

Le rôle du marquage lexicométrique des anglicismes dans un corpus franco-canadien

Liliane Rodriguez

Department of Modern Languages and Literatures

The University of Winnipeg – R3B 2E9 Canada

Résumé

Cette présentation s'appuie sur nos trois enquêtes de disponibilité lexicale réalisées au Manitoba (Canada) entre 1990 et 2006. La première a donné une occurrence totale de l'ordre de 130 000 items conduisant à un corpus lemmatisé de 19.250 lemmes. L'occurrence des deux autres enquêtes est de l'ordre de 30.000. Chacun des items dispose d'un indice lexicométrique de disponibilité. A la saisie des données, chaque item a été marqué de plusieurs balises (âge, sexe, langue parlée à la maison, etc.). C'est la balise des anglicismes qui fait l'objet de cette présentation, ainsi que trois manières d'utiliser ce balisage pour mettre au jour les processus de l'anglicisation, et de la résistance à l'anglicisation, dans le cas d'une langue minoritaire. Loin d'être d'un seul recours numérique, le marquage lexicométrique des anglicismes ouvre à la dynamique du contact des langues.

Abstract

This paper is based on our fieldwork in Manitoba (Canada), with three Availability Surveys taking place between 1990 and 2006. From the first fieldwork, a lemmatized corpus of 19.250 lemmas was created, based on a total occurrence of over 130 000 items. The total occurrence for the other two is over 30.000 items. Each lemma has an Availability Index. When processed, each item from the raw data was tagged with several parameters (age, sex, home language, etc.). Anglicism tagging is the subject of this paper, along with three different ways of examining it. Such examination reveals the Anglicization process, and the resistance it meets in the context of a minority language. Far from being strictly quantitative, Availability Indexes are a tool in exploring languages in contact.

Keywords: lexicometry, lexical availability, languages in contact, anglicisms

1. Langues en contact et lexicométrie

La lexicométrie est un outil précieux pour étudier une langue en situation minoritaire (le français) dans un milieu où la langue de contact (l'anglais) se trouve être aussi la langue de la mondialisation. Nos enquêtes de disponibilité lexicale, réalisées au Manitoba entre 1990 et 2006, en 12 points différents de cette province canadienne, ont donné des corpus lexicométriques qui permettent une description du français en usage au Manitoba, dans sa variation diatopique et diastratique. Ces corpus ouvrent des possibilités d'analyse très variées, notamment en lexicologie, lexicographie et pédagogie. Ici, nous les mettons à l'œuvre pour cerner des effets de contact entre langues (anglicismes) et des phénomènes linguistiques liés à la survie d'une langue minoritaire.

Le Manitoba, au centre du Canada, compte 50.000 personnes de langue maternelle française, soit 5% de sa population, et 15% de bilingues français-anglais. L'entrée du Manitoba dans la

Confédération canadienne, en 1870, a été marquée par la reconnaissance constitutionnelle de ses deux langues officielles, le français et l'anglais. Mais, peu après, deux lois provinciales antifrancophones ont rayé le français des pratiques parlementaires (en 1890), et des écoles (en 1915). Ce n'est qu'entre 1970 et 1980 que l'usage officiel du français fut rétabli, par les lois linguistiques fédérales, et par la Cour suprême du Canada, qui jugea anticonstitutionnelles les lois provinciales de 1890 et de 1915. C'est dire l'endurance sociale et linguistique dont firent preuve les Franco-Manitobains, pendant près d'un siècle, pour sauver leur langue.

Aujourd'hui, la cohabitation des deux langues sur un même territoire conduit inévitablement à des interférences linguistiques. Comme le français est ici langue minoritaire, en nombre sinon en droit, ses locuteurs ne peuvent échapper à l'anglais ambiant, dans les milieux du travail, de la santé, des loisirs, etc. Même en faisant abstraction des médias anglophones prédominants, de simples situations quotidiennes exposent à la langue majoritaire. Il est donc important de mesurer scientifiquement l'ampleur du contact entre les deux langues. C'est pourquoi cette présentation porte spécifiquement sur les indices lexicométriques des anglicismes.

2. Le concept de disponibilité

La *disponibilité* est un concept mis au point par Michéa (Gougenheim et al., 1964)¹, afin de cerner le vocabulaire fondamental, le plus utile à enseigner. Ce concept et la recherche afférente sur le français fondamental se sont révélés très féconds, et ont été salués par plusieurs linguistes².

Comparée aux autres méthodes de statistique lexicale (fréquence absolue, répartition, rang associatif et valence), la recherche de la disponibilité (ou fréquence relative au contexte) semblait la plus appropriée pour enquêter sur le vocabulaire d'une langue en situation minoritaire. Pour saisir la spécificité de la disponibilité, il faut comprendre sa relation avec la fréquence. La disponibilité s'oppose à la fréquence et la complète. La fréquence se recherche en s'appuyant sur des banques de données, pouvant atteindre des millions d'unités, généralement écrites, extraites de textes de genres variés. Pour chaque texte sont calculés le nombre total d'items lexicaux (N) et le nombre de leurs occurrences (n), dit fréquence absolue. La fréquence est idéale pour mettre en évidence la variation stylistique chez un auteur (Muller, 1979).

Or, il se trouve que les items lexicaux les plus fréquents sont les auxiliaires, les articles, les connecteurs et actualisateurs, les modaux et les génériques. Michéa remarqua que certains items, connus de tous, généralement des noms concrets, restent absents ou de très faible fréquence dans les corpus de fréquence (même les corpus quantitativement importants). En outre, ces items sont instables, comme le prouvent leurs indices de répartition. Pour cerner cette partie du lexique, si connue, mais de fréquence faible et instable, Michéa opposa le concept d'*usuel* (et thématique) à celui de *fréquent* (Michéa, 1950). Les mots disponibles se présentent à nous selon la situation : « un mot disponible est un mot qui, sans être particulièrement fréquent, est cependant toujours prêt à être employé et se présente immédiatement et naturellement à l'esprit au moment où l'on en a besoin » (Michéa, 1953 : 342). Une liste de *centres d'intérêt* (ou C.I., qui représentent des champs lexicaux sur le plan lexicologique) a alors été établie (par linguistes, sociologues et psychologues). Ils sont liés aux champs d'expérience quotidiens des enfants (parties du corps, vêtements, école, jeux, etc.). Ces centres d'intérêt sont utilisés lors

¹ Adrienne Lehrer (1974 : 11) semble ignorer que ce concept revient à Michéa, car elle attribue à tort la méthode de *saliency* (saillance d'un mot dans une situation donnée, c'est-à-dire disponibilité) à Battig et Montague (1968).

² Entre autres, Muller (1992 : 41-42), Picoche (1992 : 49-52), Benveniste et Jeanjean (1987 : 81), etc.

des d'enquêtes de disponibilité pour susciter dans la mémoire le rappel des mots disponibles. Fondée sur l'association d'idées, la disponibilité lexicale a donc l'avantage d'abolir la distinction entre langue orale et écrite. L'indice de disponibilité se calcule séparément pour chaque C.I. en divisant 100 par le nombre total (T) de témoins ayant participé à un C.I., puis en multipliant ce résultat par la fréquence absolue (n) de ce mot dans ce centre d'intérêt, soit: $(100 : T) n$; ou bien: $(n : T) 100$.

Très longues à réaliser (depuis l'enquête de terrain initiale jusqu'au corpus final), les enquêtes de disponibilité sont rares, mais utiles, car elles ont le mérite de retrouver les mots les plus connus, qui échappent aux enquêtes de fréquence. Elles les complètent parfaitement. Enfin, avec les ressources informatiques actuelles, la disponibilité présente un potentiel de développement pour lequel nous avons opté dans nos enquêtes manitobaines.

3. Les enquêtes lexicales

La première enquête a été réalisée en 1990-1991 dans 9 écoles manitobaines, auprès de 344 témoins. Une fois traités, les 131.144 items recueillis constituent un corpus lexicométrique lemmatisé de 19 250 lemmes (Rodriguez, 2006). La seconde enquête a eu lieu en 1993 en trois autres points du Manitoba, avec 48 témoins, et celle de 2004-2006, avec 68 témoins, aux mêmes points qu'en 1993, à fin d'étude diachronique. Pour les trois enquêtes, les 460 témoins³ ont fourni une occurrence totale de plus de 160.000 items⁴. Les détails (nom des écoles, lieu, type d'école, date des enquêtes et nombre de témoins) figurent dans Tab. 1:

<i>nom de l'école</i>	<i>lieu</i>	<i>type</i>	<i>date</i>	<i>total</i>
La Broquerie	r	f	mai 1991	42
Provencher	u	f	février 1991	55
Powerview	r	i	mai 1991	42
Notre-Dame-de-Lourdes	r	f	novembre 1990	48
Saint-Pierre-Jolys	r	f	janvier 1991	39
Taché	u	f	mars 1991	40
Laurier	r	f	juin 1991	24
Sainte-Rose-du-Lac	r	f	juin 1991	39
Sacré-Cœur	u	i	janvier 1991	15
Saint-Laurent	r	mf	juin 1993	19
Saint-Lazare	r	mf	juin 1993	15
Saint-Eustache	r	mf	juin 1993	14
Saint-Laurent	r	mf	décembre 2004	10
Saint-Lazare	r	mf	juin 2005	29
Saint-Eustache	r	mf	décembre 2006	29
effectif total			1990-1991	344
effectif total			1993	48
effectif total			2004-2006	68

Tableau 1 : Enquêtes 1990-2006 : Écoles, dates et effectifs. Lieu: r (zone rurale) ou u (zone urbain).
Type: f (école française), i (école d'immersion), mf (population scolaire métisse francophone).
Total : nombre de témoins par école.

³ Nombre de témoins représentés dans les corpus.

⁴ Le nombre exact d'items sera connu quand tous les corpus des C.I. de 2004-2006 seront établis. Nous présentons ici des exemples inédits des C.I. de 2004-2006 déjà finalisés.

Le déroulement de chaque enquête a été systématique : dans la salle de classe, les participants, sans accès à aucun document, ont inscrit sur des fiches individuelles les mots leur venant à l'esprit sur un thème donné (champ lexical, dit *centre d'intérêt*). Les 16 centres d'intérêt enquêtés font appel au vocabulaire fondamental de la vie quotidienne. Ce sont: 1. les parties du corps; 2. les vêtements; 3. la maison; 4. les meubles de la maison; 5. la nourriture; 6. les objets posés sur la table; 7. la cuisine; 8. l'école; 9. le chauffage et l'électricité; 10. la ville; 11. le village; 12. les moyens de transport; 13. les travaux des champs; 14. les animaux; 15. les jeux et les distractions; 16. les métiers.

Les trois enquêtes ont été conduites de la même façon: en salle de classe, parmi des élèves d'une moyenne d'âge de 11 ans, les filles et garçons étant en nombre similaire, la zone géographique étant urbaine ou rurale, et les écoles françaises pour la plupart (deux étant écoles d'immersion). L'âge des témoins (8 à 13 ans) est crucial dans la recherche de la disponibilité. C'est à cet âge que la *langue fondamentale* des enfants s'affirme et s'ancre dans leur mémoire. Cette langue fondamentale est dépourvue de modes linguistiques et de vocabulaires professionnels ou techniques. Elle n'est pas l'apanage des enfants, mais elle forme leur langue dominante, et elle perdure dans l'âge adulte, bien que moins visible.

Les fiches relevées sur le terrain constituent les données brutes. Tous les mots recueillis ont été saisis, lemmatisés et indexés de balises (étiquettes) attachées à des paramètres linguistiques et sociolinguistiques. Ce processus a abouti à la construction d'un *corpus de données*, puis d'un *corpus lexicométrique* pour chacune des enquêtes.

4. Corpus de données et corpus lexicométriques

Un programme informatique de saisie des données a été conçu, où chaque mot (venant des données recueillies sur le terrain) a été saisi analytiquement par un marquage (balises) représentant des paramètres linguistiques (régionalisme, anglicisme et, par défaut, français transnational ⁵) et des paramètres sociolinguistiques (âge, rural/urbain, fille/garçon, langue(s) parlée(s) à la maison, type d'école fréquentée). Cette saisie résulte en un corpus informatisé, composé des fiches remplies par les témoins (lemmatisées et balisées), et classées par C.I.

Le programme d'analyse lexicométrique, relié informatiquement au précédent (programme de saisie), a permis d'extraire les données du corpus de saisie, pour établir la disponibilité globale des items recueillis. L'indice de disponibilité est ainsi calculé pour chaque item ⁶. Comme il n'existe aucun logiciel évaluant la disponibilité, il fallut créer un outil informatique sur mesure. Pour les deux programmes (de saisie et de calcul lexicométrique), nous avons opté pour RapidFile, conçu par Philippe Borland pour Ashton Tate et pour The Proximity/Merriam-Webster Linguabase, qui permet d'écrire un programme spécifique en DOS. Ce programme fut reconfiguré ensuite pour un système ultérieur compatible avec DOS. Depuis, ce programme s'est avéré performant, même avec un nombre élevé de balises.

En plus de la disponibilité, ainsi rendue calculable même pour une occurrence élevée, le balisage des items rendait possible une description typologique (anglicismes, régionalismes) et sociolinguistique, selon des critères aussi nombreux qu'on le souhaite. Grâce aux balises

⁵ Tronc commun de la langue en usage en France et dans la francophonie. Rodriguez (2006 : 90-91).

⁶ En sémantique lexicale, chaque item représente un mot, employé au sens de lexie (Pottier, voir Ibrahim, 1989), car il s'agit d'unités mémorisées, fonctionnelles et de morphologie variable – mots simples, composés, locutions, etc.

placées à la saisie des données, des indices peuvent être calculés pour un seul paramètre (par exemple les anglicismes), ou pour plusieurs (par exemple, les régionalismes chez les filles en milieu rural, ou les anglicismes chez les garçons des écoles françaises, etc.). Les résultats obtenus pour des paramètres groupés sont nombreux. Cette communication se concentre sur les anglicismes et les résultats obtenus par leur balisage.

Les exemples analysés ci-après proviennent du corpus des 19.250 lemmes de la première enquête, et des deux autres corpus inédits. La disponibilité globale est désormais établie ⁷ pour le premier ensemble de neuf points d'enquête (1990-1991), ainsi que pour chacun de ces points séparément. Le corpus de données et le corpus lexicométrique sont établis pour l'enquête de 1993. Le corpus de données de l'enquête 2004-2006 est terminé, et son corpus lexicométrique en cours de préparation. Nous donnerons des exemples provenant de la partie terminée de ce dernier corpus. Pour les trois corpus, le total des données (l'occurrence) et le nombre des témoins de l'enquête sont très élevés. Les indices lexicométriques obtenus sont donc fiables: statistiquement stables et représentatifs de l'usage.

Nous ferons maintenant trois démonstrations fondées sur le marquage des anglicismes. Dans la première, les indices lexicométriques des anglicismes démarqueront la place globale des anglicismes dans un corpus. Dans la seconde, ils établiront la distinction entre anglicismes de langue et anglicismes de discours (alternance codique). Dans la troisième, mis en corrélation avec les indices des autres types d'items (régionalismes et mots transnationaux), ils révèlent la dynamique existant entre anglicismes et autres composantes du lexique franco-manitobain.

5. Place globale des anglicismes dans un corpus

Tab. 2 représente le nombre de lemmes (items différents obtenus), le nombre et le pourcentage d'anglicismes pour chaque centre d'intérêt (C.I.):

<i>C.I. N°</i>	<i>Total</i>	<i>Lemmes Anglicismes</i>	<i>%</i>
1	392	143	36,47
2	747	314	42,03
3	911	332	36,44
4	939	392	41,74
5	1.080	430	39,81
6	931	312	33,31
7	1.298	539	41,52
8	1.883	619	32,87
9	1.167	452	38,63
10	1.850	586	31,60
11	1.621	462	28,50
12	1.020	448	43,92
13	998	313	31,36
14	895	393	43,52
15	1.793	641	35,75
16	1.725	670	38,84
Total	19.250	7.046	36,60
Moyenne	1.203	440	36,57

Tableau 2 : Anglicismes (corpus 1990-1991)

⁷ La disponibilité lexicale est publiée (Rodriguez, 2006), celle pour chaque point d'enquête est inédite.

Pour le C.I. n°14, les animaux, par exemple, le total des items recueillis (occurrence totale) s'élève à 12.058. Ces items comptent 895 lemmes, dont 393 sont du type anglicisme. Ces anglicismes sont de différents types, dont les deux principaux sont le type lexical (*fox*, pour *renard*; *dragonfly*, pour *libellule*), et le type hybride. Les hybrides sont syntagmatiques (**grizzly ours*⁸, de *grizzly bear*, pour le français *grizzly*), sémantiques (**poisson d'or*, de *goldfish*, en français *poisson rouge*; ou bien **oiseau noir*, de *blackbird*, en français *merle*), ou d'autres types non abordés ici.

L'étude du tableau complet de disponibilité d'un C.I. permet d'éviter l'erreur qui consisterait à penser que les témoins de l'enquête connaissent certains mots en français et des mots autres en anglais. En fait, ce n'est pas le cas. Si plusieurs mots ne sont connus qu'en français, et d'autres qu'en anglais, la majorité des anglicismes de ce C.I. ont des équivalents français qui eux aussi ont été cités. Par exemple, l'anglicisme *squirrel*, avec 4,9% de disponibilité, a son équivalent français *écureuil*, avec 30,9% de disponibilité. Les indices lexicométriques renseignent sur l'état synchronique actuel du lexique, et montrent aussi la concurrence lexicale entre les deux langues – une concurrence qui annonce des changements linguistiques. Dans l'exemple donné, le mot français devance de loin son équivalent anglais, pour le moment.

L'étude d'un tableau complet permet aussi d'établir les proportions d'items français, anglais, et des deux langues. Par exemple, le centre d'intérêt n°1, les parties du corps, de l'enquête de 2004-2006, a recueilli une occurrence totale de 1.614 items, réduite à 188 lemmes, dont 102 sont français, soit 54,3% du corpus, et 86 sont des anglicismes, soit 45,7% de ce champ lexical. Tab. 3 précise le volume et les types d'anglicismes de ce C.I.

C.I. n°1	nombre	%	exemples
1 lemmes	188	100	
2 F (lemmes français)	102	54,3	
3 A (anglicismes)	86	45,7	
4 F sans A	48	25,5	bassin
5 F avec A	54	28,7	cil / <i>eyelash</i>
6 A lexicaux.	75	39,9	
7 A sans F	29	15,4	<i>dimple</i> (fossette)
8 A avec F	57	30,3	<i>thigh</i> / cuisse
9 A hybrides	109	5,3	
10 A hybrides sans F	6	3,1	<i>cellule rouge</i> (globule rouge)
11 A hybrides avec F	4	2,1	<i>gomme</i> / gencive

Tableau 3 : Centre d'intérêt n° 1: les parties du corps (2004-2006). Comparaison des anglicismes et des mots français. Les pourcentages de la quatrième colonne se réfèrent à la totalité des lemmes (188).

Les 86 anglicismes de ce champ lexical sont de types divers. Dans Tab. 3 figurent des exemples des deux types, anglicismes lexicaux et hybrides. La question suivante se pose: combien de mots anglais ont leur équivalent français dans ce C.I., et vice-versa. Sur les 102 mots français du corpus, 54 (28,7% du C.I.) ont un co-occurent synonyme anglais cité par les témoins, soit 52,9% d'entre eux; et 48 (25,5% du C.I.) n'en ont pas, soit 47% d'entre eux.

Voici quelques exemples: *nez* est un exemple de lemme ayant un co-occurent anglais cité dans ce C.I. (*nose*). De même pour *poignet/wrist*, et *coude/elbow*. Par contre, la plupart des 102 lemmes français sont restés sans équivalent anglais cité dans ce C.I.

⁸ L'astérisque indique une forme non attestée.

Le champ lexical n°1 de 2004-2006 contient 86 anglicismes dont 75 sont des lemmes anglais, soit 39,8% du corpus de ce C.I., et 87,2% du corpus des anglicismes de ce C.I. Par contre, seulement 10 sont des hybrides morphologiques ou sémantiques, soit 5,3% du corpus, et 11,6% des anglicismes de ce C.I.

Parmi les 75 anglicismes lexicaux (mots anglais), 57 ont des co-occurents français dans le corpus de ce C.I., soit 30,3% du corpus, et 66,2% du corpus des anglicismes de ce C.I. Il s'agit d'items comme *elbow*, *ankle*, *hip*, *knuckle*, etc. Les 29 items anglais restant, soit 15,4% du C.I., et 33,7% du corpus anglais de ce C.I., n'ont pas de co-occurents français cités. Ce sont des items comme *eyelid*, *gum*, *dimple*, *freckle*, etc.

Quant aux anglicismes hybrides, leur importance lexicométrique peut aussi se calculer. Nous avons dit que sur les 86 anglicismes du corpus, 75 sont des mots anglais et 10 des hybrides. Le tableau complet de disponibilité de ce C.I. permet de vérifier si les hybrides ont un co-occurent français dans le corpus de ce C.I., ou pas. Sur les 10 hybrides, 6 (3,1% du corpus du C.I., et 51,6 % des hybrides) sont sans co-occurent français dans ce corpus (**cellule rouge* pour *globule rouge*). Par contre, 4 hybrides ont un co-occurent français dans ce C.I., soit 2,1% du C.I. et 40% du corpus des hybrides (**petit intestin* pour *intestin grêle*).

Ces proportions sont statistiquement représentatives du redoublement partiel du vocabulaire disponible en milieu bilingue, et sont corroborées par les proportions et statistiques fondées sur notre premier corpus de 130.000 items. Cette utilisation des indices de disponibilité avec leurs balises identifiant les anglicismes permet de cerner de manière objective, quantitative, le volume et le type d'interférences en jeu entre langue majoritaire et langue minoritaire.

6. Langue et discours: l'alternance codique

Le tableau lexicométrique complet d'un C.I. apporte aussi d'autres précisions sur l'emploi des anglicismes dans une langue minoritaire. Il est possible de mettre en relief tous les anglicismes en leur conservant leur rang, sans les regrouper en un sous-ensemble. On peut ainsi repérer la place du premier anglicisme et sa disponibilité. On peut aussi calculer combien d'anglicismes ont une disponibilité forte, moyenne, faible, ou au-dessous du seuil de l'aléatoire. Dans cette perspective, nous avons identifié le nombre d'anglicismes n'ayant été cités que par un seul témoin. Chacun ayant 0,3% de disponibilité, ils ne représentent pas la langue en usage. Par contre, ils représentent une pratique discursive, celle de l'alternance codique, fréquente en milieu bilingue.

Dans cette pratique langagière, qui incorpore les deux langues dans un même échange verbal, le locuteur inclut, dans ses phrases en une langue, des mots ou des membres de phrases en une autre langue. Cette interaction en deux langues a été judicieusement qualifiée d'*alternance codique* par Shana Poplack dans un article de sociolinguistique intitulé « Sometimes I'll start a sentence in English y termino en Español. Toward a Typology of Code-Switching » (Poplack, 1982). Ce n'est pas notre propos ici d'élucider en détail ce qui motive l'alternance codique, mais disons qu'elle montre que ceux qui la pratiquent prennent pour acquis le bilinguisme de leurs interlocuteurs. Elle n'indique pas nécessairement une lacune de la langue, ni une lacune chez les locuteurs. Autrement dit, placés en milieu unilingue, ces mêmes locuteurs pourraient converser en une seule langue.

Quelle est donc la proportion d'alternance codique dans le corpus lexicométrique du C.I. n°1, en 1990-1991 et en 2004-2006 ? Tab. 4 montre la comparaison entre les deux enquêtes. En 1990-1991, les 392 mots du corpus de ce champ lexical, 148 mots sont d'occurrence unique, dont

143 anglicismes. En revanche, seuls 5 mots français sont d'occurrence unique. Parmi les 143 anglicismes, 55 sont des occurrences uniques, soit 38% des anglicismes. Mais la proportion qui nous intéresse est celle des anglicismes à occurrence unique par rapport à l'ensemble des mots du champ lexical. Les 55 anglicismes à occurrence unique composent 14% des 392 mots du champ lexical. En tenant compte des rapports entre langue et discours, il y a de fortes chances pour que ces 14% représentent également le pourcentage d'alternance codique dans beaucoup de conversations qui se tiennent dans le Manitoba bilingue. En 2004-2006, on observe une composition similaire:

<i>C.I.</i> <i>n° 1</i>	1990- 1991	2004- 2006
lemmes (français et anglais)	392	188
lemmes sans occurrences uniques	244	125
lemmes d'occurrence unique	148	63
anglicismes	143	102
anglicismes sans occurrences uniques	88	66
anglicismes d'occurrence unique	55	36
% d'anglicismes d'occurrence unique	38% (de 143)	35,2% (de 102)

Tableau 4 : *C.I. n° 1 : Occurrences uniques et alternance codique*

Que pouvons-nous dire de la « progression » de l'anglais dans le lexique disponible manitobain ? Une autre lecture des indices lexicométriques apporte une réponse à cette question.

7. La dynamique des types lexicaux

Le lexique disponible de chaque C.I. se divise en trois types lexicaux: items français transnational, items régionaux (néologismes canadiens, archaïsmes et dialectalismes de France, amérindianismes et emprunts à diverses langues) et items étrangers (anglicismes). Les proportions entre ces trois types d'items varient plus ou moins d'un C.I. à l'autre, mais sont toujours présentes.

La comparaison des indices lexicométriques des co-occurents (items transnationaux, régionaux et anglicismes) a mis au jour une sorte de « règle des co-occurents » (« synonymes sur plusieurs langues ») par laquelle, dans un même C.I., quand un mot français est accompagné d'un co-occurent français (que ce soit un co-occurent synonyme R, régional, ou un synonyme T, du français transnational), le co-occurent anglais A ne dispose que d'une faible disponibilité, voire est absent de ce C.I. Comme tous les autres C.I. des trois corpus, le n°2 du premier corpus, consacré aux vêtements, en offre des exemples probants.

Le mot *pantalon* (français T) a obtenu une disponibilité de 87,3%, et son synonyme *culotte* (français R), 11,5%, alors que le synonyme anglais, *pants*, n'a obtenu que 0,3%.

De même, le mot *manteau* (français T) a 79% de disponibilité, et son synonyme R, *capot*, 2,7%, alors que l'anglais *coat* n'a obtenu que 0,3%.

Le mot *foulard* (français T) a 39,9% de disponibilité, et son synonyme R, *crémone*, est à 2,7%, tandis que le synonyme anglais *scarf* a obtenu 6,5%. Cette tendance où les mots français résistent mieux aux mots anglais quand ils sont renforcés par un synonyme, nous l'avons analysée pour

les C.I. de l'enquête de 1990-1991 (Rodriguez, 2006). Nous la poursuivons actuellement pour les deux enquêtes ultérieures, sur le plan synchronique et diachronique, car nous souhaitons savoir si cette règle des co-occurents (synonymes) se confirme à 15 ans d'intervalle.

Le marquage des anglicismes et des autres items lexicaux ouvre donc des perspectives d'analyse de la dynamique des langues en contact.

8. Conclusion

Cette analyse du marquage (balisage) des anglicismes dans des corpus de français disponible au Manitoba, a ciblé trois manières d'explorer les indices lexicométriques issus d'enquêtes de terrain. Ces trois utilisations des indices lexicométriques ont mené aux résultats suivants: une évaluation quantitative proprement dite des anglicismes (pourcentages dans chaque C.I., ou champ lexical); une distinction quantitative et qualitative entre anglicismes de langue et anglicismes de discours (alternance codique); et l'analyse du rôle des co-occurents (y compris des synonymes régionaux) dans la lutte contre l'anglicisation d'une langue en situation minoritaire. Ainsi, non seulement les mots disponibles peuvent être cernés par les enquêtes de disponibilité lexicale, mais les raisons de leur survie peuvent aussi être cernées par le biais des indices lexicométriques. Seule la statistique lexicale (notamment l'enquête et les corpus de disponibilité) permet de dresser un profil objectif des anglicismes, et de leur mouvance dans le lexique réellement en usage dans une communauté francophone isolée en milieu anglophone.

References

- Battig W. and Montague W. (1968), *Category Norms for Verbal Items in Fifty-six Categories*. Boulder : The University of Colorado.
- Benveniste C.-B. and Jeanjean C. (1987). *Le français parlé, transcription et édition*. Paris : Didier Erudition.
- Gougenheim G., Rivenc P., Michéa R. and Sauvageot A. (1964). *L'élaboration du français fondamental*. Paris : Didier.
- Ibrahim A. H. (coord.) (1989). *Lexiques*. Paris : Hachette.
- Lehrer A. (1974). *Semantic Fields and Lexical Structures*. Amsterdam-New York: Elsevier.
- Michéa R. (1950). Vocabulaire et culture. *Langues modernes*, fascicule B, mai-juin : 187-192.
- Michéa R. (1953). Mots fréquents et mots disponibles, un nouvel aspect de la statistique lexicale. *Langues Modernes*, volume 47, 4 : 338-344.
- Muller Ch. (1979). *Langue française et linguistique quantitative, Recueil d'articles*. Genève : Slatkine.
- Muller Ch. (1992). *Principes et méthodes de statistique lexicale*. Paris : Hachette.
- Picoche J. (1992). *Précis de lexicologie française: l'étude et l'enseignement du vocabulaire*. Paris : Nathan.
- Poplack Sh. (1982). Sometimes I'll start a sentence in English y termino en Espanol. Toward a typology of code-switching. In Amastae, J. and Elias-Olivares, L., editors, *Spanish in the United States: Sociolinguistics Aspects*, Cambridge : Cambridge University Press, pp. 230-263.
- Rodriguez L. (2006) *La Langue française au Manitoba (Canada): histoire et évolution lexicométrique*. Tübingen : Niemeyer.

