dictionnaire et les divers aspects typographiques pertinents.

En revanche, les répondants s'opposent à l'imposition d'une norme qui aurait l'apparence d'une norme officielle. Ils ne veulent pas d'un dictionnaire « étatique » ni « national », comme ceux de l'*Académie française* ou du *Trésor de la langue française*, qui aurait une quelconque autorité contraignante. À ce sujet, le projet développé à Sherbrooke est le fruit d'une équipe de chercheurs et de chercheuses interuniversitaire qui veulent proposer différents produits lexicographiques qui seront soumis à l'appréciation du public québécois, comme c'est le cas pour les autres productions dictionnairiques québécoises (passées ou actuelles).

2. Élaborer une nomenclature et une description originales du français standard en usage au Québec en exploitant notamment la Banque de données textuelles de Sherbrooke (BDTS)

Cette deuxième caractéristique constitue sans aucun doute la partie la plus importante de la communication et des travaux du groupe de recherche. Elle concerne l'élaboration et l'exploitation de la Banque de données textuelles de Sherbrooke (BDTS), l'extraction de cette banque de diverses nomenclatures, soit la nomenclature du français québécois pour ce qui est de la langue générale (noms communs et noms propres); la nomenclature des noms communs à caractère normatif (emprunts critiqués à l'anglais, avis de l'OLF, etc.); la nomenclature du français québécois pour ce qui est du vocabulaire de base de divers domaines spécialisés et enfin, la validation de ces nomenclatures grâce à la comparaison des nomenclatures extraites de la BDTS avec celles des principaux dictionnaires usuels français et québécois.

2.1 L'originalité et l'intérêt de la BDTS

Sa structure

La BDTS est constituée d'une série de sous-ensembles qui donnent accès aux différents registres de la langue, notamment au registre standard, actuellement en usage au Québec, et ce, tant en langue générale qu'en langue plus spécialisée. Elle permet en outre des études selon les situations de communication, les domaines d'activités, etc. Par sa taille relativement modeste, la BDTS permet une exploitation plus efficace des données lexicales du français québécois.

Son contenu et la représentativité de ses textes

La BDTS est un corpus informatisé et indexé de plus de quelque vingt-six millions de mots (juillet 2001); 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 : 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) de même que la langue littéraire, journalistique et relativement plus spécialisée utilisée au Québec dans différentes situations de communication. 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.

La BDTS est composée de cinq grandes catégories de textes réparties 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.), afin d'avoir un échantillon de textes oraux de manière à représenter ce type de discours et à hiérarchiser les usages à l'aide d'un système approprié de marques;
- Vingt pour cent d'ouvrages didactiques (20 %), comprenant des échantillons de textes tirés de manuels scolaires (de niveau 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);
- 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.). Cet échantillon 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;
- Trente pour cent de textes spécialisés (30 %) enfin, comprenant des échantillons de textes techniques, scientifiques, sociopolitiques, administratifs, juridiques, médicaux, environnementaux, et autres. Ces écrits sont nombreux et très diversifiés; la plupart de ces textes reflètent une langue relativement soignée et valorisée; ils fournissent des mots de langue générale et le vocabulaire de base de très nombreux domaines spécialisés.

Son caractère évolutif propre au modèle adopté

Conformément au modèle théorique adopté, la BDTS est une banque de taille réduite, mais à caractère évolutif, c'est-à-dire qu'elle augmente d'une manière progressive selon les besoins de représentativité des textes et compte tenu de certains tests de validation. Quatre corpus sont en cours d'élaboration, représentatifs

- du domaine très vaste des relations de travail (conventions collectives, textes de lois, manuels, procédures, etc.) (plus de 3 millions de mots);
- des domaines de la faune, de la flore et de l'environnement naturel du Québec (entre 3 et 5 millions de mots);
- des domaines linguistique, culturel, social, politique, juridique, etc. (entre 3 et 5 millions de mots);
- du domaine de la santé et de l'éducation (environ 3 millions de mots).

Nous prévoyons en outre garder à jour la Banque grâce à l'entrée de textes reflétant l'actualité socioculturelle, économique, politique et autres.

Ses corpus périphériques

À cette banque s'ajoutent enfin les cédéroms *Actualités/Québec* et *Repère*, *Internet*, etc., ainsi que le *Réseau des corpus lexicaux québécois* (un corpus réduit de la BDTS en fait partie), géré par le Secrétariat à la politique linguistique du Québec; c'est la partie disponible en périphérie de la BDTS.

2.2 La nomenclature du français québécois pour ce qui est de la langue générale

Après indexation des premiers 20 millions de mots stockés dans la BDTS (jusqu'en septembre 2000), on avait relevé 170 000 formes présentes. Toutes ces formes ont été comparées à celles du dictionnaire-machine de *WordPerfect*. Comme suite à cette comparaison, il est apparu que 100 000 formes sont communes aux deux nomenclatures et 70 000 formes, spécifiques à la BDTS. Les formes spécifiques à la BDTS se répartissent en nombre à peu près égal entre les noms communs et les noms propres.

2.2.1 Les noms propres

Il y a environ 35 000 formes absentes du dictionnaire-machine de *WordPerfect* (voir en bibliographie Cajolet-Laganière, Martel et Langlois, 2001). Ce sont surtout des noms propres de personne : *Ferland*, *Fortier*, *Gauvreau*, *Labrecque*, etc.; des toponymes et gentilés : *Boucherville*, *Drummondville*, *Estrie*, *Granby*, etc., *Abitibien*, *Adélois*, *Chicoutimien*, *Forestois*, etc.; des raisons sociales et marques de commerce : *Alcan*, *Avon*, *Bell Canada*, etc.; des sigles, acronymes et autres dénominations (plus de 1700 sont recensés dans la BDTS). Après analyse, nous constatons que 1100 sigles sont spécifiques à la BDTS. En voici quelques exemples :

Santé	AHQ	Association des hôpitaux du Québec
	CHUM	Centre hospitalier de l'Université de Montréal
Éducation	CRÉPUQ	Conférence des recteurs et des principaux des universités du Québec
Culture	BNQ	Bibliothèque nationale du Québec
Ministères, conseils et commissions	CRD	Conseil régional de développement
	RAAQ	Régie de l'assurance automobile du Québec
Langue	GDT	Grand dictionnaire terminologique
	OLF	Office de la langue française
Partis politiques	BQ	Bloc québécois
	PCM	Parti civique de Montréal

2.2.2 Les noms communs

Quelque 35 000 formes ne sont pas reconnues par *WordPerfect* (voir en bibliographie Cajolet-Lagnière et Martel, 2001). Les mots de la langue générale sont groupés, aux seules fins de la présente présentation, en deux catégories : les mots simples et les mots composés ou dérivés à partir d'éléments préfixaux et suffixaux. Nous avons vérifié leur présence dans trois dictionnaires français et québécois : le RÉ, le DFP et le DQA.

Voici quelques exemples tirés des deux catégories mentionnées :

Vocable	RÉ	DFP	DQA	BDTS/fréquence
Ruralité	-	-	-	14
Impartition	-	-	-	11
Pulperie	-	-	-	16
Tomaté	-	-	-	5
Infocentre	-	-	-	3
Sociodémographique	-	-	-	29
Subpolaire	-	-	-	6
Surpêche	-	-	-	14
Téléthéâtre	-	-	-	15
Verbo-moteur	-	-	-	5
Amérindianisme	-	X	X	9
Duplessisme	-	-	X	14
Francisme	-	X	-	33
Créditisme	-	-	-	2
Étapisme	-	-	-	6
Normativisme	-	-	-	1
Rangisme	-	-	-	2
Dépendamment	-	-	X	18
Supposément	-	-	X	43
Démographiquement	-	-	-	5
Émotivement	-	-	-	13
Minoritairement	-	-	-	1
Nationalement	-	-	-	2
Paritairement	-	-	-	1
Publicitairement	-	-	-	1
Syndicalement	-	-	-	3

2.2.3 Les noms communs à caractère « normatif »

La BDTS est une source intéressante pour renouveler et justifier la nomenclature de ces tranches de vocabulaire souvent problématiques au regard de l'usage québécois. Nous avons divisé cette catégorie de noms communs en trois groupes :

– les emprunts critiqués à l'anglais,

– la féminisation des titres de fonction,

– les avis de normalisation et de recommandation de l'OLF (voir le texte de Geneviève Labrecque dans ce même ouvrage).

Nous avons élaboré un dictionnaire électronique des anglicismes (DEA) à la suite du dépouillement des principaux relevés et répertoires québécois colligeant ces formes critiquées. Il s'agit de chroniques et de répertoires publiés par l'Office de la langue française, des deux premières éditions des *Anglicismes au Québec* de Gilles Colpron, du *Multidictionnaire* de Marie-Éva de Villers, du *Dictionnaire des canadianismes* de Gaston Dulong, et enfin, de la liste des anglicismes répertoriés dans le corpus oral de Sherbrooke, à partir du dictionnaire usuel *Le Petit Robert* (Beauchemin, Martel et Théoret, 1992). Le DEA contient 4227 vocables ainsi que toutes les formes (10 682) sous lesquelles ces vocables peuvent être rencontrés dans les textes. Aux fins de cette analyse, nous avons regroupé les emprunts critiqués en deux grandes catégories : les emprunts de forme (simples et complexes) et les emprunts de sens (simples et complexes); voir en bibliographie Cajolet-Laganière, Martel et Théoret, 1998 et 2000. Il importe de noter qu'à cette première étape du travail, qui consistait à rassembler dans un même ensemble tous les emprunts critiqués par les sources sélectionnées, nous n'avons pas remis en cause l'étiquette *anglicisme*, *emprunt*, etc., accolée par l'un ou l'autre des auteurs dépouillés, même si certaines recherches montrent qu'il y aurait lieu de modifier ou de nuancer plusieurs d'entre elles. Ces vérifications et cette hiérarchisation des usages seront faites ultérieurement, compte tenu de la présence et de la fréquence des emprunts relevés dans les textes.

En balayant la BDTS à l'aide du DEA, nous constatons que quelque 50 % des formes anglaises (simples et complexes) sont carrément absentes et la très grande majorité des formes présentes ont des fréquences très basses. Nos études nous ont fourni la liste des emprunts de forme, en contexte, par fréquence décroissante et par type de discours. Nous relevons par exemple que certains emprunts critiqués de forme anglaise ne sont

répertoriés que dans les discours oral et littéraire (littérature populaire, dialogues) :

Emprunt de forme critiqué	Type de texte et sous-fréquence	Expression
Batch	Oral : 3	
Boileur	Oral : 3	
Bumpeur	Oral : 4	
Jumper	Oral : 27	*jolly jumper*
Crinquer (verbe)	Oral : 7, Litt. : 7	
Flyer (verbe)	Oral : 7, Litt. : 19	

D'autres emprunts se trouvent dans les textes oraux, les textes littéraires et aussi dans les textes journalistiques :

Emprunt de forme critiqué	Type de texte et sous-fréquence	Expression
Gyproc	Oral : 3, Journ. : 3	
Break	Oral : 17, Journ. : 4, Litt. : 17	*tie-break*
Cash	Oral : 6, Journ. : 4, Litt. : 11	*cash-flow, passer au cash; prendre pour du cash*
Lousse	Oral : 5, Journ. : 3, Litt. : 5	

Au contraire des précédents, certains emprunts de forme anglaise se trouvent dans tous les types de textes, y compris dans les textes spécialisés :

Emprunt de forme critiqué	Type de texte et sous-fréquence	Expression
Bag	Oral : 7, Journ. : 1, Litt. : 4, Spéc. : 1	*punching bag, sleeping bag*
Flash	Oral : 4, Journ. : 6, Litt. : 3, Didact. : 2	*flash-back*
Hobby	Oral : 5, Journ. : 1, Litt. : 3, Didact. : 1, Spéc. : 3	
Kit	Oral : 12, Journ. : 10, Litt. : 4, Didact. : 1, Spéc. : 5	

Il en va différemment dans le cas des emprunts de sens; voir Escayola, 2000. La fréquence des emprunts de sens dans la BDTS est nettement supérieure à celle des emprunts de forme. Il s'agit là d'une difficulté réelle et importante pour les rédacteurs et rédactrices, qui, le plus souvent, étant donné leur forme française, les emploient d'une manière inconsciente. Le DEA recense 758 emprunts de sens fournis par les auteurs des ouvrages dépouillés. L'étude de ces vocables est très longue, puisqu'elle exige d'analyser chacun des contextes de manière à distinguer les emplois français ou « acceptés » de ceux qui sont « critiqués » par l'un ou l'autre des ouvrages lexicographiques sélectionnés. Nos études nous ont fourni la liste des emprunts de sens, en contexte, par fréquence et par dispersion selon les types de discours. Ces emprunts de sens représentent pour nous un élément très important de la nomenclature des noms communs à caractère normatif.

La féminisation des titres de fonction

La BDTS nous fournit l'attestation des formes féminines préconisées par l'OLF et des formes en concurrence : 330 formes féminisées sont attestées sur 950 formes préconisées. À titre d'exemple, mentionnons la coexistence des formes féminines en *-euse*, *-teuse*, *-trice*, et celles moins régulières en *-eure* et *teure*. Des 330 formes féminisées dans la BDTS, on trouve 91 formes se terminant en *-eur*, dont 69 sont féminisées en *-euse* et 22 en *-eure*. Certaines sont préconisées par l'OLF : *auteure, professeure, assureure, metteure en scène, réviseure, sculpteure, superviseure, ingénieure*, etc. D'autres sont spécifiques à la BDTS et représentent des créations libres des rédacteurs et rédactrices : *chercheure, compositeure, entraîneure, entrepreneure, acquéreure, agresseure, amateure, arpenteure, arrangeure, défenseure, inventeure, professeure-chercheure, recenseure et relieure.*

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

Les avis de normalisation et de recommandation de l'OLF

L'Office de la langue française a émis un peu plus de 1630 avis de recommandation ou de normalisation. Dans ces avis, on trouve non seulement des termes « recommandés » ou « normalisés », mais également des mots en concurrence, comme des synonymes, des québécismes, des régionalismes, des mots utilisés essentiellement en France, des emplois critiqués, etc.

Il importe pour nous de vérifier la présence et la dispersion de tous ces termes selon les registres de langue et les types de discours de la BDTS. Une de nos assistantes de recherche a effectué une étude en ce sens (voir Labrecque, 2001b).

Il ressort notamment de l'ensemble de l'étude que les avis devraient être diffusés davantage et qu'il y aurait lieu de favoriser leur utilisation en langue générale et spécialisée. Leur apparition dans un dictionnaire usuel contribuerait à un plus grand usage des termes officialisés dans différentes situations de communication; il serait souhaitable que les termes officialisés et leurs concurrents, recommandés et critiqués, soient hiérarchisés en fonction des différents registres de langue.

2.3 L'enrichissement de la nomenclature du français québécois pour ce qui est des langues de spécialité

Ce sont le plus souvent les termes techniques passés dans l'usage qui viennent enrichir la nomenclature des dictionnaires de langue. Cela concerne l'ajout de mots, de groupes complexes, de collocations, de cooccurrents, de contextes d'emploi illustrant les sens, les référents culturels, etc. La BDTS est constituée de sous-corpus représentatifs de plusieurs domaines de spécialité et des ajouts constants continuent de l'enrichir selon les besoins ciblés.

Un exemple : celui du domaine de l'environnement

Les langues de spécialité sont constituées essentiellement de vocabulaires spécifiques et leur description relève d'études terminologiques. Mais bon nombre de mots et de syntagmes, en raison de la place qu'occupent ces domaines dans la vie quotidienne des locuteurs d'une langue, pénètrent dans l'usage général de la langue et méritent d'être décrits dans les dictionnaires usuels. D'où l'importance et l'intérêt de repérer les mots les plus fréquents et les plus importants dans un domaine donné, ceux notamment présents dans les journaux et autres textes vulgarisés à l'intention du grand public.

Nous avons constitué un corpus représentatif de 250 000 mots dans le domaine de l'environnement (voir en bibliographie Tremblay, 2000); il est intégré dans la BDTS. À titre indicatif et aux fins de la présente communication, nous avons examiné quelques termes et vérifié leur description dans les dictionnaires usuels français et québécois. L'information nouvelle apportée par les contextes du corpus québécois touche les synonymes, les groupes complexes lexicalisés ou en voie de l'être, les collocations, les cooccurrents et phraséologismes, les citations et contextes illustrant les

sens et les référents culturels, diverses informations permettant d'enrichir les définitions et enfin des notes encyclopédiques. Voici quelques exemples d'unités complexes lexicalisées, c'est-à-dire répertoriées comme entrées dans les dictionnaires spécialisés du domaine, mais absentes des dictionnaires usuels consultés :

acidification :	*acidification des eaux de surface* (f 19):
	anthropique (f 13)
	des eaux (f 4)
	des eaux douces (f 3)
	des eaux lacustres (f 1)
	expérimentale (f 1)
polluant :	*polluant atmosphérique* (f 14)
	acide (f 3)
	acidifiant (f 1)
	précurseur (f 1)
	primaire (f 1).
pollution :	*pollution acide* (f 10)
	des eaux (f 4)
	physique (f 2)
	des eaux de surface (f 1)
précipitation :	*précipitations acides* (f 165).
sédiments :	*sédiment de fond* (f 1).

Voici maintenant des exemples de collocations et de concurrents, c'est-à-dire des expressions ou groupes de mots non lexicalisés ou en voie de l'être :

acidification :	*acidification du sol/des sols* (f 9)
	des lacs (f 7)
	des précipitations (f 4)
	des eaux de drainage (f 3)
	des rivières (f 2)
	accélérée (f 1), etc.
écosystème :	*écosystème forestier* (f 22)
	lacustre (f 5)
	naturel (f 5)
	fluvial (f 3)

marin (f 3)
terrestre (f 2)
lotique (f 1), etc.

milieu : *milieu aquatique* (f 94)
forestier (f 15)
récepteur (f 14)
acide (f 13)
agricole (f 13)
pollué (f 10)
lacustre (f 4), etc.

pollution : *pollution agricole* (f 21)
diffuse (f 11)
bactériologique (f 9)
organique (f 9)
industrielle (f 8)
chimique (f 7), etc.

espèce : *espèce intolérante/tolérante* (f 40)
piscicole (f 13)
piscivore (f 19)
prédatrice (f 8)
omnivore (f 7)
faunique (f 6), etc.

Il ressort de cette analyse que l'apport d'un corpus de ce type en ce qui a trait à la description lexicographique de mots spécialisés selon leur usage en français québécois est riche et important, et ce, pour tous les aspects de la description lexicographique (élaboration de la nomenclature, analyses contextuelles aux fins de définition, exemples et citations, marques technolectales, référents culturels, données encyclopédiques, etc.).

3. Établir une nouvelle grille de marques d'usage

Plusieurs études ont montré un manque de cohérence et d'uniformité dans le système de marques utilisé dans les dictionnaires usuels (voir Thiboutot, 1998 et Cajolet-Laganière et Martel, 1996). De plus, un large consensus existe quant à la mauvaise adaptation à la situation sociolinguistique du Québec du système de marques utilisé dans les dictionnaires fabriqués en France (voir *Actes du Colloque sur les marques lexicographiques* tenu en novembre 1994 à Montréal). Nombre d'intervenants ont alors affirmé l'importance de travailler avec de vastes corpus et par

type de difficultés (par exemple, des listes par catégories d'anglicismes, de régionalismes, etc.). Il ne fait donc aucun doute pour nous qu'il faille établir une nouvelle grille adaptée au contexte sociolinguistique québécois et que celle-ci doive être appliquée avec cohérence et uniformité (marques normatives, sociolectales, topolectales, technolectales, etc.).

Afin de tester une nouvelle grille de marques sociolectales qui serait mieux adaptée au contexte de communication du Québec, nous avons effectué une enquête par questionnaire, auto-administrée et effectuée par la poste, auprès d'une population de 100 professionnels et professionnelles de la langue répartie dans les régions de Montréal, Québec, Sherbrooke et Trois-Rivières (voir en bibliographie Langlois, 1999). Une liste de 200 mots et expressions, avec leur sens, ont été proposés aux répondants; ces derniers devaient les classer, selon leur usage courant, compte tenu des six registres de langue suivants : *écrit soutenu*, *standard écrit*, *standard oral*, *oral familier*, *oral très familier* et *vulgaire*. Deux autres cases étaient disponibles, soit : « Je pense que cet emploi ou cet anglicisme est critiquable. » et « Je ne connais pas ce mot ou ce sens. ».

Nous avons parallèlement vérifié la présence de ces mêmes 200 mots et expressions dans la BDTS de manière à vérifier l'emploi de ces mots et expressions selon les registres de discours représentés dans la BDTS (oral, littéraire, journalistique, didactique et spécialisé). Globalement, l'étude montre une symétrie entre la perception des répondants et la présence des mots selon les registres de langue dans la BDTS.

Enfin, selon les réponses fournies à la suite d'une question ouverte placée à la fin du questionnaire, formulée comme suit : « Que pensez-vous d'une telle classification des situations de communication? Permet-elle de rendre compte adéquatement du français québécois? Y apporteriez-vous des modifications? Si oui, lesquelles? », les répondants se sont déclarés très favorables à l'application d'une telle grille. Selon ces derniers, il y aurait lieu de marquer les mots des registres *soutenu*, *familier*, *très familier* et *vulgaire*. Les mots relevant des niveaux *standard écrit* et *standard oral* ne devraient pas être marqués, puisqu'il s'agit, selon eux, d'un registre neutre. Les répondants ont en outre demandé que les usages propres au Québec soient distingués de ceux des autres aires de la francophonie, ils ont également demandé que l'on marque les emplois « critiqués », les avis officiels de l'OLF ainsi que les mots et sens propres aux domaines spécialisés.

En guise d'exemple, nous reproduisons quelques résultats obtenus par Mme Langlois lors de son enquête. Aux fins de sa recherche, l'auteure a convenu qu'un mot devait avoir recueilli l'adhésion de 50 % et plus des répondants à l'intérieur d'un même registre pour affirmer qu'il se dégageait un certain consensus ou un point de convergence.

Exemples de mots classés dans les différents registres de langue par les répondants de l'enquête :

Écrit soutenu	Écrit standard+ Oral standard	Oral familier	Oral très familier	Vulgaire
Ire	*Kiwi*	*Avoir de la jasette*	*Coudon*	*Manger de la marde*
Firmament	*Fin de semaine*	*Blonde*	*Garrochage*	*Crossage*
Âtre	*Poudrerie*	*Achaler*	*Câline*	*Fourrer*
Idylle	*Banc de neige*	*Party*	*Baveux*	*Frais chié*
Nordicité	*Caribou*	*Placotage*	*Cheap*	*Pisseuse*
	Tarte au sucre			
	Décrocheur			
	Bois franc			
	Ouananiche			
	Automobile			

4. Développer une méthode de travail informatisée pour toutes les étapes d'élaboration du dictionnaire

Un des objectifs visés par le renouvellement de la pratique lexicographique de notre groupe de recherche consiste en l'informatisation de toutes les étapes de travail, à partir de la saisie du premier mot dans la Banque de données textuelles jusqu'à la diffusion du produit lexicographique dans sa forme finale. Pour ce faire, il nous faut mettre au point une plateforme logicielle intégrée de traitement des données lexicographiques assurant l'informatisation complète du traitement de données textuelles, linguistiques et lexicographiques. Le défi que nous voulons relever est de rendre accessibles de manière directe et simple toutes les sources de consultation (banques de données textuelles, linguistiques, terminologiques, dictionnaire principal, dictionnaires personnalisés, etc.) et la saisie standardisée de textes et de données lexicographiques. Le principe à la base de l'interface logicielle découle d'un choix technologique fondamental en faveur du langage de balisage XML. Cela conditionne l'ensemble des opérations de traitement, de l'acquisition des textes à la diffusion des résultats.

Pour réaliser cette plateforme logicielle intégrée de traitement des données lexicographiques, il nous faut concevoir un moteur de recherche qui permette l'utilisation compatible des diverses composantes logicielles suivantes, dont le noyau principal consistera en un module d'assistance à la saisie des données et à la rédaction des articles du dictionnaire. Ce module est gouverné par un processus intelligent déterminé sous forme de fiche de saisie automatisée et « experte » (une DTD). En clair, les rédacteurs et rédactrices des articles du dictionnaire seront guidés par un processus intelligent de résolution de problèmes; leur travail de rédaction sera fait conformément aux règles édictées et insérées dans ce module de saisie, ce qui assurera à l'ouvrage une grande constance et rigueur et accélérera d'autant le travail de révision et de correction.

Concrètement, la plateforme doit prendre en charge les trois aspects principaux de la pratique lexicographique, soit :

- l'élaboration, la mise à jour et l'exploitation de la Banque de données textuelles de Sherbrooke (BDTS), l'extraction et l'organisation de la nomenclature à partir de la BDTS ainsi que la réorganisation de la nomenclature sous forme de thésaurus;
- la rédaction des articles de dictionnaire et leur organisation interne (entrée, définition, exemples, citations, etc.),
- la préparation, l'édition et la diffusion sur support papier et électronique du produit dictionnairique.

Les rédacteurs et les rédactrices pourront consulter simultanément divers outils dictionnairiques, notamment la BDTS, extraire des données de cette banque (la nomenclature et le thésaurus, les contextes, les citations, etc.) et stocker les informations dans la fiche de saisie. Ils pourront en outre valider les données lexicographiques grâce à un module d'assistance à la rédaction. Enfin, cette plateforme servira à la production et à la diffusion du produit dictionnairique sur de multiples supports (écrits et électroniques).

5. Assurer un transfert technologique

Le dernier objectif visé par le projet concerne le transfert technologique. Nous souhaitons que l'élaboration de notre plateforme logicielle puisse donner lieu à un programme de transfert de technologie, visant à étendre l'application de la méthodologie et des produits à d'autres utilisations. Par

exemple, la constitution de banques de données textuelles dans divers secteurs spécialisés, l'extraction de nomenclatures visant à élaborer ou à mettre à jour les terminologies de pointe, la préparation et l'édition sur support papier et électronique de ces lexiques et vocabulaires dans différents domaines techniques et scientifiques. Un projet est notamment en cours dans ce sens dans le domaine des relations de travail en collaboration avec les différentes centrales syndicales du Québec.

Ajoutons enfin que les données lexicologiques et lexicographiques rassemblées ainsi permettront leur exploitation au regard des nouvelles technologies de l'information et de la communication (TIC). Cela concerne l'enrichissement et la prise en compte de la variété culturelle et linguistique québécoise dans les dictionnaires-machines voués à la correction orthographique des textes. Elles serviront également à la mise à jour et à l'adaptation au contexte québécois et nord-américain des logiciels de traduction, de dictée vocale et de reconnaissance de la parole et enfin aux moteurs de recherche et autres logiciels spécialisés.

Conclusion

À titre de conclusion, rappelons brièvement les principales caractéristiques de la pratique lexicographique que nous tenterons de mettre à profit au cours des prochaines années. Les principaux éléments propres à renouveler la pratique lexicographique québécoise sont :

- l'élaboration d'un dictionnaire conçu en fonction des attentes et des besoins réels des Québécois et des Québécoises;
- la mise à jour constante d'une banque de données textuelles originale et représentative de l'ensemble des usages du français au Québec;
- l'extraction de cette Banque de diverses nomenclatures de la langue générale, lesquelles, une fois validées avec celles des autres dictionnaires usuels français et québécois, pourront être utilisées lors de la confection ou de la mise à jour de divers outils langagiers en format papier ou informatisé;
- la hiérarchisation des différents usages correspondant à la variation sociolectale au Québec;
- l'enrichissement de la nomenclature en ce qui a trait aux domaines de spécialité : définition, synonymes, collocations, cooccurrents, exemples, etc.;

- l'élaboration d'une nouvelle grille de marques sociolectales qui corresponde davantage à la situation sociolinguistique du Québec;
- un développement important de l'informatisation de l'ensemble des étapes allant de la saisie des données à la rédaction du produit dictionnairique, comprenant, entre autres, un module d'assistance à la rédaction des articles de dictionnaire (DTD);
- un souci enfin d'assurer un transfert technologique.

Ces aspects guideront nos réflexions et notre travail lors de l'élaboration de la nomenclature, la validation de la nomenclature avec les dictionnaires usuels existants, l'élaboration du thésaurus, la rédaction des articles du dictionnaire, la mise en forme des données sur support papier et électronique et leur diffusion.

Bibliographie

CAJOLET-LAGANIÈRE, Hélène et Pierre MARTEL (1996). « Les marques de niveaux de langue dans les dictionnaires : synthèse et proposition », dans : Colloque Français du Canada–français de France, *Actes du quatrième Colloque international de Chicoutimi, Québec, du 21 au 24 septembre 1994, Lexicologie canadienne d'hier à aujourd'hui*, Tübingen, Niemeyer, Canadiana Romanica, nº 12, p. 185-197.

CAJOLET-LAGANIÈRE, Hélène et Pierre MARTEL (2001). « L'apport de la Banque de données textuelles (BDTS) : des nomenclatures enrichies », dans : *Colloque Français du Canada–français de France*, Magog (du 26 au 29 septembre 2000), (à paraître).

CAJOLET-LAGANIÈRE, Hélène, Pierre MARTEL et Marie-France LANGLOIS (2001a). « La pertinence culturelle et linguistique dans les outils linguistiques informatisés couramment utilisés au Québec », dans : *Revue belge de Philologie et d'Histoire*, (à paraître).

CAJOLET-LAGANIÈRE, Hélène, Pierre MARTEL et Marie-France LANGLOIS (2001b). « Les textes journalistiques québécois sont-ils "envahis" par les emprunts critiqués à l'anglais? », dans : *Terminogramme*, nº 97-98, p. 47-72.

CAJOLET-LAGANIÈRE, Hélène, Pierre MARTEL et Michel THÉORET (1998). « Les emprunts critiqués à l'anglais : un fait linguistique à évaluer », dans : *Colloque La lexicographie canadienne à l'aube de l'an 2000*, ACFAS, Université d'Ottawa, (à paraître).

CAJOLET-LAGANIÈRE, Hélène, Pierre MARTEL et Michel THÉORET (2000). « Des emprunts à l'anglais différents des deux côtés de l'Atlantique », dans : Colloque Français du Canada–français de France, *Actes du cinquième Colloque international de Bellême, du 5 au 7 juin 1997, Contact de langues-Histoire des contacts*, Tübingen, Niemeyer, Canadiana Romanica, nº 13, p. 207-215.

ESCAYOLA, Inès (2000). *Étude quantitative et qualitative de 45 emprunts de sens dans sept types de discours québécois*, mémoire de maîtrise, Université de Sherbrooke.

LABRECQUE, Geneviève (2001a). « La pratique lexicographique québécoise : le cas des avis officiels émis par l'Office de la langue française », dans : Actes du colloque *La représentation de la norme dans les pratiques terminologiques et lexicographiques*, 69e congrès de l'ACFAS, Université de Sherbrooke, (à paraître).

LABRECQUE, Geneviève (2001b). *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, Université de Sherbrooke.

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

THIBOUTOT, Sylvie (1998). « Inventaire des pratiques lexicographiques françaises et québécoises en matière de marques d'usage », dans : *Les marques lexicographiques en contexte québécois. Actes de la Table ronde tenue à Montréal en novembre 1994*, Coll. « Études, recherches et documentation », Office de la langue française, p. 253-294; volume publié par Louis MERCIER et Claude VERREAULT.

TREMBLAY, Steeve (2000). *Le vocabulaire de l'environnement aquatique : description de quelques unités lexicales*, mémoire de maîtrise, Université de Sherbrooke.

– Dictionnaires

BEAUCHEMIN, Normand, Pierre MARTEL et Michel THÉORET (1992). *Dictionnaire de fréquence des mots du français parlé au Québec. Fréquence, dispersion, usage, écart réduit*, New York, Peter Lang.

BOULANGER, Jean-Claude (DQA) (1993). *Dictionnaire québécois d'aujourd'hui : langue française, histoire, géographie, culture générale*, 2e édition revue et corrigée, (Saint-Laurent), Dicorobert inc.

Centre d'analyse et de traitement informatique du français québécois (CATIFQ) (1996). *Dictionnaire électronique des anglicismes*, Sherbrooke, Université de Sherbrooke.

SHIATY, A.E. (dir.) (1988). *Dictionnaire du français plus à l'usage des francophones d'Amérique*, Montréal, Centre éducatif et culturel.

DULONG, Gaston (1989). *Dictionnaire des canadianismes*, Montréal, Larousse Canada.

FOREST, Constance et Denise BOUDREAU (1998). *Le Colpron. Le dictionnaire des anglicismes*, Montréal, Beauchemin, 4e édition.

OFFICE DE LA LANGUE FRANÇAISE. *Le Grand dictionnaire terminologique*. Adresse électronique : www.granddictionnaire.com

REY, Alain et Josette REY-DEBOVE (réd.) (1993). *Le Petit Robert : dictionnaire alphabétique et analogique de la langue française* (version électronique), Dictionnaire Le Robert (Paris).

VILLERS, Marie-Éva de (1997). *Multidictionnaire de la langue française*, Montréal/Québec Amérique, 3e édition.