
Les apports et les limites de la Banque de données textuelles de Sherbrooke au regard de la description lexicographique du français en usage au Québec : l'exemple du mot *cœur*

Geneviève Labrecque

Université de Sherbrooke

Résumé

Le présent article vise à montrer les apports et les limites de la Banque de données textuelles de Sherbrooke (BDTS) au regard de la description lexicographique du français en usage au Québec (FUQ). La BDTS est un corpus de plus de 52 millions d'occurrences tirées de textes représentatifs des discours spécialisés, littéraires, journalistiques, didactiques et oraux du FUQ. Cette étude s'appuie sur la pratique d'une linguistique à base de corpus et vise à connaître les avantages que procure l'exploitation de corpus à des fins lexicographiques et les difficultés découlant d'une telle démarche. Cet article porte plus particulièrement sur les apports et les limites de la BDTS pour le mot *cœur*. Pour ce mot, les éléments extraits des contextes de la BDTS ont été comparés avec ceux présents dans les articles lexicographiques de dictionnaires. Les résultats montrent comment l'exploitation de la BDTS est utile à la description lexicographique du FUQ.

1. Introduction

Nombre de chercheurs et de linguistes ont écrit sur l'importance de décrire le français en usage au Québec (FUQ) de manière objective et complète, et ce, à partir de données textuelles représentatives de cette variété de langue. Afin que la description lexicographique du FUQ soit un juste reflet de la réalité langagière du Québec, plusieurs soulignent la nécessité de dépouiller un corpus représentatif de la langue écrite et orale québécoise. L'importance d'une banque de données textuelles dans laquelle sont exprimés divers types de discours semble réunir l'avis de plusieurs linguistes.

Nos travaux s'appuient ainsi sur la pratique d'une linguistique à base de corpus, selon laquelle la description lexicographique d'une langue ne peut reposer sur l'étude des mots de façon isolée, mais doit plutôt tenir compte du discours dans lequel les mots s'insèrent. Les corpus sont des traces concrètes de la langue en usage. Leur utilité a maintes fois été démontrée dans des études linguistiques et philologiques de toutes sortes (études lexicologiques, grammaticales, sémantiques, syntaxiques, stylistiques, etc.). Mais qu'en est-il de leur valeur en lexicographie, et plus particulièrement en lexicographie québécoise? Quels avantages procure l'exploitation de corpus à des fins lexicographiques et quelles sont les difficultés ou les limites d'une telle démarche? C'est à ces questions que nous avons tenté de répondre dans notre thèse de doctorat (Labrecque 2005). Le présent article expose quant à lui une partie des résultats de cette recherche.

Il existe à travers le monde différentes banques de données textuelles de langue française (par exemple, Frantext, Suistext, Beltext, Québétext, Lexiquum) et anglaise (par exemple, Bank of English, British National Corpus, International Corpus of English, American Corpus of English). Certaines de ces banques ont servi à l'élaboration d'ouvrages lexicographiques, notamment Frantext pour le *Trésor de la langue française* (TLF) et la Bank of English pour le dictionnaire de langue anglaise *Collins Cobuild English Dictionary*. Comment ces corpus ont-ils été exploités à des fins lexicographiques? Une recension des écrits portant sur les différentes banques de données textuelles françaises et anglaises démontre que rares sont les études ayant été publiées dont l'objet premier est d'explicitier la valeur ajoutée de l'une ou l'autre de ces banques quant à l'élaboration d'un ouvrage lexicographique.

Mentionnons tout de même une publication de langue anglaise, intitulée *Looking Up - An account of the Cobuild Project in lexical computing and the development of the Collins Cobuild English Language Dictionary*, laquelle relate l'expérience de John Sinclair et de ses collègues de l'Université de Birmingham (Sinclair 1987) lors de la rédaction du dictionnaire *Collins Cobuild*. Le point de départ du dictionnaire *Cobuild* est la banque de données textuelles bâtie aux fins de la rédaction du dictionnaire (aujourd'hui appelée Bank of English). L'étude de Sinclair a démontré que la Bank of English a été une source d'information importante pour les articles du *Cobuild*, principalement en ce qui concerne les exemples et les critères de distinction des acceptions des mots. Mise à part cette publication portant sur la langue anglaise, il apparaît qu'aucune étude exhaustive n'a été menée relativement aux apports et aux limites de corpus textuels sur le plan de la description lexicographique du français, notamment du FUQ. Pourtant, l'analyse de corpus est devenue une étape essentielle au processus de description lexicographique.

2. Objectif

Notre thèse s'inscrivait dans le cadre des travaux du Centre d'analyse et de traitement informatique du français québécois (CATIFQ) de l'Université de Sherbrooke, dont l'objectif principal est la description globale du français tel qu'il a cours dans le contexte québécois contemporain. Le CATIFQ dispose d'une banque de données textuelles représentative du FUQ, la Banque de données textuelles de Sherbrooke (BDTS). Le corpus de référence utilisé aux fins de notre étude, la BDTS, correspond au corpus exploité par le groupe de recherche FRANQUS¹ dans le cadre du projet « Le français standard en usage au Québec : nomenclatures, description et application dans les nouvelles technologies de l'information et de la communication »² dirigé par Hélène Cajolet-Laganière et Pierre Martel, professeurs à l'Université de Sherbrooke.

L'objectif de notre thèse consistait à montrer les apports³, mais aussi les limites³, de cette banque au regard de la description lexicographique originale du FUQ⁴, incluant la description de son niveau standard. Le présent article expose une partie des résultats de cette recherche, c'est-à-dire la partie consacrée aux apports et aux limites de la BDTS sur le plan de la description (sémantique, combinatoire, etc.) des mots appartenant au FUQ. Nous avons voulu connaître

¹ Le français au Québec : usage standard.

² Ce projet de recherche vise à décrire le français standard en usage au Québec (FSUQ) à partir de données textuelles représentatives et à montrer qu'il existe au Québec un français de qualité et que ce niveau doit être le modèle proposé à l'ensemble des Québécois. Ce projet se concrétisera par la production d'un dictionnaire général et normatif du FSUQ. Il faut entendre ici par « général » une description de la langue commune aux sujets parlants et par « normatif » une hiérarchisation des usages autour d'un usage valorisé socialement au Québec.

³ Nous entendons par « apport » la contribution positive, la valeur ajoutée de la BDTS. Les « limites » de la BDTS, quant à elles, font référence aux aspects pour lesquels elle ne peut fournir d'information.

⁴ Dans le cadre de notre étude, le FUQ correspond à l'ensemble des usages linguistiques constituant la variété de français utilisée sur le territoire du Québec.

quels types d'informations fournissent les contextes d'emploi de la BDTS (en ce qui a trait, par exemple, aux acceptions des mots, aux cooccurrents avec lesquels les mots sont employés, aux constructions syntaxiques, aux éléments définitoires tirés de certains contextes de la BDTS, etc.). Pour ce faire, nous avons étudié dans notre thèse un échantillon de 25 mots. Cet article porte plus précisément sur les résultats obtenus pour un mot de cet échantillon, *cœur*.

3. Description de la BDTS

Bien que la BDTS totalise à ce jour plus de 52 millions d'occurrences, le corpus utilisé aux fins de nos recherches correspond au noyau de base de la BDTS, lequel contient 37 millions d'occurrences. La BDTS est une banque de textes représentatifs des différents usages du français au Québec, de diverses situations de communication et de divers registres de langue. Afin d'interpréter adéquatement les résultats de notre étude, il est essentiel de bien connaître la composition du corpus de référence. Les 37 millions d'occurrences du noyau de base de la BDTS sont tirés de plus de 10 000 textes contemporains, couvrant les années 1960 jusqu'à 2002, et dans cinq types de discours : 59 % de textes spécialisés, 16 % de textes littéraires, 14 % de textes journalistiques, 6 % de textes didactiques et 5 % de textes oraux.

L'originalité de la composition de la BDTS ressort nettement lorsqu'on la compare aux autres banques de données textuelles francophones existantes, qui sont essentiellement composées de textes littéraires. La typologie de la BDTS s'apparente plutôt au modèle des banques anglo-saxonnes (British National Corpus et Bank of English), lesquelles sont axées vers une représentativité et une variété de types de textes. Par conséquent, il est entendu que tous les emplois lexicaux ou sémantiques rencontrés dans les textes de la BDTS n'appartiennent pas à la langue dite « standard »⁵. La diversité des sources de la BDTS amène forcément la présence de mots, de sens, d'emplois critiqués selon le jugement que portent les usagers sur leur langue. Les textes de la BDTS reflètent donc l'ensemble des usages. Enfin, la BDTS est une banque de taille relativement modeste comparativement à celles des grandes banques européennes de langue française et anglaise (lesquelles peuvent atteindre jusqu'à 500 millions d'occurrences, c'est le cas notamment de la Bank of English), mais sa composition la rend davantage représentative du FUQ.

4. Méthodologie

Dans le but d'illustrer les apports et les limites de la BDTS quant aux différents éléments d'information pouvant être extraits de ce corpus grâce à l'analyse contextuelle et permettant ainsi d'enrichir la description lexicographique actuelle du FUQ, nous avons analysé un échantillon aléatoire de 25 mots. Les éléments extraits des contextes de la BDTS ont été comparés avec ceux consignés dans la description lexicographique de cinq dictionnaires :

- *Petit Robert électronique* 2001 (PRÉ);
- *Petit Larousse illustré électronique* 2002 (PLI);
- *Dictionnaire du français Plus* 1988 (DFP);
- *Dictionnaire québécois d'aujourd'hui* 1993 (DQA);
- *Multidictionnaire de la langue française électronique* 2001 (Multi).

Ces cinq dictionnaires ont été retenus en fonction de certains critères : ce sont des dictionnaires usuels du français en un seul volume, dont deux du français de France (PRÉ, PLI) et deux du français en usage au Québec (DFP, DQA). Le PRÉ est un dictionnaire usuel de langue et, plus précisément, un « dictionnaire alphabétique et analogique de la langue française ». Le PLI,

⁵ La notion de « standard » fait référence aux emplois neutres, non marqués, qui, sans être soutenus, ne donnent prise à aucune évaluation négative.

quant à lui, est un dictionnaire encyclopédique, c'est-à-dire qu'il informe non seulement sur les mots mais aussi sur la connaissance du monde. Le DFP, paru en 1988, est destiné « à l'usage des francophones d'Amérique ». Cet ouvrage se présente comme un dérivé d'un dictionnaire publié à Paris en 1987 par la Librairie Hachette, le *Dictionnaire du français*. De son côté, le DQA est un dictionnaire québécois (1^{re} édition en 1992 et 2^e édition en 1993) qui dérive d'un dictionnaire de la maison Robert : le *Robert Micro*. Dans le DQA, les mots particuliers à la langue française au Québec ne sont pas marqués géographiquement. À la fois dans le DFP et dans le DQA, ce sont les mots caractéristiques du français de France qui sont marqués. Enfin, le *Multidictionnaire* est un ouvrage québécois de difficultés de la langue, dont la première édition remonte à 1988. À noter que cet ouvrage présente une nomenclature réduite par rapport à celle des dictionnaires usuels.

Notre recherche exigeait à la fois quantité (un nombre suffisant de mots à l'étude) et qualité (la finesse de l'analyse des apports et des limites de la BDTS). La taille de l'échantillon devait nous permettre de montrer les différents apports et limites de la BDTS tout en respectant le cadre d'un projet de thèse de doctorat. Comme nous visons un grain d'analyse assez fin des apports et des limites de la BDTS, nous nous sommes restreinte à analyser un petit nombre de mots, en l'occurrence 25.

Par ailleurs, la sélection aléatoire de l'échantillon de mots à l'étude a permis d'éliminer les risques de subjectivité, de présenter un regard plus critique sur l'exploitation de la BDTS en vue de la description lexicographique du FUQ et de montrer non seulement les apports de la BDTS, mais également ses limites et lacunes. La sélection des mots a été dirigée grâce à une table de nombres aléatoires⁶.

L'échantillon ainsi constitué inclut des mots dont la complexité du traitement lexicographique varie (monosémiques et polysémiques); des mots dont la fréquence dans la BDTS est variable (entre 1 et 7529 occurrences); des mots de catégories grammaticales variées; des mots décrits dans un ou plusieurs dictionnaires; des mots relevant de la langue générale ou spécialisée; des mots appartenant à différents domaines d'activités, etc. L'échantillon montre une variété intéressante, compte tenu de sa taille limitée. Aux fins de notre étude, nous avons divisé l'échantillon de 25 mots en trois groupes de mots, selon leur fréquence, afin d'observer progressivement les apports et les limites de la BDTS pour des mots de fréquence faible, moyenne et élevée.

Tableau 1 - Échantillon de 25 mots

hippique	adj.	(4)	abîmer	v.	(194)
Campanule	n.	(58)	regretter	v.	(877)
Permissivité	n.	(20)	placer	v., n.	(4266)
analyser	v.	(2634)	dessous	adv., prép., n.	(1525)
Réévaluer	v.	(155)	transformé	n.	(14)
carthame	n.	(2)	histologie	n.	(3)
anti-inflationniste	adj.	(11)	velléité	n.	(114)
Palynologie	n.	(3)	cœur	n.	(7529)
anglaiser	v.	(1)	désherbage	n.	(13)

⁶ La table de nombres aléatoires est tirée de Moore 1979 : 344.

Soutenable	adj.	(34)	philippique	n.	(5)
blastomère	n.	(1)	idéographique	adj.	(15)
préserver	v.	(1503)	duplicata	n.	(22)
Stéphanois	adj. et n.	(1)			

5. Résultats pour le mot *cœur*

Nous présentons ici les résultats de la recherche, c'est-à-dire les apports et les limites observés à la suite de la comparaison des données extraites des contextes de la BDTS avec celles consignées dans les articles des dictionnaires, et ce, pour le mot le plus fréquent et le plus polysémique de notre échantillon, le mot *cœur*.

Tableau 2 - Fréquence et dispersion du mot *cœur* dans la BDTS

	Spéc.	Journ.	Littér.	Didact.	Oral	Total
Fréquence dans la BDTS	1811	1135	3937	240	406	7529

Notons d'abord que la BDTS fournit des informations sur les types de discours dans lesquels un mot est le plus souvent utilisé. Dans le cas du mot *cœur*, plus de la moitié des occurrences sont attestées dans les textes littéraires, et ce, malgré le fait que les textes littéraires ne représentent que 16 % de l'ensemble des occurrences de la BDTS. La répartition des occurrences dans les différents types de discours constitue un apport en soi de la BDTS, puisque les types de textes permettent de distinguer, par exemple, des emplois fréquents à l'écrit ou à l'oral, des emplois plus souvent utilisés dans des textes spécialisés ou de langue générale, etc. Les types de discours indiquent l'appartenance à un registre de langue en particulier ou à un domaine spécialisé. Ils permettent d'orienter les décisions des lexicographes concernant la pertinence d'une marque de domaine ou d'une marque sociolectale. Dans le cadre de la description lexicographique du FUQ, les lexicographes devraient tenir compte, dans leurs réflexions, non seulement de la fréquence des emplois, mais surtout de leur dispersion dans les différents types de discours.

En outre, parmi les 7529 occurrences du mot *cœur* dans la BDTS, seules 298 sont utilisées au pluriel. Un corpus textuel comme la BDTS montre donc les particularités morphologiques associées à ce mot.

5.1 Les apports de la BDTS pour le mot *cœur*

5.1.1 Actualisation des emplois décrits dans les dictionnaires

Les articles des cinq dictionnaires consultés totalisent un nombre important d'acceptions différentes pour le mot *cœur*. Nous présentons ici les principales acceptions de ce mot (selon la structure de l'article du PRÉ).

Tableau 3 - Principales acceptions du mot *cœur*

<p>I.</p> <p>A.</p> <p>1. Organe central de l'appareil circulatoire. Chez l'homme, Viscère musculaire situé entre les poumons et dont la forme est à peu près celle d'une pyramide triangulaire à sommet dirigé vers le bas, en avant et à gauche. ◇ Cet organe chez certains animaux, faisant partie des abats.</p> <p>2. La poitrine.</p> <p>3. (En loc.) Estomac.</p>
<p>B. Par anal.</p> <p>1. Ce qui a ou évoque la forme du cœur. ◇ Aux cartes, Une des couleurs représentée par un cœur rouge.</p> <p>2. La partie centrale ou active de qqch. — Nucl. Partie (d'un réacteur nucléaire) contenant le combustible et où s'opèrent les réactions de fission.</p>
<p>II.</p> <p>1. Par métaph. Le siège des sensations et émotions.</p> <p>2. Loc. (Le siège du désir, de l'humeur).</p> <p>3. Le siège de l'affectivité (sentiments, passions). ◇ La personne considérée dans ses affections, ses sentiments. — Spécialt Intuition.</p> <p>4. Bonté, sentiments altruistes.</p> <p>5. Vx ou littér. Source des qualités de caractère, siège de la conscience.</p> <p>6. La vie intérieure; la pensée intime, secrète.</p> <p>7. PAR CŒUR : de mémoire.</p>

À la suite de l'analyse contextuelle de ce mot, il ressort que toutes les acceptions de *cœur* décrites dans les ouvrages consultés sont attestées dans la BDTS, ce qui constitue en soi un apport non négligeable. Seules certaines cooccurrences et expressions sont absentes de la Banque (voir le point 4.2 ci-dessous). Même certaines acceptions plus spécialisées, notamment « Partie (d'un réacteur nucléaire) contenant le combustible et où s'opèrent les réactions de fission », sont présentes dans la Banque.

« Tout risque de contamination ou d'irradiation sont écartés, a assuré le capitaine Leutard. Il n'y a aucun contact physique ni géographique entre le cœur nucléaire et le compartiment turbo-alternateurs. » « Le combustible nucléaire est contenu dans des gaines étanches qui chauffent un circuit de liquide dit primaire, lequel sert lui-même à vaporiser un deuxième circuit également totalement isolé que l'on appelle eau secondaire » [...]
 Journalistique 008, Le Monde - La Presse - 31-03-94

Il arrive en outre que des emplois marqués « vx » ou « vieilli » dans les dictionnaires aient des attestations dans la BDTS, c'est le cas pour le mot *cœur* de l'acception « Vx ou littér. Source des qualités de caractère, siège de la conscience ». Il s'agit ici d'un apport intéressant de la Banque, compte tenu de sa nature synchronique. À noter qu'un des deux contextes cités ci-dessous provient d'un texte littéraire.

En effet, nous avons lutté et nous luttons encore. L'obéissance est parfois dure aux **cœurs bien nés**.
 Littérature 019, Les Plouffes (1987) - Roger Lemelin

Le don de soi et le sacrifice pour la patrie sont glorifiés. La **noblesse du cœur** et le stoïcisme règnent au détriment de la sensibilité et de l'angoisse qui seront l'apanage des romantiques.
 Culture et comm 006, Francine Girard - Apprécier l'œuvre d'art - 1995

L'apport le plus important de la BDTS réside dans l'actualisation des emplois répertoriés dans les dictionnaires. En plus de valider l'utilisation des mots eux-mêmes, la BDTS valide aussi les acceptions consignées dans les ouvrages, les cooccurrents illustrés dans les exemples des dictionnaires, les sèmes présents dans les définitions (par le biais notamment de contextes métalinguistiques), les réseaux analogiques, les domaines spécialisés auxquels certaines acceptions appartiennent (validation des marques par le biais des types de discours), les niveaux de langue (validation par le biais des types de discours), etc. À la suite de l'analyse du mot *cœur*, nous pouvons donc affirmer que la BDTS joue un rôle important d'actualisation, et ce, même pour un mot très polysémique et traité de façon complexe dans les dictionnaires. Ce premier constat a également été démontré lors de l'analyse de l'ensemble des 25 mots à l'étude dans notre thèse, autant pour les mots monosémiques que pour les mots polysémiques de l'échantillon.

5.1.2 Données originales de la BDTS

En plus d'actualiser les données des articles des dictionnaires, la BDTS ajoute des données originales, autant par rapport aux dictionnaires français que québécois. Il ressort que la BDTS est une importante source d'enrichissement pour la description lexicographique des mots appartenant au FUQ. De fait, les textes de la BDTS fournissent des apports directs, c'est-à-dire des données qui peuvent servir à la rédaction lexicographique ou qui peuvent directement être intégrées à un article lexicographique. L'exploitation du corpus soulève également des pistes de réflexion pouvant servir à établir les règles d'une politique éditoriale d'un dictionnaire et orienter ainsi certaines décisions touchant des aspects de la pratique lexicographique.

De façon plus précise, les apports « différentiels » de la BDTS que nous avons relevés pour le mot *cœur* touchent les cooccurrents, les expressions, les extraits de textes pouvant servir de citations, les acceptions, les éléments définitoires, les constructions syntaxiques, les particularités morphologiques, les actants, les marques (spécialisées, sociolectales, temporelles, etc.), les emplois en concurrence et les emplois critiqués, etc. À noter que ces types d'apports ont également été observés pour les autres mots de l'échantillon à l'étude dans notre thèse.

5.1.2.1 Cooccurrents

Parmi tous ces types d'apports, la richesse de la BDTS en termes de cooccurrents est nettement ressortie (par rapport à ceux illustrés dans les exemples des dictionnaires). Par exemple, dans le corpus, l'expression *de bon cœur* est souvent utilisée avec le verbe *rire*, ce que n'illustrent pas les ouvrages dans l'article *cœur* (seule une citation du PRÉ sous *rire* atteste ce cooccurrent). De plus, la BDTS montre que l'emploi de *au cœur de* a un sens plus étendu que celui décrit dans les dictionnaires, puisque les cooccurrents sont beaucoup plus variés dans les contextes que les exemples des dictionnaires, qui n'illustrent que *au cœur de l'été*, *de l'hiver*.

La foule applaudit et rit **de bon cœur**.
Journalistique 017, André Noël - La Presse - 11-09-93

Cette halte de tranquillité, **au cœur de** la ville si nerveuse, me calmait presque toujours.
Littérature 035, La détresse et l'enchantement (1996) - Gabrielle Roy

Cependant, devant l'accroissement des besoins, l'accessibilité des services sociaux et de santé demeure **au cœur de** leurs préoccupations.
Santé 002, Financement système public de santé - Juin 2000

Pour cela, il faut reconnaître que la langue française est **au cœur de** l'identité québécoise et qu'elle est aussi le fondement de la cohésion de la société québécoise.

Métalinguistique 003, Le français québécois : usages et aménagement (1996)

La BDTS recense donc une grande variété de cooccurrents. Toutefois, il paraît évident qu'un corpus textuel puisse attester un plus grand nombre de cooccurrents que ceux illustrés dans les exemples des ouvrages de référence. Les apports de la BDTS sur le plan des cooccurrents doivent donc être nuancés en tenant compte du fait que, dans notre étude, deux différents types de données ont été comparées : les dictionnaires, qui présentent une description limitée des mots, et la BDTS, qui renferme des textes bruts non traités. Les contextes de la BDTS illustrent les mots en action dans le discours. Dans le cadre de la description lexicographique du FUQ, les cooccurrents usuels fréquents tirés de la BDTS pourraient être mentionnés dans les articles sous forme d'exemples. Évidemment, il peut être nécessaire de les reformuler ou de les neutraliser avant de les insérer dans un article. Les exemples servent à montrer véritablement l'usage et l'utilisation du mot, de l'expression ou de l'acception en question. En s'inspirant des cooccurrents fréquents de la BDTS, les lexicographes s'assurent de forger des exemples réalistes, courants, qui constituent les lieux communs de l'emploi du mot.

5.1.2.2 Expressions

L'apport de la BDTS est riche non seulement en ce qui a trait aux cooccurrents nouveaux, mais également pour ce qui est des expressions nouvelles (plus figées que les simples cooccurrents) non répertoriées dans les dictionnaires. Pour le mot *cœur*, c'est le cas notamment de :

- *peser sur le cœur,*
- *en plein cœur de,*
- *avoir le cœur gai,*
- *le cœur lui débat,*
- *entre les deux mon cœur balance,*
- *se marcher sur le cœur,*
- *ça fait chaud au cœur,*
- *la mémoire du cœur,*
- *le cœur n'y est plus (ou pas),*
- *lever le cœur,*
- *tomber sur le cœur,*
- *avoir le cordon du cœur slack (ou trop long),*
- *avoir le cœur là où la poule a l'œuf.*

Voici quelques contextes d'expressions attestées dans la BDTS, mais non répertoriées dans les dictionnaires consultés.

Les chagrins d'amour. Tout cela fait une boule, une boule dure qui, avec le temps, **pèse** de plus en plus lourd **sur le cœur** et nous empêche de respirer.
Littérature 013, La memoria (1996) - Louise Dupré

Le feu dévaste deux millions d'acres de forêts dans le parc de Yellowstone **en plein cœur des** Rocheuses.
Littérature 078, L'assaut contre les vivants (1994) - Paul Chamberland

Sur la plus haute branche, le rossignol chantait Chante, rossignol, chante, Toi qui **as le cœur gai**... Il y a longtemps que je t'aime, Jamais je ne t'oublierai...
Littérature 056, Maria Chapdelaine (1983) - Louis Hémon

Mais partout où il trouvait à s'asseoir, il s'effondrait, le pauvre! et **le cœur lui débattait** à l'idée de ce mariage qui s'en venait détruire le dernier espoir de son sang.

Littérature 057, Menaud maître-draveur (1937) - Félix-Antoine Savard

Louer ou acheter une auto d'occasion? **Entre les deux, votre cœur balance?**

Divers travail 002, OPC Guide pratique de l'automobiliste (1999)

En attendant, il dit à Marise que, le jour où Erik apprendrait la vraie nature de leurs relations, il reconnaîtrait sa méprise. - Comment peux-tu me parler ainsi? Tu **te marches sur le cœur** pour m'encourager de la sorte. Merci, Brian !

Littérature 031, Neige (1998) - Florence Nicole

Ça fait chaud au cœur d'être entouré de tendresse comme ça.

Littérature 012, Tit-Coq - Théâtre (1994) - Gratien Gélinas

Les accents qui avaient bercé mes premiers émois, tout au long des années soixante-dix, **la mémoire du cœur** y restait fidèle, mais je ne m'y connaissais guère et je m'en foutais pas mal.

Littérature 036, Vamp : roman (1995) - Christian Mistral

Guéri, Jean Lapointe reprend le collier, mais **le cœur n'y est plus**.

Actualité 003, Micheline Lachance - L'Actualité 15 octobre 1995 - 82

Précisons que l'expression *soulever le cœur de qqn* est illustrée dans les dictionnaires et présente dans la BDTS. Toutefois, *lever le cœur* ou *le cœur me lève* (non répertoriées dans les dictionnaires) sont plus fréquentes que *soulever le cœur* dans la Banque. On trouve même quelques occurrences de *tomber sur le cœur*, une nouveauté de la BDTS absente des ouvrages.

L'odeur du cigare froid me **lève le cœur**, mais je suis si impressionnée que je n'ai pas la hardiesse de changer d'endroit.

Littérature 048, Une enfance à l'eau bénite (1985) - Denise Bombardier

Le cœur me lève maintenant à la seule vue d'un pot.

Littérature 050, Juliette Pomerleau (1989) - Yves Beauchemin (Partie 1)

Une odeur de propreté qui commence à me **tomber sur le cœur**.

Littérature 003, Le Cassé et autres nouvelles - Jacques Renaud (1990)

À noter que, parmi toutes les expressions nouvelles attestées dans la BDTS (énumérées ci-dessus), aucune n'est illustrée dans un grand dictionnaire comme le TLF. S'agit-il alors d'expressions caractéristiques du FUQ? Comment s'assurer que ces expressions relèvent du niveau de langue standard? La dispersion dans les types de discours de la BDTS est un indice dont il faudrait tenir compte afin de déterminer leur « recevabilité ». Certaines sont attestées dans des textes littéraires, par de bons auteurs, tandis que d'autres sont attestées uniquement à l'oral ou dans des textes reflétant l'oral. Dans ce dernier cas, les expressions sont souvent de niveau familier et certaines peuvent être calquées sur l'anglais.

Par exemple, pour certaines expressions, les contextes de la BDTS montrent bien qu'il s'agit d'emplois familiers et caractéristiques du FUQ, c'est le cas de *avoir le cordon du cœur slack* (ou *trop long*), attestée deux fois dans le corpus, dont une occurrence dans un texte métalinguistique,

suivie d'une courte définition, et de *avoir le cœur là où la poule a l'œuf*, présente une fois dans un texte métalinguistique. Rappelons que ces expressions ne sont pas décrites dans les dictionnaires à l'étude. Dans ces deux cas, la BDTS propose des contextes définitoires par le biais de répertoires métalinguistiques stockés dans la Banque.

Y a le mal imaginaire pis le système à l'envers Et des gaz dans l'estomac, j'y ai donné un p'tit peu d'soda Comme il s'pensait consommation, y a eu une consultation Le docteur dit: « Mon garçon, **t'as l'cordon du cœur trop long.** »

Littérature 027, Madame Bolduc (1995) - Lina Remon

Avoir* le cordon* du cœur* slack* (aussi: ... la corde* du cœur) déf : se dit de qqn de paresseux à l'ouvrage [...]

Textes divers 002, Expressions figurées Québec (1982) - Normand Beauchemin

Avoir* le cœur* là où la poule* a l'œuf* déf : Se dit de qqn qui n'a pas de cœur, c.à.d. qui n'a pas de courage ou qui n'a pas de sentiments humains. Exp : la dernière partie « ... là où la poule a l'œuf » s'utilise comme euphémisme, pour remplacer le mot « cul » ou « derrière », dans un certain nombre d'autres expressions [...]

Textes divers 002, Expressions figurées Québec (1982) - Normand Beauchemin

5.1.2.3 Constructions syntaxiques

La BDTS apporte des informations supplémentaires sur la syntaxe de certains emplois par rapport aux informations données dans les dictionnaires. Par exemple, les contextes du corpus montrent que les expressions *de gaieté de cœur* et *avoir le cœur à faire qqch.* ou *avoir le cœur de faire qqch.* sont attestées surtout à la négative dans le corpus. Pour cette dernière expression, les deux prépositions *à* et *de* sont utilisées dans la BDTS.

Ce n'est pas **de gaieté de cœur** que j'ai quitté un emploi permanent à la fonction publique du Québec [...]

Mémoires 111-140, Gilles Hébert - Montréal - 14 novembre 2000

Mais Elisabeth n'**a pas le cœur à** la fête.

Littérature 040, Quelques adieux (1997) - Marie Laberge

Je pense que ça aurait été une belle maison! À c't'heure, j'**ai pus le cœur de** la finir.

Littérature 020, Les Filles de Caleb I (1995) - Arlette Cousture

De plus, la BDTS illustre l'emploi des prépositions, notamment après un verbe (ex. : dans l'expression *serrer qqn sur son cœur* ou *contre son cœur*). Dans le PLI, le DFP et le Multi, seule la préposition *sur* est illustrée dans les exemples (*serrer qqn sur son cœur*). Le PRÉ et le DQA illustrent en plus l'emploi de la préposition *contre* (*serrer qqn contre son cœur*). Dans la BDTS, les deux prépositions sont présentes (*sur* étant un peu plus fréquente que *contre*).

Également, pour l'expression *serrer qqn sur, contre son cœur*, on constate dans les contextes qu'on peut serrer « quelque chose » et non seulement « quelqu'un » sur son cœur, comme l'illustrent tous les exemples des dictionnaires. La BDTS montre que les emplois de cette expression ne sont pas restreints aux exemples fournis dans les dictionnaires. L'apport de la

Banque touche ici les actants du verbe, les types d'objets du verbe, soit une personne, soit une chose.

Les invités repartis, la fillette s'endort en serrant **sur son cœur** son précieux cadeau.

Arts visuels 001, Pascale Bréniel - La Presse 12 décembre 1992 - E5

[...] mais Philippe était nerveux, inquiet et la respiration régulière de Marcel réfugié **contre son cœur** lui fournissait cette confiance en soi qui lui manquait tant [...]

Littérature 058, La grosse femme... (1986) - Michel Tremblay

5.1.2.4 Marques

Par ailleurs, l'exploitation de la BDTS amène une réflexion quant au niveau de langue de certaines expressions, selon la répartition de leurs occurrences dans les types de discours; par exemple, *crise de cœur* n'apparaît qu'en langue orale et *avoir une crotte sur le cœur* est attestée dans un monologue. La Banque permet parfois de remettre en question la pertinence de certaines marques sociolectales, puisque certaines expressions ne semblent pas familières dans les contextes, bien qu'elles soient marquées ainsi dans certains dictionnaires, c'est le cas de *la bouche en cœur* (marque « fam. » dans le PRÉ et le DQA), *avoir le cœur sur la main* (marque « fam. » dans le PRÉ) et *à cœur de jour, de journée* (marque « fam. » dans le DQA). Par le biais des différents types de discours, les contextes de la BDTS fournissent des informations quant au niveau de langue des expressions.

Il inclina le buste et **la bouche en cœur** et il susurra : - Ces demoiselles sont bien bonnes de s'occuper de moi.

Littérature 026, Jules Faubert, le roi du papier (1991) - Ubald Paquin

Aussi, le brave monsieur Flavel, en homme bien élevé et accueillant, **le cœur sur la main**, avait fait des frais.

Littérature 030, Marie-Calumet (1990) - Rodolphe Girard

Joséphine songeait avec terreur à ces chansons d'amour passionné qu'elle fredonnait **à cœur de journée** sans songer à mal, et se demandait si ce n'était pas là l'origine de ses malheurs.

Littérature 019, Les Plouffes (1987) - Roger Lemelin

5.1.2.5 Emplois en concurrence

La BDTS apporte aussi des informations sur la fréquence des emplois en concurrence ou critiqués, notamment entre *crise de cœur* et *infarctus*. L'expression *crise de cœur* est répertoriée uniquement dans le DQA, avec la marque « fam. » et suivie de la remarque « Cet emploi est critiqué ». Comme les seules occurrences de *crise de cœur* dans la BDTS se trouvent en langue orale (quatre attestations), le niveau de langue familier est bien ressenti. L'emploi standard, *infarctus*, est quant à lui attesté 48 fois dans tous les types de textes de la BDTS.

Mon père avait fait une **crise de cœur** au début décembre, pis on est allé le chercher à l'hôpital le 24 décembre, à six heures.

Enquêtes orales 001, Enquête 11 - Princeville Femme A21 S16 TV8

Le traitement des emplois non standards est une problématique importante de l'exploitation de la BDTS. Rappelons que cette banque reflète l'ensemble des usages des Québécois, y compris des usages non standards. La fréquence d'un emploi dans la BDTS n'est donc pas toujours un indice de son caractère standard. Le fait de trouver dans la BDTS des emplois appartenant à

divers registres de langue permet d'enrichir la description lexicographique du FUQ. La BDTS soulève en effet des pistes de réflexion à savoir comment traiter les emplois non standards, notamment au regard de leur fréquence mais aussi de leur dispersion dans les types de discours.

5.2 Les limites de la BDTS pour le mot cœur

Dans un autre ordre d'idées, l'analyse du mot *cœur*, et des autres mots de l'échantillon, a permis de dégager des limites de la BDTS, lesquelles s'illustrent par l'absence d'attestations de certaines cooccurrences et expressions mentionnées dans l'un ou l'autre des dictionnaires consultés :

- *greffe du cœur* (PRÉ, Multi),
- *tant que mon cœur battra* (PRÉ; littér.),
- *avoir le cœur barbouillé* (PRÉ, DQA),
- *j'ai encore mon souper sur le cœur* (DQA),
- *cœur-de-pigeon* (PRÉ),
- *cœur à la crème* (PRÉ),
- *un fromage fait à cœur* (PRÉ, PLI),
- *cœur de cible* (PRÉ; fig.),
- *came en cœur* (DFP; tech.),
- *cœur épris* (PRÉ),
- *c'est un cœur d'artichaut* (PRÉ, DFP),
- *redonner du cœur à qqn* (PLI).

Certains proverbes ou maximes sont aussi absents du corpus :

- *loin des yeux, loin du cœur* (PRÉ),
- *le cœur a ses raisons que la raison ne connaît point* (PRÉ),
- *à cœur vaillant rien d'impossible* (PRÉ).

Comment expliquer l'absence de toutes ces expressions dans la BDTS? La plupart d'entre elles ne sont consignées que dans un dictionnaire, ce qui montre qu'il ne s'agit pas d'emplois très courants. L'absence de certaines expressions dans la Banque peut s'avérer un élément d'information pertinent qui ne découle pas d'une lacune du corpus. De fait, si certains emplois sont absents de la Banque, on peut supposer qu'ils ne sont pas courants dans le FUQ. Cette absence renseigne quant à l'intérêt de les inclure ou non dans la description lexicographique du FUQ. Par contre, d'autres expressions semblent plus courantes, mais sont tout de même absentes de la BDTS. C'est le cas notamment des trois proverbes ou maximes mentionnés ci-dessus. Les raisons qui expliquent cette absence sont dans ces cas plus difficiles à expliquer. La BDTS se présente comme un corpus représentatif de l'usage du français au Québec, mais il serait faux de croire qu'il couvre la totalité de la langue. Nous en avons la preuve ici. L'absence d'attestations d'une expression dans la BDTS ne permet pas de conclure que celle-ci n'est pas en usage dans le français au Québec.

Certaines expressions absentes de la BDTS sont répertoriées uniquement dans un ou des dictionnaires français (le PRÉ ou le PLI) et peuvent être considérées comme étant caractéristiques du français de France (mais pas dans tous les cas), par exemple *tant que mon cœur battra*, *cœur-de-pigeon*, *cœur à la crème*, *un fromage fait à cœur*, *cœur de cible*, *cœur épris* et *redonner du cœur à qqn*. La question des emplois caractéristiques du français de France (décrits uniquement dans le PRÉ ou le PLI) doit être prise en considération lors de l'interprétation des lacunes de la BDTS. On peut s'attendre à ce que ce type d'emplois soit absent d'un corpus de textes québécois comme la BDTS. Ces absences ne peuvent donc pas être

identifiées comme des limites de la BDTS; elles doivent plutôt être interprétées comme des informations intéressantes à analyser de la part des lexicographes au sujet des emplois propres à certaines aires de la francophonie. Des recherches complémentaires dans des corpus français pourraient également apporter des informations additionnelles.

À noter que, parmi les expressions absentes de la Banque, *j'ai encore mon souper sur le cœur* n'est répertoriée que dans le DQA. Comme il s'agit d'un dictionnaire adapté pour le Québec, on peut supposer que cet emploi est caractéristique du FUQ, même si aucune marque ne l'accompagne dans le DQA. Son absence dans la BDTS montre alors les lacunes du corpus pour des emplois québécois et amène à approfondir les recherches. En outre, comme *came en cœur* n'est consigné que dans le DFP, on peut s'interroger sur le fait qu'il s'agisse d'un emploi caractéristique du FUQ ou plutôt d'un « héritage » du dictionnaire Hachette, à partir duquel le DFP a été adapté. Mentionnons que *came en cœur*, « excentrique en forme de cœur », relève d'un vocabulaire technique. Étant donné que la BDTS est essentiellement un corpus de langue générale, certains domaines spécialisés ne sont pas couverts. Le manque de textes traitant de domaines spécialisés particuliers dans la BDTS peut expliquer l'absence ou la faible fréquence de vocabulaires spécialisés. Rappelons que la composition de la Banque vise à couvrir les différents types de discours du FUQ et non l'ensemble des vocabulaires spécialisés. L'interprétation des limites de la BDTS doit donc tenir compte du contexte global de la Banque, au regard des objectifs visés par les auteurs, et de la composition du corpus, laquelle peut expliquer certaines absences. De façon générale, lors de l'analyse des 25 mots échantillonnés dans le cadre de notre thèse, les principales limites observées touchent les emplois marqués (spécialisés, vieilliss, littéraires, etc.), qui s'avèrent souvent décrits dans une minorité de dictionnaires (la plupart du temps le PRÉ ou le PLI).

6. Conclusion

Pour atteindre notre objectif, qui consistait à montrer les apports et les limites de la BDTS au regard de la description lexicographique du FUQ, nous avons comparé, pour un échantillon de 25 mots, les informations recueillies dans les dictionnaires consultés avec celles extraites de la BDTS. Le nombre élevé de contextes sur lequel s'est appuyée l'analyse du mot *cœur* a permis de tirer des conclusions intéressantes et valables. Rappelons que le mot *cœur* est le mot le plus fréquent et le plus polysémique de notre échantillon; il est donc possible de généraliser les observations faites pour ce mot.

Lors de cette étude, la valeur ajoutée de la BDTS a été mise en évidence, de même que les limites de son exploitation. L'intérêt d'exploiter des corpus textuels à des fins lexicographiques réside d'abord et avant tout dans l'actualisation dans le discours des emplois déjà consignés dans les dictionnaires. En effet, pour le mot *cœur*, les emplois extraits des textes de la BDTS correspondent, dans l'ensemble, aux emplois répertoriés dans les dictionnaires consultés. Ce constat, dégagé non seulement pour le mot *cœur* mais également pour les autres mots à l'étude, montre que le français utilisé dans les textes de la BDTS correspond, de façon générale, au français décrit dans les ouvrages de référence.

Les apports de la BDTS que nous avons observés pour le mot *cœur* concernent principalement les cooccurents, les expressions, les éléments définitoires, les constructions syntaxiques, les particularités morphologiques, les marques de divers types, les emplois en concurrence, etc. Ajoutons que les contextes de la BDTS constituent non seulement une source intéressante d'exemples, mais également, grâce aux œuvres littéraires qu'elle contient, une source de citations littéraires pouvant être insérées dans un article lexicographique. Par le biais

d'attestations d'emplois caractéristiques du FUQ, la BDTS permet d'enrichir la description du français dans un contexte québécois ou nord-américain.

À la suite de notre analyse, les nouveautés de la BDTS ne seraient évidemment pas toutes retenues dans le cadre d'une description lexicographique. Tous les apports mis en lumière dans cette étude ne pourraient, à l'intérieur d'un ouvrage lexicographique, être reflétés dans leurs moindres détails. On ne peut évidemment rendre compte de toutes les nuances d'emploi tirées des contextes de la Banque. Dans le cadre de la description du niveau standard du FUQ, il faudrait s'en tenir à ce qui est récurrent et conforme à l'usage standard. De fait, la BDTS contribue à dégager la norme par l'observation des usages; toutefois, il faut faire preuve de prudence, puisque la fréquence élevée d'un emploi dans le corpus ne témoigne pas nécessairement de son caractère standard. Il importe ainsi de réfléchir à l'intégration concrète des apports de la BDTS dans la description lexicographique du FUQ.

Étant donné que les principales limites auxquelles nous avons fait face, autant pour le mot *cœur* que pour les autres mots de l'échantillon, sont liées à des emplois décrits dans un ou deux dictionnaires seulement (parfois des emplois marqués) ou décrits uniquement dans les dictionnaires français, nous croyons que la BDTS est un corpus utile pour appuyer la description lexicographique usuelle et contemporaine du FUQ (plutôt que la description lexicographique d'un vocabulaire spécialisé ou historique). Les limites de la BDTS (l'absence de certaines données) constituent en elles-mêmes des éléments de réflexion pour les lexicographes, puisqu'elles orientent leurs choix lors de la description des mots.

Au terme de cette étude, nous concluons que la description lexicographique du FUQ doit s'appuyer sur une banque de données textuelles représentative du FUQ. L'analyse du mot *cœur* et des autres mots de l'échantillon a montré qu'un tel corpus, comme la BDTS, est une importante source d'enrichissement pour la description lexicographique du FUQ. Évidemment, comme tout corpus ne saurait être exhaustif, son exploitation se doit d'être complémentaire à la consultation d'ouvrages de référence et d'autres corpus.

7. Bibliographie

- CAJOLET-LAGANIÈRE, Hélène, Geneviève LABRECQUE, Pierre MARTEL, Louis MERCIER et Michel THÉORET (à paraître). « Dictionnaires usuels du français et Banque de données textuelles de Sherbrooke (BDTS) : convergence et divergence des nomenclatures », Actes du 7^e colloque international *Français du Canada - français de France*, Lyon, juin 2003.
- Dictionnaire du français Plus : à l'usage des francophones d'Amérique* (1988). Rédacteur principal : Claude Poirier, avec le concours de Louis Mercier et de Claude Verreault, Montréal, Centre éducatif et culturel inc.
- Dictionnaire québécois d'aujourd'hui : langue française, histoire, géographie, culture générale* (1993). Rédaction dirigée par Jean-Claude Boulanger et supervisée par Alain Rey, 2^e édition revue et corrigée, Paris-Montréal, Le Robert-Dicorobert inc.
- LABRECQUE, Geneviève (2005). *Les apports et les limites de la Banque de données textuelles de Sherbrooke au regard de la description lexicographique du français en usage au Québec*, Thèse de doctorat, Université de Sherbrooke, 387 p.
- Le Nouveau Petit Robert : dictionnaire alphabétique et analogique de la langue française* (2001). Nouvelle édition du *Petit Robert* de Paul Robert, texte remanié et amplifié sous la direction de Josette Rey-Debove et Alain Rey, Paris, Le Robert (version 2.1 sur CD-ROM 2001).
- Le Petit Larousse illustré 2002* (2001). Paris, Larousse (version sur CD-ROM 2001).
- MOORE, David S. (1979). *Statistics : Concepts and Controversies*, New York, W. H. Freeman and Company, 350 p.
- SINCLAIR, John M. (1987) (éd.). *Looking Up - An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary*, Londres, HarperCollins, 182 p.
- Trésor de la langue française : dictionnaire de la langue du XIX^e et du XX^e siècle (1789-1960)* (1971-1994). Publié sous la direction de Paul Imbs, Paris, Éditions du CNRS.
- VILLERS, Marie-Éva de (2001). *Multidictionnaire de la langue française*, Montréal, Québec Amérique (version 1.0 sur CD-ROM 2001)