

Evolution lexicométrique des régionalismes dans un corpus franco-canadien (1993–2006)

Liliane Rodriguez

Department of Modern Languages and Literatures
The University of Winnipeg – R3B 2E9 Canada

Résumé

Cet article compare les résultats de nos deux enquêtes de *disponibilité lexicale* (1993 et 2004–2006) auprès de jeunes franco-manitobains. Le *corpus de terrain* de 2004–2006 est de l'ordre de 15 000 items lexicaux. Une fois analysées, ces données de terrain forment un *corpus lexicométrique* de 4 000 lemmes. L'enquête de 1993, d'une occurrence de 11 000 items lexicaux, constitue un corpus lexicométrique de 2 500 lemmes. A la saisie informatique, chaque item a été marqué de plusieurs balises liées au témoin (âge, etc.), au contexte (type d'école, etc.) et au type linguistique (régionalisme, anglicisme, etc.). Nous analysons ici les résultats lexicométriques obtenus avec la balise *régionalisme* (*item lexical régional*) pour divers champs lexicaux (maison, nourriture, objets posés sur la table). Dans une perspective diachronique, les indices de disponibilité des lemmes régionaux de 2004–2006, leur proportion et leur dispersion, sont comparés à ceux de l'enquête jumelle de 1993, conduite dans les mêmes lieux et conditions (les deux corpus sont à ce jour inédits). Cette comparaison montre l'évolution, sur quinze ans, de ces items chargés en valeur identitaire, et caractéristiques de la langue minoritaire (tout officielle soit-elle) qu'est le français au Manitoba.

Mots-clés : lexicométrie, disponibilité lexicale, corpus, régionalisme, langue minoritaire.

Abstract

This paper compares the results of two investigations of lexical availability (1993 and 2004–2006) among young Franco-Manitobans. The 2004–2006 corpus includes about 15000 lexical items. These items, once analyzed, yielded a lexicometric corpus of 4000 lemmas. The 1993 investigation, with 11000 lexical items, yielded a lexicometric corpus of 2500 lemmas. During data entry, each item was tagged according to the test subject (age, etc.), the context (type of school, etc.), and linguistic type (Regionalism, Anglicism, etc.). Here we analyze lexicometric results obtained with the *regionalism* tag (regional lexical item) for a variety of lexical fields (Food, The House, Objects on the Dinner-table). In a diachronic perspective, the availability indices for the regional lemma of 2004–2006, in both their proportions and dispersion, are compared with those of the 1993 twin study, conducted in the same places and conditions (both corpora are unpublished at this date). This comparison traces the change, over fifteen years, in regional lexical items, an enduring component in the French vocabulary of Manitobans – and an identity component as well, in a province where French is a minority language, however official it may be.

Keywords : Lexicometry ; Lexical Availability ; Corpora ; Regional Lexical Item ; Minority Language.

1. Introduction : enquêtes et corpus lexicométriques

L'analyse quantitative qui fait l'objet de cette communication appartient au domaine de la statistique lexicale (Michéa 1950, 1953 ; Muller 1992) et se fonde sur deux enquêtes de *disponibilité lexicale* que nous avons réalisées de 2004 à 2006 auprès d'élèves francophones (souvent bilingues) de 8 à 13 ans, dans trois villages manitobains du Canada. Elle porte sur six des seize champs lexicaux enquêtés (la maison, la nourriture, etc.). En un premier temps, les données relevées sur le terrain ont été saisies en un *corpus de données* d'une occurrence d'environ 15 000 items lexicaux, où chaque item saisi est marqué de plusieurs balises liées au témoin (âge, etc.), au contexte (type d'école, milieu rural, etc.) et au type linguistique (régionalisme, anglicisme, etc.). En un second temps, un *corpus lexicométrique* (inédit) de l'ordre de 4 000 lemmes a été établi en analysant le corpus de données selon les balises souhaitées, et en soumettant ces données à un calcul lexicométrique programmé qui attribue un *indice de disponibilité* à chaque lemme de ce second corpus.

Nous commentons ici les indices lexicométriques obtenus pour la balise *régionalisme* (ou *item lexical régional*), et leur place dans le corpus global pour chaque champ lexical donné. Dans une perspective diachronique, nous comparons les indices lexicométriques de 2004–2006 à ceux de notre enquête jumelle, conduite en 1993 dans les mêmes lieux et conditions. L'enquête de 1993 a abouti à un corpus de données de 11 000 items lexicaux et à un corpus lexicométrique (lui aussi inédit) de 2 500 lemmes. Cette comparaison montre l'évolution quantitative, sur quinze ans, des lemmes régionaux selon leur disponibilité, leur proportion dans le corpus et leur dispersion. Ces items lexicaux régionaux (chargés en valeur identitaire), avec leur déclin ou leur développement, sont caractéristiques de la langue minoritaire (tout officielle soit-elle) qu'est le français au Manitoba.

2. Contexte historique et linguistique manitobain

La situation minoritaire de la langue française au Manitoba et ses caractéristiques se comprennent mieux avec un aperçu historique. Le Manitoba est l'une des provinces canadiennes où le français et l'anglais sont langues officielles, comme elles le sont aussi au niveau fédéral. Toutefois, au niveau provincial, la situation linguistique du Manitoba résulte de rebondissements constitutionnels, parfois violents, antérieurs et postérieurs à l'entrée de la province dans la Confédération en 1870. Nous les résumerons brièvement.

La langue française fut introduite dans l'Ouest au XVIII^e siècle par l'explorateur Pierre Gaultier de La Vérendrye, mandaté par le Gouverneur de Nouvelle-France, et avec les «coureurs de bois», venus à titre individuel de Nouvelle-France, puis sous le nom de «Voyageurs», engagés par des compagnies de traite des fourrures : Compagnie du Nord-Ouest, française, et Compagnie de la Baie d'Hudson, anglaise. Cette dernière finit par absorber la première, mais grâce à ses employés en majorité francophones, « le français resta la langue "officielle" [dans Nord et l'Ouest de l'Amérique du Nord] aussi longtemps que la traite des fourrures y fut prospère. » (Nute 1955: 5)

Une fois terminés leurs contrats avec ces compagnies, beaucoup de Voyageurs restèrent dans l'Ouest, avec leurs épouses amérindiennes et leurs enfants métis. Le centre administratif de la Compagnie de la Baie d'Hudson (CBH ci-après) était établi au carrefour des axes est-ouest et

nord-sud de la traite, dans ce qui allait devenir le Manitoba. Au milieu du XIX^e siècle, des colons écossais, puis anglais, continuèrent à peupler cette région qui devint Colonie de la Rivière-Rouge. Quand Ottawa voulut distribuer des terrains aux colons anglophones, les francophones et Écossais déjà installés sur ces terres données par la CBH à l'Église catholique, s'insurgèrent. Après une rébellion menée par Louis Riel, un métis français, amérindien et irlandais, la région entra dans la Confédération sous le nom de Manitoba. Louis Riel, auteur principal de la constitution provinciale, l'*Acte du Manitoba* (1870), fut exécuté en 1885 par le gouvernement fédéral pour avoir dirigé une rébellion similaire dans la province voisine (Saskatchewan).

L'*Acte du Manitoba*, par lequel la Confédération canadienne octroya un statut constitutionnel bilingue (anglais et français) à la province du Manitoba. Pourtant, le gouvernement provincial abolit l'emploi du français au Parlement en 1890, et des écoles en 1915. Votées seulement par le gouvernement provincial, sans accord fédéral, ces lois antifrancophones s'avérèrent inconstitutionnelles (voir Bray 1987). Entre 1970 et 1980, les droits linguistiques initiaux furent rétablis dans le sillage de la *Loi sur le bilinguisme officiel* (loi fédérale) de 1970. Cela signifie que pendant près d'un siècle les Franco-Manitobains furent privés de leurs droits linguistiques, mais qu'ils eurent la ténacité de les faire rétablir par la Cour Suprême du Canada. Aujourd'hui, le Manitoba compte 55 000 personnes se déclarant de langue maternelle française, soit au moins 5% de sa population. A ce chiffre s'ajoutent les familles bilingues dont les pratiques linguistiques sont plus difficiles à recenser, et 10% d'allophones (environ 100 000 personnes).

Cette situation linguistique ouvre deux champs de réflexion. En premier lieu, les contacts quotidiens entre deux langues, parlées sur un même territoire, sont sources d'interférences linguistiques, donc d'anglicismes pour les locuteurs minoritaires francophones. Nous avons traité plusieurs aspects de cette question (voir Rodriguez 2005). En second lieu, une langue aussi anciennement établie que l'est le français au Manitoba, mais aussi éloignée de ses sources (la France à 8 000 km, le Québec à 3 000 km), comporte des caractéristiques régionales. Pour les cerner quantitativement et qualitativement (ce qui n'avait pas été fait), nous avons réalisé trois enquêtes de terrain au cours des vingt dernières années : en 1990–1991 (en douze points de la province), en 1993 (en trois points), et en 2004–2006 (aux mêmes points qu'en 1993). Les détails de ces enquêtes et la description du programme informatique spécifiquement créé pour les analyser, ont été présentés aux *Jadt* 2010 (Rodriguez 2010).

3. Les enquêtes manitobaines

Rappelons que notre première enquête (1990–1991) et le corpus qui en a découlé concernent 9 points du Manitoba et 344 informants. Les 131 144 items recueillis constituent un corpus lexicométrique de 19 250 lemmes (corpus et analyse publiés dans Rodriguez 2006). Notre deuxième enquête, menée en juin 1993, comptait 48 informants. Nous nous sommes alors intéressée à des points du Manitoba ne figurant pas parmi ceux de 1990. Bien que très éloignés les uns des autres, les trois villages choisis se signalaient par leurs traditions linguistiques et culturelles françaises et métisses bien préservées. Il s'agit de Saint-Eustache, Saint-Lazare et Saint-Laurent. Ce dernier village fut, par la suite, élu par le *Smithonian Institute* comme étant le village métis canadien ayant le mieux conservé ses traditions, et bénéficie d'une exposition permanente à Washington (voir *The Métis Nation* 2006). Dans un dessein comparatif, nous avons choisi, pour la troisième enquête, ces trois mêmes villages : Saint-Laurent (décembre 2004),

Saint-Lazare (juin 2005) et Saint-Eustache (décembre 2006). Les données des trois enquêtes ont été lemmatisées de la même façon, et associées aux mêmes balises : la balise *régionalisme* et toutes celles liées au lieu, à l'informant et au contexte ont aussi été intégrées à la saisie de chaque corpus. Combinées, ces diverses balises permettent des analyses différenciées. Des résultats, obtenus de 1993 à 2006 pour la balise *régionalisme*, font l'objet du présent article.

4. Concept de *disponibilité lexicale*

Pour étudier la variété géolinguistique du français manitobain, nous avons opté pour la méthode de *disponibilité lexicale* qui donne une mesure précise de la disponibilité lexicale réelle des témoins. Le concept de *disponibilité*, établi par René Michéa (Gougenheim *et al*, 1964), est le moyen le plus efficace d'identifier le vocabulaire le plus employé par les jeunes. L'enquête a lieu en salle de classe et recueille les réponses spontanées des répondants sur des thèmes essentiels à leurs intérêts immédiats et à leur développement cognitif (thèmes dits *centres d'intérêts* ou champs lexicaux). Comme les productions des répondants sont issues d'associations d'idées, la notion de disponibilité, relevant de la psycholinguistique, abolit la distinction entre langue parlée et écrite (Picoche 1993: 59-63 ; Gougenheim 1964: 69-113). Les données recueillies (par écrit) auprès des informants sont à la fois des attestations écrites (textes minimaux) et des attestations orales (conversations minimales). Les travaux inauguraux de Michéa (1950 et 1953) et Gougenheim *et al* (1964) visaient à cerner le *français fondamental*, c'est-à-dire celui qui est le plus utile à enseigner en langue seconde. Pour nous, il s'agissait de décrire une langue minoritaire.

Rappelons qu'« 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). Il se distingue donc des *items fréquents*, issus des analyses de banques de données textuelles. Alors que les items les plus fréquents d'un texte sont des connecteurs, des actualisateurs, des adverbes grammaticaux, etc., les *items disponibles* sont thématiques, et ont le trait sémantique *concret*. La disponibilité s'oppose donc à la fréquence, et la complète (Michéa 1950). L'indice de disponibilité d'un item se calcule, pour chaque champ lexical, en divisant la fréquence absolue (n) de cet item par le nombre total (T) d'informants, puis en multipliant ce résultat par 100 ($n : T \times 100$). Cet indice représente donc la proportion en pourcentage des informants ayant cité un item lexical, dans ce champ lexical.

Partie de cette méthode et de ce concept fondamentaux, nous avons choisi seize champs lexicaux et introduit des balises pertinentes à une langue en milieu minoritaire (et non seulement majoritaire, comme c'était le cas pour l'enquête de Michéa, qui portait uniquement sur le français en France). L'une de ces balises, celle qui signale un *item lexical régional*, permet de retrouver dans le vaste champ lexicométrique de l'ensemble des items, ceux qui caractérisent la variété géolinguistique du français manitobain. Dans le corpus lexicométrique chaque lemme régional a donc un indice de disponibilité, comme tout autre lemme du corpus. Les indices présentés ici proviennent de six des champs lexicaux des deux corpus (1993 et de 2004–2006).

5. Définition du terme *item lexical régional*

La *régionalité* se définit généralement par rapport à un centre, soit un centre géographique, soit un centre d'influence économique, financier, culturel, social, etc. Ainsi le français de Marseille

est-il considéré comme régional par rapport à celui de la capitale nationale, Paris. Celui de Chicoutimi (au Québec) peut être considéré comme régional par rapport à celui de la ville de Montréal. Celui du Manitoba est régional par rapport à celui de Paris et du Québec, celui de Saint-Laurent (Manitoba) est régional par rapport à celui de Saint-Boniface, etc. Mais plutôt que de rechercher un « centre » d'influence géographique ou autre, nous préférons décrire le lexique manitobain comme une variété de français dont nous avons recueilli des attestations, et élargir la notion de « centre » à l'ensemble de la francophonie. En effet, l'ensemble des variétés de français dispose d'un tronc commun, lexical, morphosyntaxique, phonologique, etc., que l'on retrouve dans toute la francophonie. C'est ce que nous appelons le *français transnational* (Rodriguez 2006 : 91), véritable « centre » linguistique autour duquel se définissent les variétés géolinguistiques. Les éléments dits « régionaux » (topolectaux) marquent une langue géographiquement, l'attachent à un territoire, la rendent reconnaissable à ses locuteurs, voire à d'autres locuteurs. S'il est vrai que certains des régionalismes attestés au Manitoba se retrouvent ailleurs au Canada, ou dans la francophonie, ils n'en demeurent pas moins des régionalismes (topolectes) dans leurs variations de fréquence, de disponibilité ou de forme, distinctes de celles d'autres régions francophones, et dans leur rapport variable au « centre » qu'est le français transnational.

La plus grande partie du corpus manitobain se compose d'items transnationaux. En plus de ce tronc commun, les deux autres types lexicaux sont les anglicismes et les régionalismes. A la saisie, ces distinctions ont été opérées manuellement, en appui sur nos connaissances en histoire des deux langues et de leurs dialectes, sur des dictionnaires historiques spécialisés (Huguet 1925, Godefroy 1982, etc.), et sur les *Atlas linguistiques et ethnographiques* du domaine d'oïl (par exemple, Horiot et Massignon 1971–1983) et l'*Atlas linguistique de France* (Gilliéron et Edmont 1902–1910). Du fait de leur origine dialectale variable, les régionalismes se subdivisent en archaïsmes, résidus d'une ancienne norme française (*maganer*, pour abîmer), en dialectalismes (*cru* est picard, *moulée*, pour farine animale, est normand), en archaïsmes-dialectalismes (*grafigner*, pour griffer), ou en amérindianismes (*pembina*, pour viorne trilobée, voir Dulong 1999), ou encore en canadianismes (*coton-ouaté*, pour sweat-shirt, *plume feutre*, pour stylo feutre, etc.). Certains emprunts anciens à l'anglais, morphologiquement intégrés au français (*cayousse*, de *horse*, pour cheval), sont aussi des régionalismes. Mais, nous ne nous attarderons pas sur cet aspect historique des régionalismes, qui sort de la visée essentiellement statistique de cet article.

Quant aux anglicismes, ils ne font pas partie des régionalismes. Ils vivent en marge du français, et se mêlent (parfois ou souvent) à la langue des francophones minoritaires, en situation de dominance lexicale, ou simplement d'alternance codique comme nous l'avons démontré (Rodriguez 2010). Contrairement à la France, où l'interférence est lexicale, au Canada les anglicismes sémantiques prédominent, mais ils sont statistiquement aléatoires (par exemple, « un ordinateur **amical* », pour *facile d'emploi*, sous l'influence du sens commercial de *friendly*). L'identification des anglicismes ne peut donc se faire que par l'observation des *deux* langues, et non d'après des listes préétablies. Ces dernières sont souvent fautives (confondant, par exemple, archaïsme et anglicisme), et incomplètes par nature puisque le phénomène de l'interférence fluctue, dû à un emploi plus idiolectal que collectif, à des comportements pragmatiques d'alternance codique, et au fait qu'il touche à toutes les composantes de la langue.

Au Manitoba, la régionalité (comme l'interférence) apparaît dans toutes les strates de la langue. En voici des exemples, extraits de nos recherches manitobaines. Sur le plan phonétique, le [r] roulé caractérise deux tiers des locuteurs. Sur le plan phonologique, l'opposition entre la voyelle antérieure de *patte* et la voyelle postérieure de *pâte*, est encore très marquée. Sur le plan syntaxique, l'inversion verbe-sujet est préférée à l'interrogation avec *est-ce que*. Sur le plan lexical, un mot comme *butin* a encore le sens de vêtement, et *cru* le sens de froid et humide, en parlant du temps, etc.

Après avoir défini le concept de *régionalité*, nous l'avons utilisé comme balise pour notre corpus de données, où chaque item régional a été analysé du point de vue de son histoire linguistique avant d'être marqué comme régionalisme. Cette analyse en amont du marquage est d'ordre historique et lexicographique (voir Rodriguez 1984).

6. Evolution quantitative des lemmes régionaux au Manitoba

Lemmes et informants

Nous avons commencé par calculer l'occurrence totale des items lexicaux recueillis et le nombre d'informants pour chacun des champs lexicaux (C.I., centres d'intérêt). Le tableau 1 rend compte de ces chiffres pour six des seize champs lexicaux enquêtés : 1. les parties du corps ; 2. les vêtements ; 3. la maison ; 5. la nourriture ; 6. les objets posés sur la table ; 7. la cuisine (les numéros des C.I. sont ceux utilisés dans le corpus).

La lecture du tableau 1 permet de constater que le nombre d'informants est passé de 33 à 62,5 en 15 ans. Comme il s'agit de la population scolaire de villages isolés (classes de 5^e et de 6^e), cette augmentation des effectifs des écoles représente un résultat positif pour ces petites communautés francophones. Le nombre d'items lexicaux recueillis à l'enquête (occurrence totale) a augmenté d'environ 50%, ce qui montre que les jeunes de 2004–2006 n'ont pas perdu de leur vocabulaire français.

C.I. numéro	1993 nombre items	1993 nombre informants	2004–06 nombre items	2004–06 nombre informants
1	769	33	1 614	65
2	644	33	1 467	64
3	607	33	1 221	63
5	1 291	33	2 077	64
6	647	33	1 012	64
7	873	33	890	54
total	4 831	----	8 281	---
moyenne	805	33	1 380	62,5

Tableau 1. Nombre d'items lexicaux (occurrence totale) et d'informants (1993 et 2004–2006).

Mais qu'en est-il du vocabulaire régional ? Combien d'informants ont-ils produit des régionalismes ? Le tableau 2 répond à ces questions et apporte des éléments de réflexion sur l'évolution des items régionaux. La lecture du tableau 2 montre que, contrairement aux anglicismes (Rodriguez 2010), les lemmes régionaux résistent au changement linguistique. En effet, la moyenne d'informants ayant donné un régionalisme reste comparable à 15 ans d'écart (57% en 1993, et 56,6% en 2004–2006), malgré la variabilité selon les champs lexicaux (le C.I. n°1 n'étant pas représentatif). Pour l'ensemble de la production lexicale, tous C.I. confondus, cette proportion atteint 100%, de 1993 à 2006. La moyenne des items régionaux cités par les informants est donc aussi restée identique (1,6 item). Une étude qualitative ultérieure nous dira si ces régionalismes sont les mêmes dans les deux corpus. Nous savons déjà que certains le sont, mais pas tous.

C.I.	1993 total informants	1993 nombre donnant région. %*	1993 nombre région. par inform.	2004–06 total informants	2004–06 nombre donnant région. %*	2004–06 nombre région. par inform.
1	33	1 3%	1	65	32 49,2%	1
2	33	32 96,9%	3	64	62 96,8%	3
3	33	15 45,4%	1	63	19 30,1%	1
5	33	26 78,7%	2	64	52 81,2%	3
6	33	14 42,4%	1	64	21 32,8%	1
7	33	25 75,7%	2	54	27 50%	1
moyenne	33	19 57%	1,6	62,5	35,5 56,6%	1,6

*proportion de répondants ayant produit un régionalisme

Tableau 2. Nombre d'informants ayant donné des régionalismes et moyenne de régionalismes par informant (1993 et 2004–2006).

Nous commenterons maintenant brièvement quelques résultats avec balises combinées, celles du type lexical (régionalisme) et du sexe. Les résultats montrent que les filles des deux corpus citent en moyenne moins de régionalismes que les garçons : 1,5 contre 2. Mais, les deux groupes de répondants citent des mots anciens (*blé d'Inde* pour maïs) comme des néologismes canadiens (*croustille* pour chips ; *craquelin* pour craker, *hambourgeois* pour hamburger). Nous remarquons que des mots français ont été introduits au Canada pour remplacer ces trois mots

anglais. Ces observations semblent indiquer un maintien, voire un renouveau, du vocabulaire régional franco-manitobain.

Place des lemmes régionaux dans les corpus

Indépendamment des informants, nous recadrerons maintenant notre attention sur les lemmes régionaux, sur leur place dans les corpus. Pour le corpus de 1993 et celui de 2004–2006, nous avons calculé le nombre total de lemmes et celui des lemmes régionaux, puis la proportion des lemmes régionaux dans chaque champ lexical (C.I.).

Le tableau 3 présente le total des lemmes, le nombre et le pourcentage de lemmes régionaux dans les six mêmes champs lexicaux (C.I.) que dans le tableau 1. Nous avons constaté qu'en 1993 le C.I. n°3, la maison, par exemple, comptait un total d'items recueillis (occurrence totale) de 607 (voir tableau 1). Ces 607 items donnent un total de 198 lemmes, dont 10 sont du type *régionalisme*. Ces lemmes régionaux composaient donc 5% du corpus de ce champ lexical. En 2004–2006, ils composent 2,3% du corpus de ce même champ (voir tableau 3).

Une étude complémentaire, additionnant les divers indices de disponibilité de chaque item régional apparaissant dans plusieurs des seize champs lexicaux (C.I.), permettra de mesurer la *vitalité* des items lexicaux (la *vitalité* se distinguant de la *disponibilité*), c'est-à-dire la place de ces régionalismes dans l'usage.

C.I. numéro	1993 total lemmes	1993 total lemmes région.	1993 % lemmes région.	2004– 06 total lemmes	2004– 06 total lemmes région.	2004– 06 % lemmes région.
1	112	1	0,9	188	3	1,6
2	160	12	7,5	324	14	4,3
3	198	10	5,0	348	8	2,3
5	368	14	3,8	473	25	5,2
6	231	11	4,7	293	11	3,7
7	290	17	5,8	276	15	5,4
total	1 359	65	4,7	1902	76	3,9
moyenne	226,5	10,8	4,7	317	12,6	3,9

Tableau 3. Pourcentages des lemmes régionaux (1993 et 2004–2006).

L'étude qualitative des items régionaux permet de comprendre ce qui a causé leur diminution ou leur maintien. Pour le C.I. n°3, les 10 régionalismes de 1993 sont : *armoire* (pour placard de cuisine), *chambre de bain* (salle de bain), *bardeau* (tuile en goudron), *gravois* (gravier), *roche* (caillou), *salle de lavage* (buanderie), *chambre à plantes* (pièce à plantes), *poêle* (cuisinière), et *salle à dîner* (salle à manger). Les 8 régionalismes de 2004–2006 sont : *roche*, *chambre de bain*, *armoire*, *chambre d'ordinateur*, *chambre à jouer*, *chambre à lire*, *chambre de travail* et *lévier* (forme agglutinée de l'évier). Les régionalismes liés à la construction de la maison et

aux travaux domestiques sont en diminution, car les enfants y participent moins qu'autrefois, mais l'item *chambre* (au sens de pièce dans une maison) s'est renforcé au lieu de perdre de sa disponibilité. Cette évolution signale, entre autre, un vécu dans une maison plus vaste qu'en 1993, et des activités plus souvent à l'intérieur.

Par contre, pour le C.I. n°7, la cuisine (il s'agit de la pièce, non des mets), le pourcentage des lemmes régionaux est resté semblable : 5,8% en 1993 et 5,4% en 2004–2006. Sept d'entre eux sont les mêmes (*armoire*, *bassin* (évier), etc.). D'autres sont différents : par exemple, *laveuse* (machine à laver), *tasse en verre* (en verre) en 1993 ; *canard* (bouilloire), *poêlonne* (poêlon ou caquelon) en 2004–2006 (voir tableau 3).

Le C.I. n°5 (la nourriture) voit une augmentation des régionalismes (de 3,8% à 5,2%), justifiée par un vocabulaire scolaire beaucoup plus nombreux aujourd'hui où les écoles sont mieux équipées, et plus ouvertement tournées vers la culture francophone (voir tableau 3). Nous constatons donc, sur le plan qualitatif, le maintien de certains éléments régionaux et la disparition de certains autres.

Les indices lexicométriques des lemmes régionaux permettent donc d'affirmer que, d'une part, le français régional ne recule pas au Manitoba et, d'autre part, qu'il se renforce ou se renouvelle dans une certaine mesure. Mais ces items régionaux représentent-ils un vocabulaire commun restreint, un sous-ensemble limité et figé à l'intérieur de la totalité du vocabulaire disponible ? Pour répondre complètement à cette question de la place des régionalismes dans l'usage, nous devons mesurer la *vitalité* des items lexicaux (solution mentionnée ci-dessus, mais qui sort du cadre de cet article). Nous pouvons aussi mesurer leur *dispersion*, ce que nous nous attacherons à faire maintenant.

7. La dispersion

La *dispersion* est le rapport entre le nombre total d'occurrences et le nombre de lemmes. Elle est « le rapport du nombre de mots différents de chaque liste au nombre total de mots » (Dugast 1980 : 43). Cette notion appartient à la *statistique statique* (Dugast 1980 : 32-55), qui analyse les moyennes, les indices (pourcentages) de dispersion d'un item, et la variance (indice de dispersion autour de la moyenne). La dispersion d se calcule en divisant le nombre de lemmes l par le nombre d'occurrences L (tous les items des données) et en multipliant ce résultat par 100, soit $d = (l : L) 100$. Par exemple, pour le C.I. n°3 de 1993, la dispersion de la totalité des lemmes, dT , est : $(198 : 607) 100 = 32,6\%$. Pour répondre à la question de l'usage posée plus haut, on peut mesurer la dispersion dT pour l'ensemble des lemmes, et la comparer à dR , celle des lemmes régionaux.

C.I. numéro	1993 dispersion totalité des lemmes		1993 dispersion lemmes régionaux		2004–06 dispersion totalité des lemmes		2004–06 dispersion lemmes régionaux	
	<i>l/L</i>	<i>dT</i>	<i>l/L</i>	<i>dR</i>	<i>l/L</i>	<i>dT</i>	<i>l/L</i>	<i>dR</i>
1	112/769	14,5	1/1	100	188/1614	11,6	3/33	9,0
2	160/644	24,8	12/110	10,9	324/1467	22,1	14/198	7,1
3	198/607	32,6	10/21	47,6	348/1221	28,5	8/24	33,3
5	368/1291	28,5	14/43	32,5	473/2077	22,7	25/132	18,9
6	231/647	35,7	11/20	55,0	293/1012	28,9	11/25	44,0
7	290/873	33,2	17/52	32,6	276/890	31,0	15/39	38,4
d moyenne 1-7*	28,2		46,4		24,1		25,1	
d moyenne 2-7	30,9		35,7		26,6		28,3	

Tableau 4. Dispersion des lemmes (1993 et 2004–2006).

La moyenne *d* 1-7* du tableau 4 est moins stable que la moyenne *d* 2-7 car le C.I. n°1 est statistiquement aléatoire (un seul lemme). Nous retiendrons donc la moyenne *d* 2-7 pour les remarques suivantes, car ce tableau appelle plusieurs commentaires.

Pour chacun des deux corpus (1993 et 2004–2006), la moyenne *dR* (35,7 et 28,3) est supérieure à la moyenne *dT* (30,9 et 26,6). Cela signifie que le vocabulaire disponible régional inclut plus d'items moins cités que les items transnationaux par les répondants, ce que confirment les indices de disponibilité plus faibles de la plupart des lemmes régionaux. Très rares sont les items régionaux de haute disponibilité, comme, pour le C.I. n°6 (les objets posés sur la table), *chandelle* (bougie) à 24,2% en 1993, et à 18,8% en 2006. La plupart sont de disponibilité moyenne ou faible : *chaudron* (marmite), *couverte de table* (nappe), *poêlonne* (poêlon), etc., sont à 3% en 1993 ; et *blé d'Inde* (maïs), *courriel* (mél), *grosse assiette* (plat à servir), etc., sont 1,6% en 2006. Ce profil lexicométrique se répète pour tous les C.I., même à 15 ans d'écart.

Nous constatons aussi que la dispersion (*dT* comme *dR*) a diminué dans les deux synchronies : 30,9 et 35,7 en 1993 ; et 26,6 et 28,3 en 2004–2006. Cette évolution des lemmes régionaux, comme de l'ensemble des lemmes, signifie qu'aujourd'hui les enfants ont un plus grand vocabulaire disponible en commun. Cela s'explique en partie par les conditions de vie : moins d'isolement dans les fermes, plus de jeux et de sports organisés (hockey, base-ball, etc.), plus d'activités et de voyages en ville, grâce à des transports améliorés.

A titre comparatif, la dispersion des anglicismes a aussi baissé. Pour les six mêmes champs lexicaux étudiés ici, elle est passée de 45,1% en 1993 à 39,5% en 2004–2006. Cependant elle est nettement supérieure à celle des autres types de lemmes (régionaux ou transnationaux), car les lemmes anglais sont toujours de très faible disponibilité, et ne constituent pas une partie statistiquement stable du vocabulaire que les enfants ont en commun. Ce sont majoritairement des cas d'alternance codique, et non des anglicismes installés dans la langue.

Le fait que les lemmes régionaux évoluent au même rythme que l'ensemble des lemmes (la dispersion a baissé de manière similaire) a une signification linguistique supplémentaire : les régionalismes font partie à part entière de l'ensemble du vocabulaire disponible. C'est signe que le vocabulaire disponible régional franco-manitobain n'est ni clos ni figé, et pas seulement composé d'archaïsmes, disparaissant peu à peu, au fil des ans. Au contraire, il se développe en phase avec le vocabulaire transnational. Cela est confirmé par l'analyse qualitative, géolinguistique et dialectale : ce vocabulaire n'est pas constitué que de formes en survie (archaïques ou dialectales), mais aussi d'emprunts à diverses langues, de néologismes canadiens, et même de dérivés sur base d'archaïsmes (ce dernier trait semble particulier aux langues minoritaires).

8. Conclusion

Le lexique disponible franco-manitobain contient trois grands sous-ensembles : les lemmes de français transnational, les lemmes régionaux (néologismes canadiens, archaïsmes et dialectalismes de France, amérindianismes et emprunts à diverses langues) et les lemmes étrangers (anglicismes). Dans les deux corpus étudiés ci-dessus, les proportions entre ces trois parties varient plus ou moins d'un C.I. à l'autre, mais restent présentes. Cette étude des lemmes régionaux montre la proportion du vocabulaire régional dans l'ensemble du corpus manitobain, et démontre que les indices lexicométriques, les proportions et la dispersion du vocabulaire disponible évoluent en phase avec la composante transnationale dominante. Le vocabulaire français des jeunes manitobains conserve donc une certaine régionalité, tout en évoluant et en intégrant des néologismes internationaux, canadiens, voire locaux. Malgré les déboires du passé, et sa situation minoritaire, la langue française du Manitoba poursuit son chemin.

Bibliographie

- Bray J. (1987). *L'article 23*. Saint-Boniface (Manitoba): Les Editions du blé.
- Dugast D. (1980). *La statistique lexicale*. Genève: Slatkine.
- Dulong G. (1999). *Dictionnaire des canadianismes*. Sillery (Québec): Septentrion.
- Gilliéron J. et Edmont E. (1902–1910). *Atlas linguistique de France*. Paris: Champion.
- Godefroy F. (1880–1902). *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XV^e siècle*. Paris: Éditions Picard. 10 volumes.
- Gougenheim G., Rivenc P., Michéa R. et Sauvageot A. (1964). *L'élaboration du français fondamental*. Paris: Didier.
- Horiot B. et Massignon, G. (1971–1983). *Atlas linguistique et ethnographique du Poitou, de l'Aunis, de la Saintonge et d'Angoumois Ouest (ALO)*. Editions du CNRS. (1971) Volume 1; (1974) Volume 2; (1983) Volume 3.
- Huguet E. (1925). *Dictionnaire de la langue française du XVI^e siècle*. Paris: Didier.
- Michéa R. (1950). « Vocabulaire et culture », in *Langues modernes*, fascicule B, mai-juin, 187-192.
- Michéa R. (1953). « Mots fréquents et mots disponibles, un nouvel aspect de la statistique lexicale », in *Langues Modernes*, volume 47, n°4, 338-344.
- Muller Ch. (1992). *Principes et méthodes de statistique lexicale*. Paris: Hachette.
- Nute G.L. (1987). *The Voyageur*, Reprint edition (1955). St Paul: Minnesota Historical Society.
- Picoche J. (1992). *Précis de lexicologie française: l'étude et l'enseignement du vocabulaire*. Paris, Nathan.

- Rodriguez L. (1984). *Mots d'hier, mots d'aujourd'hui*. Saint-Boniface: Editions des Plaines.
- Rodriguez L. (2005). « Interférences phonétiques et variation topolectale, in *Cahiers franco-Canadiens de l'Ouest*, volume 16, numéros 1 et 2, 99-109.
- Rodriguez L. (2006). *La Langue française au Manitoba (Canada): histoire et évolution lexicométrique*. Tübingen: Niemeyer.
- Rodriguez L. (2010). « Le rôle du marquage lexicométrique des anglicismes dans un corpus franco-canadien ». *Jadt2010, Statistical Analysis of Textual Data*, Proceedings of the 10th International Conference Journées d'Analyse statistique des Données Textuelles, 9–11 June 2010 – Sapienza University of Rome, Bolasco S., Chiara I. et Giuliano L. (editors), 1295-1303.
- The Métis Nation* (2006). « Smithsonian National Museum Features Métis Display », Issue 2, Volume 3, November, 29.