

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MESURE DE LA TAILLE DU VOCABULAIRE ORTHOGRAPHIQUE ET
PHONOLOGIQUE CHEZ DES ÉLÈVES DE FRANÇAIS LANGUE SECONDE DE
SECONDAIRE 1 À 5

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAITRISE EN DIDACTIQUE DES LANGUES

PAR
JENNIFER VACHON

Août 2017

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.10-2015). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Plusieurs personnes ont contribué à ce que ce mémoire soit rédigé avec succès et je souhaiterais les en remercier chaleureusement. C'est un grand accomplissement pour moi, mais il y a aussi un peu de chacun d'eux dans cette réalisation.

Tout d'abord, merci à ma directrice Philippa Bell qui m'a supportée tout au long de cette périlleuse rédaction et ses précieux conseils m'ont toujours ramenée à l'essentiel. Une mention aussi à mes lectrices, Véronique Fortier et Ophélie Tremblay, qui ont gracieusement accepté de lire et commenter ce mémoire. Soulignons l'essentielle contribution de Jill Vandermeerschen, statisticienne au SCAD, qui était toujours disponible pour offrir son aide lors des analyses statistiques.

Ensuite, je veux remercier l'école qui m'a accueillie, les enseignantes qui ont accepté de collaborer avec moi, ainsi que les participants à cette étude. Sans vous, la cueillette de données n'aurait pu être aussi facile.

Je souhaite aussi dire merci à Édith, ma mère, qui m'a poussée à aller jusqu'au bout, et sans qui, bien honnêtement, je n'y serais jamais arrivé. Tu me permets d'être à mon meilleur à chaque instant, et ton aide fut plus que précieuse. Tu y croyais plus que moi et tu vois, j'y suis arrivé !

Finalement, je souhaite remercier de tout cœur Rafael, mon complice de tous les jours, et Liana qui a vu le jour alors que j'étais en pleine rédaction de ce mémoire. Vous êtes ce qui me permet d'avancer, tout simplement, et c'est avec amour et fierté que je vous dédie ce mémoire.

AVANT-PROPOS

Ce mémoire a commencé à prendre forme dès le début de ma carrière d'enseignante du français langue seconde, il y a 10 ans, alors que je me demandais quel serait le vocabulaire le plus utile à enseigner. Je remarquais que non seulement mes élèves avaient de la difficulté à comprendre des termes que je pensais simples, mais aussi qu'une conversation simple pouvait parfois les embrouiller. Le vocabulaire de base semblait occasionnellement mal maîtrisé, malgré le fait que les élèves utilisaient d'autres termes, quelquefois même rares. De plus, les écarts présents en salle de classe étaient réels et les besoins des apprenants bien différents. Je me demandais, d'une part, par où je devais commencer dans mon enseignement du vocabulaire et, d'autre part, si les textes que j'utilisais leurs étaient accessibles.

À l'hiver 2014, le cours d'acquisition et d'enseignement du vocabulaire, offert au programme de maîtrise en didactique des langues de l'UQAM, m'a montré qu'il était possible de classer les mots selon leur fréquence et que ceux très fréquents formaient une base essentielle à l'apprentissage de la L2. C'est alors que je me suis demandé si mes élèves possédaient ces mots fréquents et comment il était possible de savoir s'ils les maîtrisaient, ou même de pouvoir mesurer la quantité de termes qu'ils savaient.

Comme chercheuse, mesurer la taille du vocabulaire m'apparait comme un outil permettant d'éclairer les enseignants sur ce qu'ils connaissent de leurs apprenants, les aidant ainsi à mieux adapter les enseignements lexicaux afin de rejoindre davantage leurs besoins. C'est dans cet esprit que ce mémoire a été créé.

TABLE DES MATIÈRES

REMERCIEMENTS.....	iii
AVANT-PROPOS.....	iv
LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX.....	viii
LISTE DES ABRÉVIATION, SIGLES ET ACRONYMES.....	ix
RÉSUMÉ.....	x
INTRODUCTION.....	1
CHAPITRE 1 PROBLÉMATIQUE.....	3
1.1 Le rôle des connaissances lexicales dans l'acquisition des langues secondes .3	
1.2 Le vocabulaire et l'apprenant.....	4
1.2.1 Les besoins lexicaux des apprenants	4
1.2.2 Les défis liés à l'apprentissage du vocabulaire	5
1.3 L'enseignement du vocabulaire	6
1.4 La taille du vocabulaire.....	7
1.4.1 La taille du vocabulaire des apprenants du français	8
1.5 Pertinence scientifique et sociale de la recherche.....	10
1.6 Objectifs.....	11
CHAPITRE II CADRE THÉORIQUE	12
2.1 Qu'est-ce que connaître un mot ?	12
2.1.1 L'accès aux connaissances lexicales.....	14
2.2 La taille du vocabulaire et la compétence langagière	16
2.2.1 La TV et la production écrite et orale	17
2.2.2 La TV et la compréhension écrite.....	18
2.2.3 La TV et la compréhension orale	19
2.3 La TV, la couverture de texte et la fréquence	19
2.4 La mesure de la taille du vocabulaire	22
2.4.1 Vocabulary Level Test	22
2.4.2 X_Lex et A_Lex	24
2.5 La TV en contexte scolaire	25
2.6 Questions de recherche et hypothèses.....	28
CHAPITRE III MÉTHODOLOGIE.....	30
3.1 Participants.....	30
3.2 Instruments de collecte de données.....	32
3.2.1 Questionnaire sociodémographique	32

3.2.2	Tests de taille du vocabulaire	33
3.3	Déroulement.....	34
3.4	Traitement des données.....	35
3.4.1	Calcul de la TVO et de la TVPh.....	35
3.4.2	Calcul de la TV	36
CHAPITRE IV	RÉSULTATS.....	37
4.1	Dresser un portrait de la TV.....	37
4.1.1	La TV globale.....	37
4.1.2	La TV par bande de fréquence	40
4.2	Comparaison entre la TVO et la TVPh.....	43
CHAPITRE V	DISCUSSION.....	46
5.1	La TV globale	46
5.2	La TV selon les bandes de fréquence	50
5.3	Le vocabulaire orthographique et phonologique	54
5.4	Les limites.....	57
5.5	Les implications didactiques.....	58
5.6	Les perspectives de recherche.....	61
CONCLUSION	64
ANNEXE A: LETTRE DE CONSENTEMENT	67
ANNEXE B: QUESTIONNAIRE SOCIODÉMOGRAPHIQUE	70
ANNEXE C: LISTE DES MOTS TESTÉS X_LEX	72
ANNEXE D: LISTE DES MOTS TESTÉS A_LEX.....	73
RÉFÉRENCES	74

LISTE DES FIGURES

<u>Figure</u>		<u>Page</u>
1	Qu'est-ce que connaitre un mot ?	13
2	Dual-Route Cascaded Model	15
3	Exemple de regroupement dans le VLT	23
4	Typologie des réponses aux tests oui/non	35
5	Résultats moyens de la taille du vocabulaire	38
6	Comparaison entre la TVO et la TVPh	44

LISTE DES TABLEAUX

<u>Tableau</u>		<u>Page</u>
1	Nombre de mots pour assurer la compétence en L2	21
2	Taille du vocabulaire par année d'étude	26
3	Résultats moyens de la taille du vocabulaire	38
4	Résultats moyens de la taille du vocabulaire par cycle	39
5	Taille du vocabulaire selon la bande la fréquence	40
6	Taille du vocabulaire par secondaire, selon la bande de fréquence	41
7	Effet global de chaque bande de fréquence	41
8	Comparaison entre les années scolaires par bande de fréquence	42
9	Résultats moyens de la TVO (X_Lex) et de la TVPh (A_Lex)	43

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

ALS	Anglais langue seconde
FLE	Français langue étrangère
FLS	Français langue seconde
TV	Taille du vocabulaire
TVO	Taille du vocabulaire orthographique
TVPh	Taille du vocabulaire phonologique
MELS	Ministère de l'Éducation, du Loisir et du Sport
1K	Première tranche de 1 000 mots
2K	Deuxième tranche de 1 000 mots
3K	Troisième tranche de 1 000 mots
4K	Quatrième tranche de 1 000 mots
5K	Cinquième tranche de 1 000 mots

RÉSUMÉ

Cette recherche dresse un portrait des connaissances lexicales phonologique et orthographique des apprenants du français langue seconde de secondaire 1 à 5 au Québec. Elle a pour but d'examiner l'évolution au fil des années scolaires et de mesurer la taille du vocabulaire à la fin de la scolarité

La recherche de Milton et Meara (1998) sur l'évolution de la taille du vocabulaire selon le nombre d'heures de cours d'élèves britanniques en français langue étrangère nous laisse croire que les élèves québécois doivent connaître 4 500 familles de mots à la fin de leur scolarité. Toutefois, ces calculs n'ont été basés que sur les connaissances du vocabulaire à l'écrit quand les programmes québécois de français langue seconde au secondaire mettent l'emphase sur le développement oral et écrit de la langue.

Alors, la taille du vocabulaire de 144 participants a été calculée à l'aide d'un test orthographique (Meara & Milton, 2003), et d'un phonologique (Milton & Hopkins, 2005). Elle n'a pas montré de croissance significative d'une année à l'autre, mais plutôt entre les deux cycles d'apprentissage au secondaire. Les apprenants de secondaire 5 atteignait 2 796 familles de mots en moyenne. De secondaire 1 à 5, les premiers deux milles mots étaient maîtrisés à 80% ou plus, et la taille du vocabulaire orthographique était toujours plus élevée que celle phonologique.

Ces résultats démontrent que l'évolution de la taille du vocabulaire est lente et que les élèves ne finissent pas l'école avec ce qui serait jugé suffisant pour communiquer dans différents contextes (Nation, 2001; 2006). Poursuivre la recherche sur la taille du vocabulaire en FLS contribuerait à mieux cerner les différents besoins lexicaux des apprenants en termes quantitatifs, et aiderait les enseignants à sélectionner le vocabulaire à enseigner et les textes à travailler pour optimiser l'évolution lexicale d'année en année.

MOTS CLÉS : français langue seconde, taille du vocabulaire, vocabulaire orthographique, vocabulaire phonologique, enseignement au secondaire

INTRODUCTION

Qu'est-ce que nos élèves connaissent ? C'est généralement l'une des premières questions qu'une enseignante se pose lorsqu'elle veut construire un cours: le point de départ d'un enseignement étant les connaissances antérieures des élèves pour pouvoir les amener plus loin. Ceci est également vrai dans l'enseignement du lexique. Nous nous sommes donc demandé, dans le cadre de ce mémoire, ce que les apprenants du français langue seconde au secondaire connaissaient comme vocabulaire, ou, en d'autres termes, qu'elle était la taille de leur vocabulaire. Pour ce faire, nous nous sommes intéressés à ce qui la composait, pour en venir à l'évidence que la taille du vocabulaire est composée du vocabulaire orthographique et phonologique et qu'il existait déjà des tests pouvant mesurer ces deux éléments.

Nous désirons atteindre deux objectifs principaux à l'aide de cette recherche. Premièrement, nous souhaitons mesurer la taille du vocabulaire, totale et selon la fréquence des mots, d'apprenants du français langue seconde de secondaire 1 à 5 dans le but de dresser un portrait de l'état actuel des connaissances lexicales des participants. Ensuite, nous voulons comparer la taille du vocabulaire orthographique et phonologique chez ces apprenants, afin de vérifier si ces deux composantes évoluent au fil des années scolaires.

Notre mémoire se divise en cinq chapitres. Le premier chapitre présentera la problématique et les objectifs de recherche. Dans le deuxième chapitre, nous préciserons les bases théoriques de notre recherche pour arriver à poser les questions de recherche de même qu'à proposer des hypothèses pour chacune d'entre elles. Le troisième chapitre sera consacré à la méthodologie. Nous y présenterons les

instruments de mesure utilisés. Au quatrième chapitre, nous rapporterons les analyses statistiques des données recueillies. Dans le cinquième chapitre, nous discuterons des résultats obtenus afin d'atteindre nos deux objectifs. Puis, nous exposerons les limites de notre étude et les implications pédagogiques de nos résultats. Finalement, nous offrirons aux chercheurs quelques pistes de recherches futures dans le même domaine. Pour terminer, nous serons en mesure de tirer des conclusions générales de notre recherche.

CHAPITRE 1

PROBLÉMATIQUE

Dans ce chapitre, il sera question du rôle des connaissances lexicales dans l'acquisition de la langue seconde, de ce que signifie « être plus compétent », des attentes ministérielles quant à l'enseignement et à l'apprentissages du vocabulaire, de ce qui est enseigné en classe et finalement de ce qu'est la taille de vocabulaire. Ces éléments permettront de formuler les objectifs de recherche.

1.1 Le rôle des connaissances lexicales dans l'acquisition es langues secondes

Acquérir une langue seconde n'est pas une mince tâche. Au début de l'apprentissage, tout de la nouvelle langue est inconnu et semble à la limite de l'étrange. L'apprentissage du vocabulaire est à la base de son acquisition, un constat souligné par Wilkins (1972) : « sans connaître la grammaire d'une langue, il y a peu de communication possible, alors que sans vocabulaire, rien ne peut être communiqué » (p. 111, traduction libre). Il semble donc essentiel de mieux comprendre les connaissances lexicales nécessaires pour communiquer, et cela semble d'autant plus évident lorsque l'on constate le rôle de celles-ci dans la compétence langagière en général. En effet, les quatre habiletés langagières (c'est-à-dire la compréhension écrite et orale ainsi que la production écrite et orale) interagissent avec l'apprentissage du lexique puisqu'il a été démontré que la quantité de mots connue par un apprenant (ou la taille de son vocabulaire) est directement corrélée au niveau de maîtrise de ses compétences langagières en L2 (Milton, Wade & Hopkins, 2010 ; Staehr, 2008 ; Meara, 1996 ; Laufer 1992).

Alors, en sachant que les connaissances lexicales d'un apprenant sont primordiales pour bien acquérir la langue, il faut aussi mettre en lumière les défis que présente son apprentissage. Cela permettra de mieux comprendre les besoins des apprenants qui pourront arriver à la décoder et à l'utiliser.

1.2 Le vocabulaire et l'apprenant

Avant d'aborder les défis que représente l'apprentissage du vocabulaire pour les apprenants de langue seconde, il est tout d'abord nécessaire de mieux comprendre ce que les apprenants doivent apprendre.

1.2.1 Les besoins lexicaux des apprenants

Les apprenants de langue seconde ont différents besoins selon ce qu'ils désirent faire avec la langue qu'ils apprennent (Laufer & Ravenhorst-Kalovski, 2010). Par exemple, le fait de connaître les mots fréquents (1-2K) d'une langue permet de l'utiliser à un niveau minimal, soit surtout au niveau conversationnel de base (Nation & Meara, 2002 ; Schmitt, 2000). De plus, pour qu'il y ait compréhension de textes écrits sans aide supplémentaire (comme un dictionnaire) les apprenants doivent maîtriser entre 95 % et 98 % des mots d'un texte (Nation, 2006). Afin de lire des écrits authentiques de sujets variés, cela pourrait correspondre à la connaissance de 8K à 9K familles de mots.

Nous savons donc que les 2K mots les plus fréquents permettent de grossièrement fonctionner dans la langue seconde, mais que, pour comprendre une grande variété de textes, il faut avoir un vocabulaire très développé. Alors, d'une perspective linguistique, identifier les défis que pose l'acquisition de la langue serait d'une grande aide aux apprenants.

1.2.2 Les défis liés à l'apprentissage du vocabulaire

La liste des difficultés qui peuvent se présenter dans l'apprentissage lexical est longue (Laufer, 1997). À titre d'exemple, il y aurait la prononciation, l'utilisation de nouveaux phonèmes, la correspondance entre l'oral et l'écrit qui est peu transparente, la proximité de la forme entre différents mots et même leur longueur. Ces éléments pourraient ajouter un degré de difficulté à l'apprentissage du vocabulaire.

Pour apprendre du nouveau vocabulaire, les apprenants pourraient commencer par les mots les plus fréquents. En effet, le fait de rencontrer souvent un mot, à l'oral ou à l'écrit, faciliterait sa rétention de même que son utilisation (Webb, 2007). Toutefois, il semblerait qu'un mot fréquent soit très complexe à acquérir puisque hautement polysémique (Tidball & Treffers-Daller, 2008), comme le verbe « fatiguer » qui comprend cinq différentes définitions qui ne se distinguent parfois que par une subtile nuance (Dictionnaire Larousse en ligne, 2017). Un mot moins fréquent serait, quant à lui, plus difficile à apprendre puisque rencontré moins souvent dans les discours et les textes écrits (Laufer & Nation, 1995).

Il ne semble donc pas y avoir de raccourci : l'apprentissage du vocabulaire en langue seconde est une lourde tâche et peut sembler décourageante pour les apprenants. Les pratiques en salle de classe doivent toutefois contribuer à faciliter ces apprentissages tout comme l'enseignement du vocabulaire doit être mieux compris.

1.3 L'enseignement du vocabulaire

Les années 1980 ont démontré à quel point le vocabulaire était un élément négligé des curriculums (Nation, 1997), mais, depuis une vingtaine d'années, l'importance de l'inclure semble faire l'unanimité (Nation, 2011 ; Schmitt, 2011 ; Staehr, 2008 ; Milton, 2006 ; Cobb & Horst, 2004 ; Meara, 1995 ; Laufer, 1992). À ce sujet, le ministère de l'Éducation, du Loisir et du Sport (MELS, 2007) précise que :

[...] l'enseignant détermine les besoins langagiers de ses élèves et établit des buts suffisamment élevés mais réalistes quant à l'étendue du vocabulaire à leur faire acquérir. Il choisit les textes à aborder en tenant compte de leurs connaissances lexicales, se souciant à la fois de l'accessibilité et de la richesse du vocabulaire utilisé. De façon générale, il est recommandé de développer au maximum le vocabulaire réceptif (en compréhension orale et écrite) [...] (p. 48)

Toutefois, le *Programme de formation de l'école québécoise* (MELS, 2007) n'offre pas de recommandations quant aux démarches à adopter pour l'enseignement du vocabulaire. Des chercheurs se sont penchés sur cette question et il est largement admis que l'enseignement planifié et explicite doit être favorisé (Nation, 2011), c'est-à-dire qu'il en revient à l'enseignant de sélectionner le vocabulaire à enseigner et de choisir les méthodes utilisées pour le faire. Ceci le cas particulier des mots les plus fréquents de la langue seconde (Nation, 2011 ; Schmitt, 2011 ; Staehr, 2008 ; Milton, 2006 ; Cobb & Horst, 2004 ; Meara, 1995), car il existe des limites à l'apprentissage fortuit (Cobb, 2007), ou en d'autres termes, l'apprentissage non intentionnel de nouveaux mots rencontrés à l'écrit ou à l'oral. Mais, afin de sélectionner le lexique à enseigner aux apprenants de français langue seconde (FLS), un enseignant devrait être en mesure de savoir, de manière fiable, où se situe le niveau de développement lexical de ceux-ci et d'identifier, par exemple, s'ils maîtrisent le vocabulaire fréquent de la langue.

Les enseignants possèdent tout de même un jugement assez juste de la richesse lexicale des productions de leurs apprenants, c'est-à-dire qu'ils sont en mesure d'identifier si les termes utilisés sont fréquents dans la langue française ou non (Tidball & Treffers-Daller, 2008). Leur jugement devrait aussi se transférer à l'analyse de la richesse lexicale des textes lus en classe et cela contribuerait à mieux cibler le vocabulaire et la littérature à l'étude. En effet, un enseignant pourrait choisir d'utiliser un texte comportant un vocabulaire plus ou moins fréquent, selon le niveau des apprenants. Mais, il leur faudra un moyen d'associer la richesse lexicale des écrits et des discours à ce que les apprenants eux-mêmes sont en mesure de comprendre, car, pour qu'il y ait compréhension sans aide supplémentaire, les apprenants doivent maîtriser entre 95 % et 98 % des mots (Nation, 2006).

Le fait de pouvoir sonder nos apprenants sur ce qu'ils connaissent dans le domaine lexical serait une information utile aux enseignants. D'ailleurs, l'utilisation des tests de mesures de la taille du vocabulaire sont répandus et ont été validés (Mochida & Harrington, 2006) et certains sont aussi disponibles en français (Meara & Milton, 2003 ; Milton & Hopkins, 2005 ; Batista, 2014). Les résultats auxquels nous avons accès en anglais langue seconde (ALS) montrent que le vocabulaire fréquent est généralement mal maîtrisé (Horst & Collins, 2006), mais, actuellement, nous avons peu de données sur la taille du vocabulaire des apprenants du FLS et nous ignorons toujours, par exemple, s'ils maîtrisent les mots les plus fréquents du français.

1.4 La taille du vocabulaire

La taille du vocabulaire (TV) se définit par le nombre de mots qu'un apprenant connaît dans une langue (Nation & Waring, 1997), ou son étendue (concepts définis au deuxième chapitre). Lorsqu'on la mesure, elle se divise par tranche de 1 000 mots et cela permet donc d'identifier si un apprenant connaît les 1K mots les plus fréquents

de la langue, 2K mots, 3K mots ou plus. Les recherches qui ont examiné l'importance de la TV ont bien démontré une corrélation significative à la compétence langagière (Laufer, 1992 ; Qian, 1999 ; Beglar & Hunt, 1999 ; Nation, 2001 ; Zimmerman, 2004 ; Milton, 2006 ; Nation, 2006 ; Stearh, 2008 ; Milton *et al.*, 2010, Milton, 2013 ; Van Zeeland, 2013). Généralement, la TV est mesurée à l'écrit (la TV orthographique) et les résultats démontrent une corrélation forte aux compétences écrites (Staehr, 2008 ; Meara, 1996 ; Laufer, 1992). Quant aux deux seules études qui, à notre connaissance, ont mesuré la TV à l'oral (la TV phonologique), la tendance dégagée montre une forte corrélation aux compétences orales (Milton *et al.*, 2010 ; Milton & Hopkins, 2006).

La TV a été fortement corrélée à la compétence langagière, mais principalement en contexte d'anglais langue étrangère. Il semble donc y avoir un réel intérêt à mesurer la TV des apprenants du FLS, et, tant en anglais qu'en français, les données sur l'aspect phonologique manquent. De plus, l'importance attribuée aux formes orale et écrite dans le programme du MELS (2007) rend ce manque à combler encore plus évident.

1.4.1 La taille du vocabulaire des apprenants du français

En français langue étrangère (FLE), dans un contexte d'apprentissage où le français n'est pas utilisé en dehors de la salle de classe, plusieurs études ont évalué la TV des apprenants à l'école secondaire (David, 2008 ; Bulté *et al.*, 2008 ; Milton & Meara, 1998), à l'aide d'une mesure à l'écrit. Milton et Meara se sont penchés sur la TV selon le nombre d'heures d'études par année. Ces apprenants britanniques du FLE avaient une TV qui progressait de trois à quatre mots par heure de cours en moyenne. En sachant que les élèves québécois ont environ 150 heures de FLS par année à l'école primaire et secondaire (*Loi sur l'instruction publique*, au 1^{er} mars 2017), donc plus de 1 500 heures d'étude du français dans le parcours scolaire obligatoire, il serait

juste de penser que la taille de leur vocabulaire, une fois en secondaire 5, atteigne au moins 4 500 mots.

À notre connaissance, seulement une étude a examiné la TV des apprenants du français au Québec, encore une fois avec une mesure à l'écrit. Blais (2014) a mesuré celle de 17 participants adultes, nouvellement arrivés au Québec et apprenant le français. Cette étude avait, entre autres, pour but de déterminer s'ils maîtrisaient les mots fréquents de cette langue, mais les résultats ont démontré qu'ils étaient mal maîtrisés par les nouveaux arrivants et que cela affectait leurs résultats aux examens d'habiletés langagières, en particulier ceux de la compréhension orale. Précisons que Blais a recommandé l'utilisation d'un test phonologique, étant donné la corrélation étroite démontrée entre la TV et la compréhension orale dans son étude. Elle souhaitait fournir des données sur la TV qui soient plus complètes.

Les résultats de cette étude sont intéressants, mais ils n'éclairent en rien la situation lexicale des apprenants dans le système scolaire. Dresser un portrait du développement lexical des élèves au secondaire semble être nécessaire étant donné que tous ceux qui fréquentent des écoles non francophones doivent apprendre le FLS dès le début de leur scolarité. De plus, ces informations pourraient être fortement utiles aux enseignants, puisqu'elles leur permettraient, d'une part, de cibler le vocabulaire à travailler de façon explicite si les mots les plus fréquents ne sont pas maîtrisés (Staerh, 2008 ; Coady, 1997) et, d'autre part, de pouvoir faire une sélection des textes à l'étude qui seraient davantage à la portée des élèves.

1.5 Pertinence scientifique et sociale de la recherche

Malgré le fait que la TV a été étudiée abondamment en anglais langue seconde et étrangère, il n'en demeure pas moins que la TV en français langue seconde a été très peu abordée et notre recherche s'avère pertinente pour différentes raisons.

D'abord, l'étude fournira de nouvelles données empiriques sur la taille du vocabulaire auprès d'apprenants qui sont en contact permanent avec la langue seconde, ce qui change les paramètres de l'apprentissage. Cela semble encore plus évident lorsqu'on la compare aux apprenants du FLE et/ou de l'ALE des études sur le même sujet. De plus, il y a peu de données confirmant la taille du vocabulaire nécessaire en français pour qu'un apprenant comprenne 95 % à 98% des mots d'un texte, seuil assurant une compréhension sans aide (Nation, 2006). Par ailleurs, nous ne savons pas non plus quelle est la TV des apprenants dans le système scolaire québécois, alors cette étude suggère une première étape à la recherche sur la TV en français.

Peu d'études ont mesuré à la fois la TV orthographique et phonologique. Toutefois, reconnaître la forme d'un mot se fait nécessairement à l'oral et à l'écrit (Milton et Hopkins, 2006). L'utilisation d'un seul test écrit pourrait sous-estimer la TV chez des apprenants ayant un problème avec l'orthographe. De plus, étant donné que les recherches antérieures (Milton *et al.*, 2010) ont démontré une plus forte corrélation entre la TV orthographique et les compétences écrites, et entre la TV phonologique et les compétences orales, il nous semble nécessaire d'inclure ces deux mesures pour mieux caractériser les connaissances lexicales des apprenants.

Dans une perspective sociale, les connaissances dont nous disposons au sujet de la TV chez des apprenants au secondaire nous viennent surtout du Royaume-Uni. Au Québec, tel qu'exposé précédemment, seule Blais (2014) a mesuré la TV

d'apprenants adultes du français, à notre connaissance. Pourtant, le MELS lui-même parle de développer le vocabulaire réceptif à l'oral et à l'écrit (2007) et il semble étonnant qu'il n'y ait pas de données qui puissent guider les enseignants pour les aider à sélectionner le vocabulaire à enseigner. Les données qui seront recueillies dans cette recherche permettraient aux enseignants d'obtenir plus de précisions quant au vocabulaire à enseigner, de sélectionner des textes et discours qui soient davantage au niveau des apprenants et d'adapter les contenus d'enseignement plus adéquatement. Cela signifie pour les apprenants que les textes lus ou entendus seraient plus accessibles et mieux compris puisque adaptés à leur besoin. Cela leur permettrait aussi de développer leurs connaissances lexicales en fonction des données obtenues sur leur maîtrise du vocabulaire, et ainsi de relever des défis à leur portée. Bref, fournir ces informations semble donc important pour aider les enseignants à adapter leur enseignement au niveau des besoins lexicaux des apprenants.

1.6 Objectifs

Étant donné le manque de données sur la TV des apprenants du FLS au secondaire, et ce, tant au niveau orthographique que phonologique, la présente étude se donne comme objectifs de:

- mesurer la taille du vocabulaire, totale et selon la fréquence des mots, d'apprenants du français langue seconde de secondaire 1 à 5 ;
- comparer la taille du vocabulaire orthographique et phonologique chez ces apprenants de secondaire 1 à 5.

CHAPITRE II

CADRE THÉORIQUE

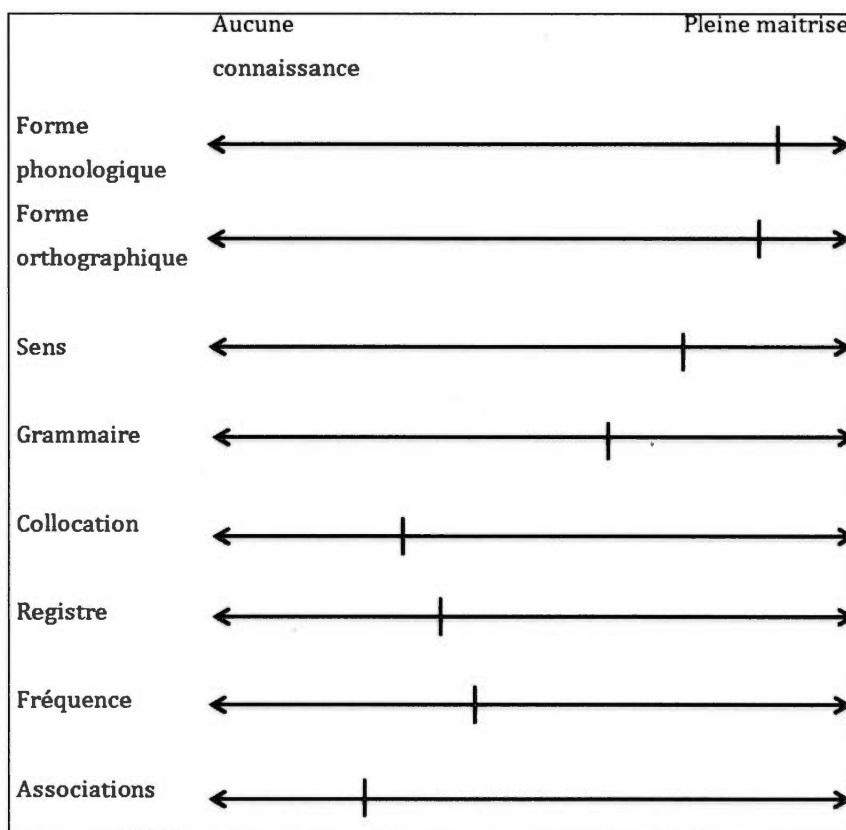
Dans ce chapitre, nous présenterons les concepts centraux de la recherche en mesure du vocabulaire : ce que signifie connaître un mot et quelques notions linguistiques, le lien direct entre la taille du vocabulaire (TV) et la compétence langagière, la couverture de texte et la fréquence des mots en français et, finalement, la mesure du vocabulaire. Ces définitions permettront de formuler les questions de recherche et les hypothèses.

2.1 Qu'est-ce que connaître un mot ?

Nation (1990) inspiré de Richards (1976) a organisé la connaissance lexicale autour de trois pôles principaux, allant de la compétence réceptive, ou ce qui est reconnu et compris d'un mot, vers la compétence productive, ou l'emploi approprié de la langue seconde. D'abord, le premier pôle réfère à la connaissance de la forme des mots et est essentiellement une compétence réceptive : sa phonologie, sa prononciation, son orthographe et la reconnaissance de ses affixes le cas échéant. Ensuite, le deuxième pôle intègre la connaissance du sens des mots et touche tant à la compétence réceptive que productive : la relation entre la forme et le sens, les concepts et les référents, puis les associations, comme la synonymie. Le troisième pôle concerne les connaissances productive d'un item, c'est-à-dire son emploi : ses fonctions grammaticales, ses collocations et ses contraintes lors de l'utilisation tel le registre. Ces trois pôles ne se hiérarchisent pas nécessairement, puisqu'il serait possible, par exemple, de connaître certaines collocations, comme « avoir faim », sans toutefois pouvoir lui associer un synonyme comme « fringale » (Schmitt & Meara, 1997).

Notons que la compétence réceptive est généralement plus grande en taille, et qu'il existe des tests qui la mesurent sans toutefois dissocier la forme (pôle 1) et le sens (pôle 2) (Milton, 2009). Dans la figure 1, Schmitt (2010) montre bien que le degré de maîtrise peut varier selon les différentes dimensions.

Figure 1 Qu'est-ce que connaître un mot ? [Traduction libre] (Schmitt, 2010, p.38)



Chacune des dimensions de la connaissance lexicale se développe simultanément et forme un tout, sans que l'une ait plus d'importance qu'une autre et même, parfois, sans qu'une dimension ne soit jamais maîtrisée (Schmitt & Meara, 1997). La maîtrise d'une, plusieurs ou toutes ces dimensions pour un terme signifie le connaître de plus en plus en profondeur. Toutefois, lors de la communication, tant en L1 qu'en L2,

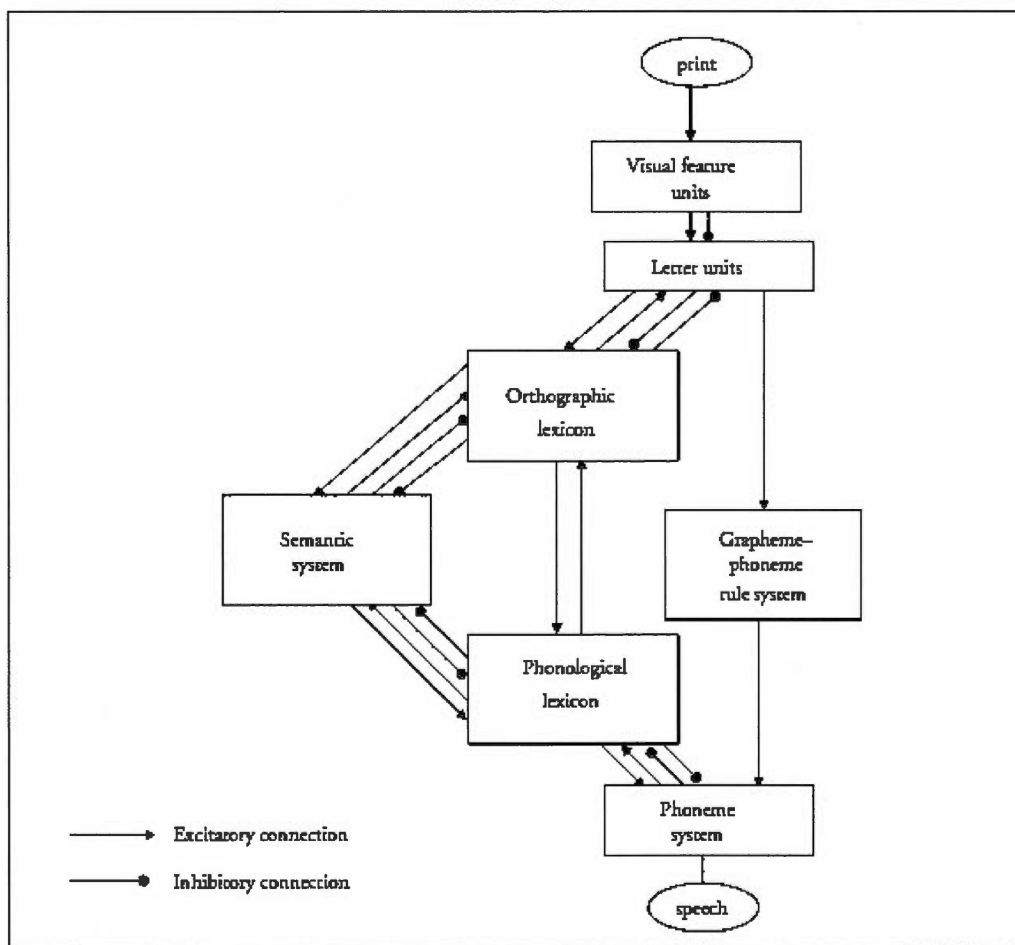
certaines tâches favorisent l'utilisation de certaines dimensions plutôt que d'autres. Par exemple, lors d'une entrevue orale, beaucoup de connaissances au niveau du registre et de la forme phonologique sont essentielles, alors qu'il y a peu de sollicitation au niveau de la forme orthographique. Laufer (1997) élabore chaque dimension en présentant certaines des difficultés qu'un apprenant de L2 peut rencontrer lors de son apprentissage. D'abord, certains phonèmes non présents dans la langue maternelle du locuteur peuvent venir compliquer l'apprentissage de la forme phonologique. Ensuite, la polysémie d'un mot peut rendre sa maîtrise, au niveau sémantique, plus complexe. Puis, au niveau du registre, certains termes s'utilisent différemment d'un pays à l'autre. Par exemple au Québec, le nom « menterie », comme dans « dire des menteries », est un usage familier, tandis qu'en France, il est archaïque et ne s'utilise pas vraiment. On lui préférera le terme « mensonge » (Dictionnaire Larousse en ligne, 2017). Alors, tel que démontré précédemment, la reconnaissance orthographique et/ou phonologique doit être mobilisée dans la maîtrise des dimensions de la connaissance lexicale et deux stratégies pour y arriver seront décrites à la section suivante. De plus, il sera question du lien entre la forme et le sens qui doit s'opérer pour utiliser les connaissances lexicales.

2.1 L'accès aux connaissances lexicales

Coltheart et al. (2001) expliquent que l'accès aux connaissances lexicales se fait par deux voies. D'abord, le transfert de la forme phonémique vers sa représentation mentale permettrait d'accéder au mot. C'est ce qu'ils nomment « la stratégie phonémique » (p.217). Nous nommerons, dans le cadre de cette recherche, le vocabulaire récupéré de la forme phonémique le vocabulaire phonologique. Ensuite, la forme orthographique peut aussi contribuer à accéder au mot et permettrait, par exemple, de différencier des homophones. Les chercheurs nomment ce processus « la stratégie visuelle » (p.217). Puisque c'est la forme orthographique qui est sollicitée, il

sera donc question du vocabulaire orthographique. Voici une représentation de ce modèle, le *Dual-Route Cascaded Model (DRC)*.

Figure 2 Dual-Route Cascaded Model (DRC) (Coltheart *et al.*, 2001, p.213)



Ainsi, la représentation mentale d'un mot peut être rappelée par sa forme orale de même que par sa forme écrite (Dahan & Magnuson, 2006), sans que les deux codes n'interagissent entre eux, mais, tel que démontré par le DRC, en accédant indépendamment au sens du mot. La représentation du lexique n'est donc pas statique, mais en changement continu, puisque, notamment, la maîtrise du

vocabulaire phonologique pourrait faire partie ou non du vocabulaire orthographique et tout de même permettre l'accès à la sémantique du mot.

Cette relation entre la phonologie, l'orthographe et la sémantique a aussi été développée par Perfetti et Hart (2002), dans un modèle appelé « l'hypothèse de la qualité lexicale ». Une représentation lexicale serait de haute qualité si elle lie trois composantes : l'information phonologique, orthographique et sémantique. De plus, l'identification d'un mot doit comprendre ces trois informations. Pour un apprenant, le fait de rencontrer un mot plusieurs fois en contexte, par exemple par la lecture, permet de solidifier cette relation. Les chercheurs ont aussi illustré cette dernière en spécifiant que si une de ces composantes était moins bien maîtrisée, la connaissance du mot n'était pas dite « de qualité ». Ainsi, on ne pourrait pas dire qu'un mot est maîtrisé s'il est mal prononcé lors de sa lecture ou s'il est mal orthographié, tout comme s'il n'est compris que partiellement. Ce modèle non seulement contribue à renforcer la relation entre la forme et le sens décrite par Nation (1990), mais également définit le fait que la forme des mots doit se présenter tant à l'oral qu'à l'écrit.

Pour terminer, la connaissance lexicale prend forme par les quatre compétences en langue seconde : la compréhension orale et écrite, ainsi que la production orale et écrite (Schmitt, 2013b). En effet, la corrélation entre les quatre compétences en L2 et la taille du vocabulaire a largement été démontrée.

2.2 La taille du vocabulaire et la compétence langagière

La taille du vocabulaire (TV) permet de montrer comment la compétence langagière chez les apprenants varie (Milton, Wade & Hopkins, 2010), et, effectivement, la corrélation entre la TV et la compétence langagière a rapidement été démontrée et

continue de l'être aujourd'hui (Laufer, 1992 ; Qian, 1999 ; Beglar & Hunt, 1999 ; Nation, 2001 ; Milton, 2006 ; Nation, 2006 ; Stearh, 2008 ; Milton *et al.*, 2010, Milton, 2013 ; Van Zeeland, 2013). Chacun de ces chercheurs ont démontré l'existence d'une corrélation entre la TV et une ou plusieurs des quatre compétences en L2.

2.2.1 La TV et la production écrite et orale

Le lien entre la qualité des productions écrites des apprenants et la richesse lexicale a été démontrée. La qualité lexicale était évaluée de façon globale à l'aide de grilles d'observation (Astika, 1993; Nation 2001; Daller & Phelan, 2007). Généralement, ils ont démontré que la richesse lexicale expliquait la variance dans les résultats aux productions écrites, même que chez Astika (1993), le vocabulaire expliquait 84% de la variance dans les résultats. Toutefois, peu ont montré un lien corrélationnel entre la TV à proprement dit et la production écrite. À l'aide d'un test qui mesure la taille du vocabulaire réceptif et de la composition d'une lettre pour postuler à un emploi, Staerh (2008) a démontré une corrélation significative entre la TV et la production écrite (0.73) chez 88 apprenants néerlandais de 15-16 ans. Il précise toutefois que ce résultat est surprenant puisque c'est le vocabulaire réceptif qui était mesuré pour le comparer à une compétence productive. Il mentionne également que cette mesure n'est pas tout à fait complète puisqu'il y a de nombreux facteurs autres que le vocabulaire qui font qu'une production écrite soit de bonne qualité. Il suggère l'utilisation de multiples mesures de la taille du vocabulaire (réceptif / productif et orthographique / phonologique) pour analyser la corrélation entre la TV et les compétences langagières. À l'aide d'un test orthographique (X_Lex) et d'un test phonologique (A_Lex), Milton *et al.* (2010) ont montré une forte corrélation entre la taille du vocabulaire orthographique (TVO) et la production écrite (0.76), mais plutôt faible avec la taille du vocabulaire phonologique (TVPh)(0.44). (voir la section 2.2.2 pour un résumé de l'étude)

Pour ce qui est de la production orale, seuls Milton *et al.* (2010) ont montré une forte corrélation entre la production orale et la TVPh (0.71), mais qu'elle était faible lorsque comparée à la TVO (0.30).

Dû à des différences méthodologiques importantes et puisque très peu de données sont disponibles au sujet du lien entre la TV et les compétences en production, ce sont les compétences réceptives qui seront davantage analysées dans cette recherche.

2.2.2 La TV et la compréhension écrite

La corrélation entre la TV et la compréhension écrite a été étudiée dans différents contextes et a montré des résultats significatifs. À l'aide du test de TV créé par Nation (1990), Laufer (1992) (n=64) a confirmé que plus un apprenant connaît de mots dans la L2, plus il est en mesure de bien performer dans des tests de lecture de niveau universitaire. Qian (1999) a même ajouté que, chez ses 74 participants universitaires d'un programme intensif d'anglais L2, la TV était en corrélation non seulement avec la lecture, mais aussi avec la profondeur de la connaissance lexicale. Staerh (2008) a rapporté des résultats tout aussi concluants auprès de 88 apprenants néerlandais de 15-16 ans, montrant une forte corrélation (0.83) entre la TV et la lecture. Dans l'étude de Milton *et al.* (2010), la TV de 30 apprenants universitaires de niveau intermédiaire-avancé a été mesurée afin de la comparer à leurs résultats à l'*International English Language Testing System* (IELTS). À l'aide des tests de TV orthographique (Meara & Milton, 2003) et phonologique (Milton & Hopkins, 2006), Milton *et al.* (2010) ont, encore une fois, montré que la TV orthographique était en forte corrélation avec la compréhension écrite et que la TV phonologique l'était elle aussi avec les composantes orales de la langue. En considérant tous les résultats rapportés, il semble évident que la TV est une bonne indication de la compétence en compréhension écrite des apprenants en langue seconde.

2.2.3 La TV et la compréhension orale

La compréhension orale semble aussi être bien corrélée avec la TV (Staerh, 2008, 2009 ; Milton & Hopkins, 2006 ; Milton, *et al.*, 2010). Staerh (2009), dans une étude auprès de 115 participants, a conclu que l'étendue du vocabulaire et la profondeur de la connaissance lexicale étaient corrélées à la compréhension orale (0.70 et 0.65). Il a aussi précisé que seule la TVO avait été mesurée dans sa recherche et que la mesure de la TVPh pourrait ajouter un degré de fiabilité aux résultats obtenus. Il proposait de poursuivre la recherche à l'aide de multiples mesures, ce que Milton *et al.* (2010) ont par la suite accompli. La TVO et la TVPh ont été mesurées séparément et la TVPh a montré une corrélation forte avec la compréhension orale (0.67) de même qu'expliqué 45 % de la variance dans les résultats aux tests de compréhension orale du IELTS.

Si la TV semble influencer la compréhension des apprenants de la L2 à l'écrit et à l'oral, il est donc primordial de comprendre le nombre de mots dont un apprenant a besoin pour décoder un texte oral ou écrit. Cette information permettrait aux acteurs en didactique des langues de se questionner sur l'enseignement et l'apprentissage de la L2, plus précisément en ce qui a trait à la couverture de texte et à la fréquence des mots.

2.3 La TV, la couverture de texte et la fréquence

Dans une perspective connexionniste, la fréquence des mots dans une langue est considérée comme un élément clé pour son apprentissage (Ellis & Robinson, 2008; Ellis & Wulff, 2014). La fréquence pourrait servir à sélectionner le vocabulaire à enseigner (David, 2008 ; Richards & Malvern, 2008), mais encore faut-il savoir où se

situent les apprenants au départ. En effet, lire un texte en langue seconde peut présenter plusieurs défis, entre autres en ce qui a trait à la connaissance des mots lus (Laufer, 1992 ; Nation, 2001). Il a été évalué que, si l'apprenant connaissait 95 % des mots lus, soit un mot inconnu sur 20, c'était encore une minorité d'apprenants qui arrivait à saisir le sens du texte, alors qu'à 98 %, soit un mot inconnu sur 50, la majorité des apprenants arrivait à une compréhension suffisante sans aide (Nation, 2001). Ainsi, Nation (2006) s'est demandé combien de mots étaient nécessaires pour arriver à lire un texte authentique en langue seconde, de même que pour comprendre un texte à l'oral. Si un apprenant désire lire des textes authentiques provenant de médias ou de romans de façon indépendante, 8K à 9K familles de mots seraient nécessaires pour atteindre le seuil de 98 % de compréhension de texte. Cependant, puisque chaque apprenant n'a pas les mêmes objectifs d'apprentissage, Laufer et Ravenhorst-Kalovski (2010) ont ajouté que, pour lire des textes de façon guidée, par exemple à l'aide de discussions, la connaissance de 4K à 5K familles de mots, ainsi que les noms propres, était nécessaire pour atteindre une couverture de texte de 95 %.

En ce qui concerne la compréhension du discours oral, il était d'abord convenu que la maîtrise de 2K familles de mots était suffisante pour permettre à un apprenant de s'engager dans une conversation de la vie quotidienne, puisque cela correspondait à 99 % de la couverture d'un texte (Nation & Meara, 2002 ; Schmitt, 2000). Cette mesure du vocabulaire oral avait été prise grâce à une étude menée en 1956 auprès de travailleurs australiens (Schonell *et al.*, 1956) et, comme le corpus utilisé était assez limité (Adolphs & Schmitt, 2003), ces derniers ont saisi l'occasion de moderniser les données au sujet de la couverture de textes oraux. Ainsi, en analysant deux corpus comprenant des conversations et des conférences totalisant plus de 4 millions de mots chacun (Adolphs & Schmitt, 2003), ils ont déterminé que, pour atteindre 96 % de couverture d'un discours oral de la vie quotidienne, environ 5K familles de mots étaient nécessaires et que cela suffisait pour comprendre. Nation (2006), quant à lui, a

utilisé des dialogues de films pour enfants et des conversations enregistrées et a découvert que, afin d'obtenir 95 % de couverture de texte, ce serait de 3K à 4K familles de mots qui seraient nécessaires, et de 6K à 7K familles de mots pour 98 % de couverture de texte (2006). Quant à eux, VanZeeland et Schmitt (2012) ont analysé la composition de quatre textes oraux narratifs de même que mesuré la compréhension orale de 36 locuteurs natifs et de 40 non natifs pour trouver que 2K à 3K familles de mots pour une couverture de 95 % étaient suffisants pour une compréhension orale adéquate. Le tableau de Schmitt (2013a) propose un sommaire de toutes ces données concernant le nombre de mots, la couverture de texte et le niveau de compétence langagière (Tableau 1).

Tableau 1 Nombre de mots pour assurer la compréhension en L2 [Traduction libre] (Schmitt, 2013a, p. 6)

TV minimum pour permettre une communication de base en L2	2K-3K familles de mots
TV permettant une conversation sur une grande variété de sujets	5K-7K familles de mots
TV permettant de lire des textes authentiques sur des sujets de base	4K-5K familles de mots
TV permettant de lire des textes authentiques de sujets variés	8K-9K familles de mots

Les données précédentes concernent la langue anglaise mais les recherches en français langue seconde à propos de la fréquence lexicale à maîtriser pour comprendre des textes écrits et oraux manquent. Par exemple, à l'aide de *VocabProfile* (Cobb, 2017 ; Heatley, Nation & Coxhead, 2002), un outil qui permet d'analyser les mots d'un texte et de les classer selon leur fréquence, Cobb et Horst (2004) ont évalué différents journaux écrits en français. Ainsi, ils ont découvert que les 2K familles de mots les plus fréquents en français permettaient une couverture de texte allant jusqu'à 85 %. Comme il s'agit de la seule donnée disponible en français

et qu'une couverture de 85 % est reconnue comme insuffisante pour assurer la compréhension de textes écrits et parlés (Nation, 2001; Adolphs & Schmitt, 2003; Nation, 2006; Laufer & Ravenhorst-Kalovski, 2010; VanZeeland & Schmitt, 2012), il est souhaitable de continuer la recherche pour comprendre la TV qui est nécessaire à la compréhension de différents textes en français, tant à l'oral qu'à l'écrit. En outre, il reste que la mesure de la TV n'a pas encore été abordée et différentes façons de la mesurer seront présentées.

2.4 La mesure de la taille du vocabulaire

Dans la mesure de la taille du vocabulaire, deux tests ont majoritairement influencé le développement de la recherche. D'une part, il y a le *Vocabulary Level Test* (VLT) (Nation, 1990) et, d'autre part, il y a le *Eurocentres Vocabulary Size Test* (Meara & Buxton, 1987) qui est devenu, par la suite, X_Lex (Meara & Milton, 2003), une version informatisée. Finalement, Milton et Hopkins (2005) ont créé un test phonologique inspiré de X_Lex, nommé A_Lex, en suivant exactement le même modèle. Le VLT, X_lex et A_Lex seront brièvement présentés.

2.4.1 Vocabulary Level Test

Ce test créé par Nation (1990) mesure la taille de vocabulaire des 2K, 3K, 5K et 10K mots les plus fréquents. Il s'agit d'un test à choix multiples où un participant doit associer trois définitions de la colonne de droite à trois des six mots de la colonne de gauche. La figure 3 en montre un exemple. Chaque niveau comprend six regroupements, pour un total de 18 mots testés.

Figure 3 Exemple de regroupement dans le VLT

1 desolate	
2 fragrant	_____ good for your health
3 gloomy	_____ sweet-smelling
4 profound	_____ dark or sad
5 radical	
6 wholesome	

Le test de Nation a été grandement utilisé (Laufer, 1992 ; Schmitt, Schmitt & Clapham, 2001 ; Staerh, 2009). Cependant, d'importants problèmes se posent lors de son utilisation. Tout d'abord, il semble que le fait de donner des réponses aléatoires (et ainsi de surestimer la TV) ne soit pas contrôlé dans ce test (Kamimoto, 2014). En effet, rien n'empêche un participant de choisir au hasard le mot qui conviendrait le mieux à une définition. Or, aucune méthode pour pallier ce problème n'a été conçue, ce qui engendrerait une possible surestimation de la taille du vocabulaire chez les participants d'environ 19 % (Kamimoto, 2014). Ensuite, en principe, la réponse à un item devrait être indépendante de celle d'un autre item, car, dans le cas contraire, la possibilité de répondre de façon aléatoire, et ce correctement, serait augmentée. Kamimoto a donc démontré comment les items du VLT étaient dépendants, contrairement à ce qu'avaient démontré Schmitt, Schmitt et Clapham (2001). Il a demandé à ses participants japonais apprenants de l'anglais langue étrangère de compléter le VLT original ainsi qu'une version modifiée où deux groupements comprenant 18 mots dans la colonne de gauche et neuf définitions dans la colonne de droite étaient proposés. Il s'agissait des mêmes mots et définitions dans les deux versions. Cette étude a montré que les participants avaient des résultats de 15 % plus élevés au VLT original. Il n'était donc pas possible de dire que les items du VLT possédaient une réponse unique.

Le VLT ne teste pas la première tranche de 1K mots, ce qui est problématique étant donné l'importance qu'ont les 2K mots les plus fréquents pour l'enseignement et l'apprentissage de la L2 (Staerh, 2008). Comme le montre l'étude de Horst et Collins (2006) chez de jeunes francophones apprenant l'anglais langue seconde, les mots les plus fréquents de la langue anglaise sont souvent évités, car les apprenants semblent préférer des termes moins fréquents et souvent congénères du français, c'est-à-dire ayant le même sens dans les deux langues. Par exemple, un de leur participant avait utilisé le terme « sensation », mot moins fréquent en anglais, connu et compris dans la L1, plutôt que « *feeling* », qui est plus fréquent en anglais. Comme il serait possible de penser que les jeunes anglophones apprenant le FLS se comportent sensiblement de la même façon, le VLT ne serait donc pas le meilleur test à utiliser afin de mesurer le vocabulaire fréquent étant donné l'absence de la première tranche de mille mots.

2.4.2 X_Lex et A_Lex

Les tests X_Lex et A_Lex mesurent la quantité de mots connus dans chaque bande de fréquence, de la première tranche de mille mots à la cinquième tranche (Meara & Milton, 2003; Milton & Hopkins, 2005). La seule différence entre ces deux tests est le médium d'exposition. Le test X_Lex présente les mots à l'écrit, tandis que le test A_Lex les présente à l'oral. De plus, il s'agit d'une liste de vérification où un participant indique s'il connaît ou non le sens d'une lexie (un test oui/non). Donc, ce test permet d'obtenir des résultats en utilisant un grand nombre de termes, en très peu de temps (Pellicer-Sanchez & Schmitt, 2012 ; Beekmans *et al.*, 2001 ; Read, 2000).

Afin d'éviter une surestimation de la TV dans un test de format oui/non, une série de pseudomots est insérée pour ainsi permettre de contrôler les résultats. Les réponses aux pseudomots sont utilisées pour confirmer que celles aux vrais mots sont justes et que le score final est sans surestimation. Beekmans *et al.* (2001) mentionnaient aussi

l'attention particulière qu'il faut porter aux pseudomots, et, à ce sujet, Meara (1996) mettait aussi en relief le fait que l'intégration de faux amis pouvait venir fausser les données. En effet, si un participant cochait « oui » pour un mot qui existe dans sa langue maternelle (sans nécessairement en avoir une connaissance dans la langue seconde), cela pourrait avoir un impact négatif sur le score d'un participant. Le fait qu'il s'agisse d'un pseudomot aurait même beaucoup plus d'effet, étant donné le calcul utilisé (voir la section 3.4.1 dans le chapitre méthodologie).

Si les tests oui/non n'étudient pas la profondeur de la connaissance lexicale (Beekmans *et al.*, 2001), la combinaison de X_Lex et de A_Lex permet tout de même de mesurer la TV dans deux dimensions de la connaissance d'un mot, soit la forme orthographique et la forme phonologique. De plus, mesurer l'étendue du vocabulaire des apprenants est tout à fait valable, surtout que, comme la TV est en forte corrélation avec la compétence langagière, cela laisse croire que la reconnaissance de la forme des mots est aussi indicatrice de la connaissance du sens de ceux-ci, à divers degrés.

Finalement, la TV peut être mesurée de différentes façons et il est possible, voire important, de mesurer la TVO et la TVPh, tout particulièrement parce que le système scolaire québécois insiste sur le fait que chacune des compétences en langue seconde doit être développée tant à l'oral qu'à l'écrit.

2.5 La TV en contexte scolaire

Dans son parcours scolaire, il semble logique de penser qu'un apprenant de FLS accroît l'étendue de son vocabulaire au fil des ans, mais peu de chercheurs se sont intéressés à la TV selon le nombre d'années d'étude du français. Milton et Meara

(1998) ont testé des apprenants du français langue étrangère (FLE) britanniques après 210 heures d'instruction à l'aide de X_Lex et ceux-ci avaient une TV de 660 mots en moyenne. Par la suite, Milton (2006) a mesuré la TV des apprenants du FLE britanniques de la 7^e à la 13^e année d'études. Il faut noter, toutefois, que les données concernant le nombre d'heures d'instruction ne semblent pas complètement fiables, puisque le chercheur indique le nombre d'heures de français suivi par année, mais ne spécifie pas depuis combien de temps ces apprenants l'apprennent.

Tableau 2 Taille du vocabulaire par année d'étude (Milton, 2006, p. 192)

year	learners	number	hours/year	mean	max	min	Sd
1	School Year 7	80	78	311	900	0	193
2	School Year 8	78	58.5	411	1150	0	228
3	School Year 9	106	58.5	460	1150	0	261
4	School Year 10	57	78	592	1350	50	263
5	School Year 11	49	78	852	1800	0	440
6	School Year 12	10	175.5	1555	2000	1100	349
7	School Year 13	69	117	1930	3100	650	475

Milton et Meara (1998) avaient estimé un progrès dans l'apprentissage lexical de trois à quatre mots par heure d'instruction, ce qui a aussi été validé par Milton (2006). De plus, plus un apprenant prend de l'expérience dans la langue seconde, plus sa TVO augmente en étendue en comparaison à la TVPh (Milton & Hopkins, 2006). De telles données nous permettent de nous questionner à plusieurs niveaux sur nos apprenants québécois de français langue seconde au sujet de leur TV et de leur progrès au cours des études secondaires. Il faut se demander si leur vocabulaire est suffisant pour satisfaire aux exigences du *Programme de formation québécoise* (MELS, 2007), qui précise que :

À la fin du deuxième cycle du secondaire, l'élève lit et comprend avec facilité une diversité de textes de forme orale, écrite, visuelle ou mixte susceptibles de répondre aux besoins scolaires, personnels, sociaux ou professionnels d'un jeune de son âge. (p. 23)

Combien de mots un apprenant doit-il connaître afin de manifester une telle compétence ? Cette information n'est pas précisée dans le *Programme de formation québécoise* (MELS, 2007), mais sachant qu'on demande à ce que les textes soient compris tant au niveau oral qu'écrit, une mesure du vocabulaire touchant à ces deux éléments de la langue doit être considérée.

Jusqu'à maintenant, à notre connaissance, une seule étude a mesuré la TV dans le système scolaire québécois, auprès d'un public adulte immigrant qui suivait un programme de francisation divisé en 6 niveaux. Blais (2014) a montré, à l'aide de l'outil de mesure X_Lex (Meara & Milton, 2003) que les 17 participants du sixième niveau avaient une TVO moyenne de 2 228 mots, que les mots fréquents du français étaient mal maîtrisés par plusieurs et qu'une corrélation forte était présente entre la TV et les tests de compréhension écrite et orale. Elle expliquait la faible maîtrise des mots fréquents du français par l'orientation thématique du programme de francisation et le manque de recyclage des termes appris qui ne favoriseraient pas l'apprentissage du vocabulaire fréquent. Elle a aussi remis en question la fiabilité de la corrélation établie entre la TV et la compréhension orale, puisque seule une mesure de la TVO a été utilisée. Elle suggérait l'utilisation d'un test phonologique pour ajouter aux données recueillies dans son étude.

Il reste donc beaucoup d'inconnu quant aux connaissances lexicales dans le système scolaire québécois en dépit de l'importance de l'apprentissage lexical pour maîtriser une langue seconde (Nation, 2011 ; Schmitt, 2011 ; Staehr, 2008 ; Milton, 2006 ; Cobb & Horst, 2004 ; Meara, 1995 ; Laufer, 1992) et de l'importance accordée au lexique dans les programmes scolaires québécois (MELS, 2007). Nous connaissons peu de nos apprenants du FLS au sujet de leur TV, orthographique et phonologique (rarement intégrée), et de leur maîtrise des mots fréquents du français. Étant donnée

l'étude menée auprès des élèves britanniques en FLE qui fournit une piste quant à la progression des connaissances lexicales au secondaire (Milton & Meara, 1998), cette étude ciblera aussi ce public d'apprenants.

2.6 Questions de recherche et hypothèses

Cette recherche a pour mission première de dresser un portrait de la TV, qui inclurait une mesure orthographique et une phonologique, chez des apprenants du français langue seconde au secondaire en mesurant d'abord la taille de vocabulaire, totale et selon la fréquence des mots. Ensuite, il s'agit de comparer la TVO et la TVPh. Afin de répondre à ces objectifs, les questions de recherche suivantes ont été formulées :

1. Quelle est la taille du vocabulaire des apprenants du français langue seconde de secondaire 1 à 5 au Québec et comment évolue-t-elle ?
2. Quelle est la taille du vocabulaire des apprenants du français langue seconde de secondaire 1 à 5 au Québec dans chacune des cinq bandes de fréquence testées et comment évolue-t-elle ?
3. Pour chaque année scolaire, comment la taille du vocabulaire orthographique et phonologique se comparent-t-elles chez ces apprenants ?

Les hypothèses pour chacune des questions de recherches sont les suivantes :

1. La TV croît d'année en année (Milton, 2006) et, en nous basant sur les calculs proposés par Milton et Meara (1998) dans une étude faite auprès d'apprenants britanniques de français langue étrangère et d'apprenants Grecs et Allemands d'anglais langue étrangère (ALE), nous faisons l'hypothèse que les élèves québécois

terminant leurs études secondaires après 11 ans d'instruction dans la langue seconde devraient avoir une TV se situant au-delà de 4 500 familles de mots.

2. Le vocabulaire plus fréquent en français devrait être mieux maîtrisé que le moins fréquent. Ainsi, le nombre de mots connus dans le 1K devrait être plus élevé que celui dans le 2K, et ainsi de suite, suivant le profil lexical de l'apprenant standard élaboré par Meara (1992).

3. Plus un apprenant avance dans le système scolaire, plus la TVO prend d'importance en comparaison à la TVPh, dû à davantage d'expositions au système orthographique de la langue (Milton, 2006).

CHAPITRE III

MÉTHODOLOGIE

Ce chapitre décrira les choix méthodologiques de cette recherche fondamentale, empirique, descriptive et corrélationnelle (Gaudreau, 2011 ; Van der Maren, 1995). Tout d'abord, nous décrirons les participants à l'étude, ensuite les instruments de mesure qui ont servi pour collecter les données ainsi que le déroulement suivi, pour terminer par le traitement des données.

3.1 Participants

Tous les groupes de français d'une école secondaire de Montréal (approximativement 800 élèves) ont été sollicités pour un total de 568 participants potentiels. La sollicitation s'est faite par le biais d'une lettre d'information concernant le projet du mémoire ainsi qu'un formulaire de consentement (voir Annexe A). Les documents ont été distribués par les enseignants de français de l'école. Seuls 239 participants avec un formulaire signé par un adulte responsable ont participé. Afin de pouvoir dresser un portrait global de la taille du vocabulaire chez les apprenants de français langue seconde au secondaire, l'échantillon de cette étude était constitué d'élèves de chaque année scolaire. Les 239 élèves, soit 70 de première secondaire, 36 de deuxième, 41 de troisième, 35 de quatrième et 57 de cinquième, venaient de divers milieux socioéconomiques où une certaine diversité culturelle est présente, à l'image des écoles publiques montréalaises (Armand, 2012). De plus, cette école secondaire anglophone comprend trois programmes d'étude : régulier, immersion et Baccalauréat International (IBO). Les participants provenaient donc de ces trois programmes et, afin d'avoir un échantillon suffisant dans chaque année scolaire, les

données ont été traitées ensemble. Cela a aussi permis de représenter la réalité sociodémographique de l'école le plus fidèlement possible.

Les participants ont rempli un questionnaire sociodémographique (voir Annexe B) qui a permis de voir s'ils répondaient au principal critère de sélection, soit le fait de ne pas communiquer en français à la maison. Seize participants en tout, plus précisément sept participants en secondaire 1, trois en secondaire 2, trois en secondaire 4 et trois en secondaire 5 ont été éliminés de la recherche, car ils ont affirmé communiquer souvent (ou toujours) en français à la maison.

Lors de la cueillette de données, les premiers participants à réaliser les tests ont malheureusement eu un problème technique et leurs données n'ont pu être enregistrées (14 de première secondaire et 13 de troisième secondaire). De plus, certains (n=12) n'ont pas complété le deuxième test (X_Lex), possiblement par manque de temps.

Lors de l'analyse des résultats, plusieurs participants ont montré qu'ils n'avaient pas complété les tests de façon sérieuse puisqu'ils avaient repéré moins de la moitié des pseudomots, voire parfois même répondu qu'ils connaissaient tous les mots du test, sans avoir reconnu un seul pseudomot. Ceci indiquerait que les réponses aux mots réels seraient donc moins fiables (Meara, 1992 ; Beekmans, 2001 ; Eyckmans, 2004 ; Milton, 2006). Par conséquent, 17 participants en secondaire 1, deux en secondaire 2, 20 en secondaire 3 et deux en secondaire 4 ont été éliminés de la recherche. Comme il ne restait que sept participants de secondaire 3 et que nous voulions conserver une taille d'échantillon assez uniforme entre les groupes comparés, nous avons préféré éliminer tout ce groupe.

Suite à l'élimination de ces participants, ce projet de recherche a donc analysé les résultats de 144 élèves du secondaire : 41 en secondaire 1, 30 en secondaire 2, 27 en secondaire 4 et 46 en secondaire 5.

3.2 Instruments de collecte de données

3.2.1 Questionnaire sociodémographique

Avant la passation des tests, les participants ont répondu au questionnaire sociodémographique. Ils devaient, premièrement, indiquer en quelle année scolaire ils étudiaient afin que nous puissions classer les résultats, puis le programme dans lequel ils étaient inscrits afin de nous assurer que chacun de ceux-ci était représenté dans les données collectées. Les participants devaient aussi indiquer depuis combien de temps ils étudiaient le français, pour assurer une certaine uniformité entre les différentes années scolaires. Finalement, l'information la plus importante concernait les langues parlées à la maison et dans le quotidien des élèves. Ils indiquaient la langue parlée avec leur mère, leur père, leurs frères et sœurs, puis avec leurs amis et à quelle fréquence ils utilisaient l'anglais, le français ou une autre langue qu'ils devaient spécifier.

Puisque Montréal est une ville bilingue où le français peut être parlé quotidiennement à l'extérieur de la maison et de l'école, il est difficile de garantir que l'apprentissage du français soit fait uniquement en salle de classe. Toutefois, il fallait s'assurer qu'il était essentiellement fait en milieu scolaire. En effet, dès qu'un participant affirmait échanger souvent (ou toujours) en français avec un parent, ses données étaient exclues.

3.2.2 Tests de taille du vocabulaire

Pour mesurer la taille du vocabulaire orthographique (TVO), le test X_Lex (Meara & Milton, 2003) a été utilisé. Une version française de ce dernier a été obtenue grâce à l'ouvrage de Milton (2009), disponible aussi sur le site de Meara (<http://www.lognostics.co.uk/tools/index.htm>). Il s'agit d'un test oui/non comportant 100 mots et 20 pseudomots auxquels un participant affirme connaître ou non le sens. Les mots y sont lemmatisés, ce qui signifie que c'est leur forme canonique qui est retenue (comme des verbes à l'infinitif présent, des noms singuliers et des adjectifs masculins singuliers). Les lemmes qu'il contient viennent d'une liste sur la fréquence d'utilisation des mots français, élaborée par Baudot (1992). La liste se divise par tranche de 1K mots. Dans X_Lex, il y a 20 mots par tranche de 1K, pour un total de 100 mots. 20 pseudomots sont ensuite intégrés afin d'éviter une surestimation de la TV des participants. Pour chaque énoncé présenté à l'écrit, le participant clique sur « oui » ou « non » avec la souris, selon s'il connaît ou non le sens du mot présenté.

Pour mesurer la taille du vocabulaire phonologique (TVPh), le test A_Lex (Milton & Hopkins, 2005) a été utilisé, qui est une version orale du test X_Lex. Le participant répond ensuite s'il connaît le sens du mot ou non. Il peut le réentendre autant de fois qu'il le veut avant de répondre, ceci dans le souci de fournir les mêmes conditions pour chaque test, puisque dans X_Lex, il est bien sûr possible de lire les mots un nombre illimité de fois (Milton & Hopkins, 2006). Dans les deux cas, il est impossible de revenir en arrière lorsqu'une réponse est déjà donnée. Le choix des mots testés suit le même procédé que pour X_Lex, et ils ont été obtenus sur le site Web Lextutor.ca (Cobb, 2017).

Toutefois, après avoir consulté les listes de lemmes utilisés pour X_Lex et A_Lex, nous avons constaté que les pseudomots semblaient poser problème. Certains étaient

une francisation d'un terme en anglais (ex. : *disabilité*). Nous avons cru important de les éliminer, car ils étaient issus de la L1 des participants et cela aurait pu avoir un impact négatif sur le score final du test. Rappelons que les pseudomots sont censés être rejetés par les participants pour deux raisons. D'une part, ils doivent prouver qu'ils ont fait le test avec sérieux et, d'autre part, montrer qu'ils ont bien distingué les vrais mots qu'ils connaissent de ceux qu'ils ne savent pas. Par conséquent, les pseudomots qui ressemblent à la langue maternelle posent problème, car un participant qui les accepte connaît ces pseudomots en réalité, même s'il s'agit de sa L1. Cela vient alors tout à fait fausser les données. Ainsi, nous avons remplacé les listes de pseudomots dans les tests X_Lex et A_Lex (Annexe C et D). À cause de ces changements, nous avons dû refaire les documents audio de A_Lex, enregistrés par la chercheuse, locutrice native du français québécois. C'est donc à l'aide d'une nouvelle programmation de X_Lex et de A_Lex, disponible sur le Web, que les données ont été collectées tout en conservant les méthodes de calcul utilisées par les créateurs des deux tests (voir section 3.4.1 à la page 30).

3.3 Déroulement

La distribution des formulaires de consentement était faite dans les classes de français afin de s'assurer que les parents acceptaient la participation de leur enfant à l'étude. Ce sont les enseignants de français qui ont accepté de donner du temps de leurs classes pour distribuer ces formulaires, les collecter, puis amener les élèves au laboratoire informatique. Chaque élève a reçu un code d'accès unique, connu d'eux seuls, pour qu'ils ne puissent faire le test plus d'une fois. Alors, après avoir entré le code d'accès, les participants remplissaient le formulaire sociodémographique avant de faire le test phonologique (A_Lex) et le test orthographique (X_Lex). Pour des raisons techniques de programmation, tous les participants ont d'abord complété le test phonologique pour ensuite poursuivre avec le test orthographique. X_Lex et A_Lex comprenaient des mots et des pseudomots différents qui étaient présentés au

hasard pour éviter des résultats liés à l'ordre de présentation. Les participants avaient reçu comme consigne d'indiquer par « oui » ou « non » s'ils connaissaient le sens du mot. Ainsi, ils ignoraient la présence de pseudomots. Il n'y avait aucune limite de temps pour compléter les épreuves, mais une vingtaine de minutes a généralement suffi. À la fin, les participants n'avaient pas accès à leurs résultats.

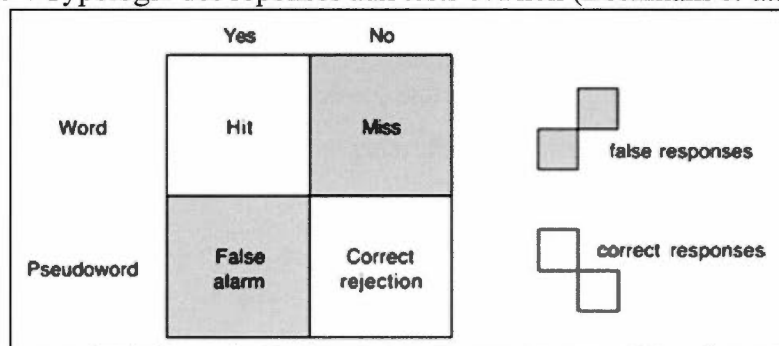
3.4 Traitement des données

Les sections suivantes présenteront certains éléments techniques qui serviront à l'analyse des données. Cela permettra, par le biais de trois analyses de variances, d'analyser les données et de répondre aux trois questions de recherche.

3.4.1 Calcul de la TVO et de la TVPh

La nouvelle programmation faite pour cette recherche a conservé toutes les méthodes de calculs utilisées dans les tests originaux de X_Lex et de A_Lex. En effet, les réponses étaient catégorisées comme montrées dans la figure 4.

Figure 4 Typologie des réponses aux tests oui/non (Beekmans *et al.*, 2001)



Afin de calculer la TVO et la TVPh, chaque réponse positive, ou *hit*, accumulait 50 points. Si un participant répondait qu'il ne connaissait pas un mot, il n'y avait aucun

point accordé. S'il affirmait connaître un pseudomot, 250 points étaient soustraits du total, tandis que s'il répondait à la négative, aucun point n'était accordé ou soustrait. Le score maximal possible était de 5K à chaque test. Le résultat obtenu à X_Lex montre la TVO et A_Lex montre la TVPh.

Pour terminer, comme les deux tests étaient proposés dans une version informatisée, le calcul de la TV se faisait automatiquement et tous les résultats étaient colligés et cryptés dans la base de données *MySQL* (www.mysql.fr), donc inaccessible à moins d'avoir les codes d'accès nécessaires. L'information recueillie était ensuite extraite dans un document *comma-separated values* (.csv) qui est compatible avec Excel. Ce fichier a été importé dans ce logiciel afin d'organiser les données par année scolaire pour l'analyse.

3.4.2 Calcul de la TV

Afin de répondre à la première et à la deuxième question de recherche au sujet de la TV, il était nécessaire de créer un score à partir des résultats aux deux tests, ce qui permettrait d'arriver à un seul résultat montrant la taille du vocabulaire réceptif des participants. En effet, une corrélation significative entre A_Lex et X_Lex confirmerait la validité d'une moyenne des résultats aux deux tests comme calcul de la TV, puisque ceux-ci seraient fortement connectés l'un à l'autre (Milton 2013). Une corrélation de Pearson a montré une corrélation significative entre A_Lex et X_Lex de 0,60 ($p=0,01$). Il existe donc un fort lien entre le vocabulaire orthographique et phonologique chez les participants, et c'est ce résultat composé qui est utilisé pour rapporter les données sur la TV globale.

Maintenant que la méthodologie utilisée a été présentée, il est possible de procéder aux analyses des données.

CHAPITRE IV

RÉSULTATS

Dans ce chapitre, les résultats seront présentés. Tout d'abord, nous analyserons la taille du vocabulaire (TV) chez les participants de cette étude de façon globale, puis par bande de fréquence pour finalement comparer les tailles du vocabulaire orthographique (TVO) et phonologique (TVPh).

4.1 Dresser un portrait de la TV

Les deux sections suivantes répondront au premier objectif qui était de mesurer la TV, totale et selon la fréquence des mots, d'apprenants du français langue seconde de secondaire 1 à 5.

4.1.1 La TV globale

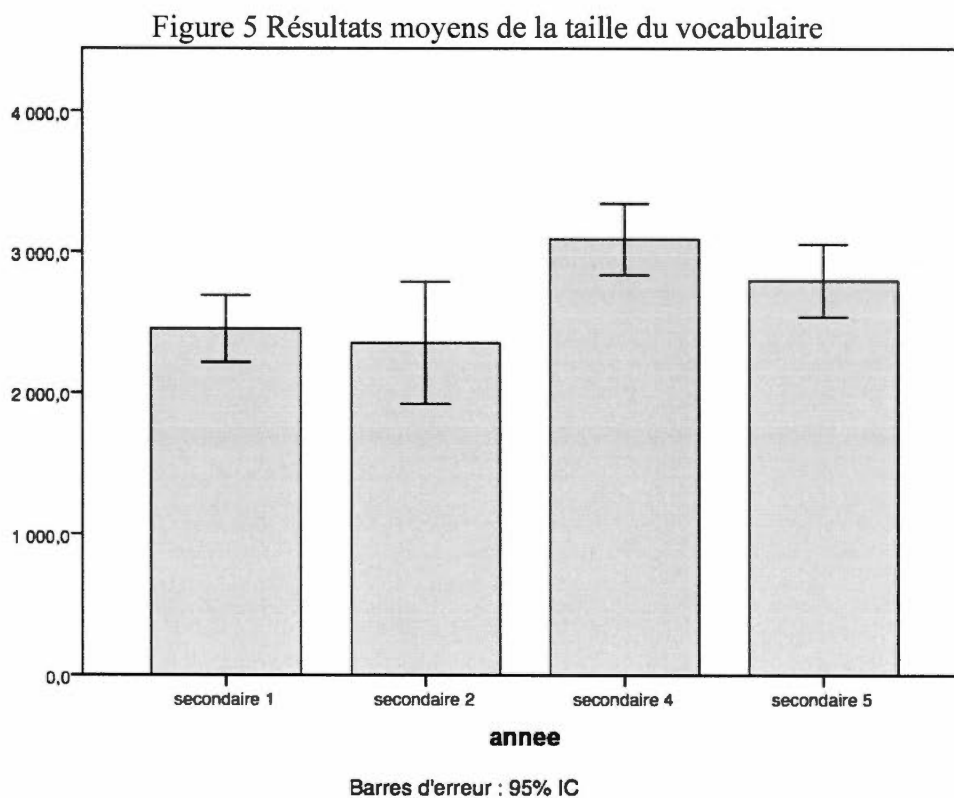
Une première question de recherche a été formulée : « Quelle est la taille du vocabulaire des apprenants du français langue seconde de secondaire 1 à 5 au Québec et comment évolue-t-elle ? » Dans ce qui suit, une première analyse par année scolaire sera effectuée, puis les cycles un et deux seront comparés.

Les résultats descriptifs pour chaque année scolaire sont présentés dans le tableau 3 ci-dessous et la figure 5 qui suit nous permet de constater les intervalles de confiance

à 95 % qui montrent la marge d'erreur dans les résultats pour chaque groupe. Le score maximum possible est de 5 000.

Tableau 3 Résultats moyens de la taille du vocabulaire

Année	N	Moyenne	Écart type	Min	Max
1	41	2 454	749	750	4 000
2	30	2 353	1 159	100	4 025
4	27	3 088	638	1 600	4 475
5	46	2 796	869	1 025	4 200



Pour comprendre si la TV des participants était significativement différente d'une année scolaire à l'autre, une analyse de variance (ANOVA) unidirectionnelle ayant comme facteur indépendant l'année scolaire et comme facteur dépendant la TV a été effectuée. Cette analyse paramétrique a été faite, car les données étaient normalement distribuées et l'homogénéité des variances était respectée. Les tests d'effets intersujets montraient une différence significative entre les années ($F(3,140) = 4.563$; $p = 0.004$) Donc, des tests *post hoc* ont ensuite pu être effectués. Pour diminuer la possibilité d'erreur de type I, la correction *False Detection Rate* a été appliquée pour chaque analyse (Benjamini & Hochberg, 1995), selon les recommandations en didactique des langues secondes (Larson-Hall, 2010). Tous les résultats de ce chapitre sont significatifs après son application.

Les tests *post hoc* de l'ANOVA ont montré que le groupe de secondaire 1 était significativement différent du secondaire 4 ($p=0,004$) et que le secondaire 2 était significativement différent du secondaire 4 et 5 (respectivement $p=0,002$ et $p=0,032$).

Les données par cycle ont ensuite été analysées, puisque la structure du système scolaire est organisée ainsi. Le résultat au test t a montré une différence significative entre les résultats du cycle 1 et du cycle 2 : $t(142) = -3,394$, $p = 0,001$. Le tableau suivant résume les résultats obtenus.

Tableau 4 Résultats moyens de la taille du vocabulaire par cycle

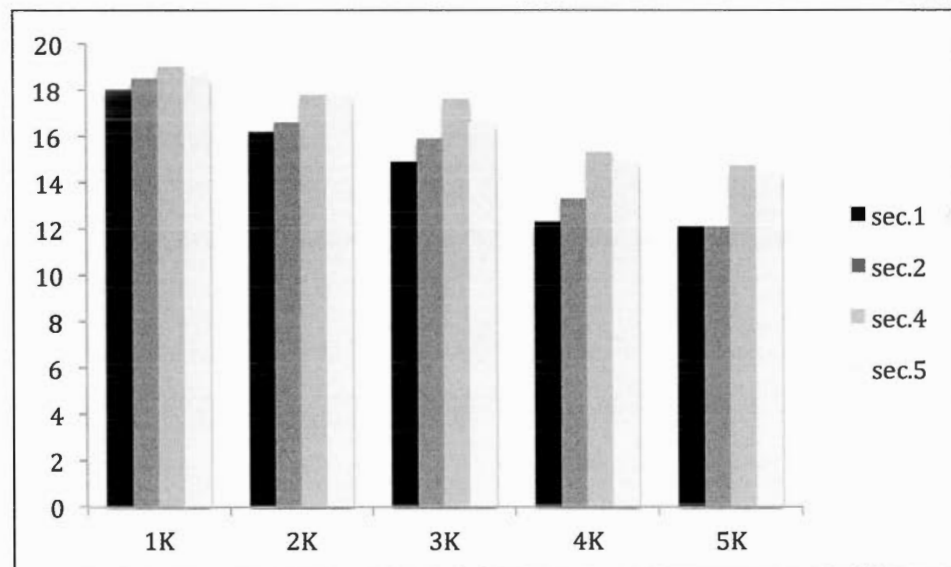
	N	Moyenne	Écart type	t	Valeur p
Cycle 1	71	2411,62	938,50	-3,394	,001
Cycle 2	73	2904,45	800,17		

Pour terminer, la première hypothèse, qui était que la TV croît d'année en année, n'est pas confirmée. Par contre, il est possible de dire que d'un cycle à l'autre, le vocabulaire croît significativement.

4.1.2 La TV par bande de fréquence

Afin de mesurer la TV par bande de fréquence, une deuxième question de recherche a été formulée : « Pour chaque année scolaire, quelle est la taille du vocabulaire des apprenants de secondaire 1 à 5 dans chacune des cinq bandes de fréquence testées et comment évolue-t-elle ? » Pour y répondre, nous avons analysé les résultats des participants pour chaque tranche de 1K mots. Une courbe normale montrerait que le 1K mots est davantage maîtrisé que le 2K, qui est plus maîtrisé que le 3K, et ainsi de suite, tel qu'illustré par un profil de la fréquence lexicale standard (Meara, 1992). Cette tendance a pu être observée, comme le montre le tableau 5.

Tableau 5 Taille du vocabulaire selon la bande la fréquence



La mesure de chaque bande de fréquence se fait sur un total de 20, puisque les tests comprenaient 20 mots dans la tranche de 1K, 20 dans la tranche de 2K, et ainsi de suite. Les résultats pour chaque secondaire, dans chaque bande de fréquence, sont présentés dans le tableau 6. Le tableau 7 présente ensuite l'effet global significatif de chaque bande de fréquence.

Tableau 6 Taille du vocabulaire par secondaire, selon la bande de fréquence

		sec.1	sec.2	sec.4	sec.5
1K	Moyenne	18	18,5	19	18,6
	Écart type	1,7	1,8	1,3	2,3
2K	Moyenne	16,2	16,6	17,8	17,7
	Écart type	2,0	2,4	2,2	2,8
3K	Moyenne	14,9	15,9	17,6	16,6
	Écart type	2,6	3,2	2,4	3,1
4K	Moyenne	12,3	13,3	15,3	14,9
	Écart type	2,7	3,2	3,1	3,2
5K	Moyenne	12,1	13,6	14,7	14,5
	Écart type	2,8	2,8	2,6	2,9

Tableau 7 Effet global de chaque bande de fréquence

Bande de fréquence	DDL	F	<i>p</i>
1K	3, 140	1,541	0,207 (n.s.)
2K	3, 140	4,029	0,009
3K	3, 140	5,150	0,002
4K	3, 140	7,556	0,000
5K	3, 140	6,889	0,000

Les tests *post hoc* de l'ANOVA ont ensuite montré qu'il y avait une différence significative dans la TV par bande de fréquence selon les années scolaires. Les résultats sont rapportés dans le tableau 8.

Tableau 8 Comparaison entre les années scolaires par bande de fréquence

Bande de fréquence	Année (I)	Année (J)	Différence moyenne (I-J)	Sig.	Intervalle de confiance à 95%	
					Borne inférieure	Borne supérieure
1K	Aucune différence entre les années					
2K	Sec.1	Sec.4	-1,5763	,010	-2,762	-,391
		Sec.5	-1,5453	,003	-2,573	-,518
3K	Sec.1	Sec.4	-2,6472	,000	-4,058	-1,237
		Sec.5	-1,6710	,003	-2,893	-,449
	Sec.2	Sec.4	-1,7241	,026	-3,234	-,214
4K	Sec.1	Sec.4	-2,9734	,000	-4,463	-1,483
		Sec.5	-2,5498	,000	-3,841	-1,259
	Sec.2	Sec.4	-2,0148	,014	-3,610	-,420
		Sec.5	-1,5913	,027	-3,002	-,180
5K	Sec.1	Sec.2	-1,5191	,025	-2,845	-,193
		Sec.4	-2,5876	,000	-3,955	-1,220
		Sec.5	-2,4024	,000	-3,588	-1,217

Pour terminer, il est possible de confirmer notre deuxième hypothèse selon laquelle le vocabulaire plus fréquent en français devrait être mieux maîtrisé que le moins fréquent. En effet, les premières 2K familles de mots sont maîtrisées à 80 % ou plus,

et ce, à tous les secondaires. De plus, il est possible d'ajouter que les lemmes les moins fréquents (4-5K) sont significativement moins connus par les participants de secondaire 1 et 2 en comparaison à ceux de secondaire 4 et 5.

4.2 Comparaison entre la TVO et la TVPh

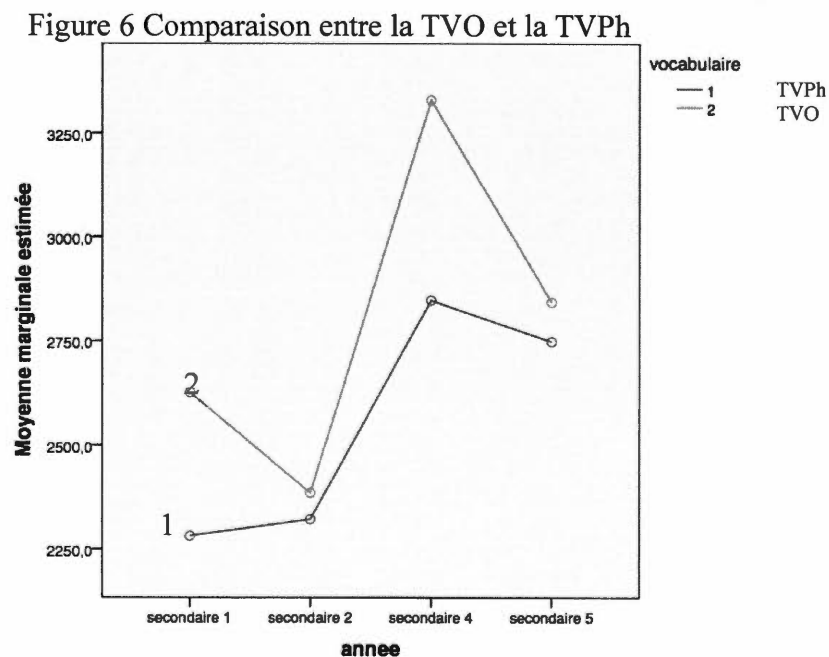
Notre deuxième objectif de recherche était de comparer la TVO et TVPh pour chaque année scolaire. Nous avons avancé l'hypothèse que, plus un apprenant avance en âge, plus il est exposé au système écrit de la langue, plus sa TVO est importante en comparaison à sa TVPh (Milton & Hopkins, 2006). Tout d'abord, nous présentons les statistiques descriptives des tests orthographique et phonologique.

Tableau 9 Résultats moyens de la TVO (X_Lex) et de la TVPh (A_Lex)

	Année	Moyenne	Erreur type	N
A_Lex	Sec.1	2281,7	910,96	41
	Sec.2	2321,66	1081,48	30
	Sec.4	2848,15	667,44	27
	Sec.5	2748,91	1006,89	46
	Total	2545,49	964,58	144
X_Lex	Sec.1	2626,83	755,41	41
	Sec.2	2385,0	1371,33	30
	Sec.4	3329,48	772,39	27
	Sec.5	2843,48	1056,34	46
	Total	2777,43	1048,04	144

Les deux tests de mesure de vocabulaire X_Lex et A_Lex sondent les mêmes participants et il est de mise de s'interroger si les résultats sont significativement différents. Alors, pour arriver à comparer la TVO (X_Lex) et la TVPh (A_Lex), une ANOVA à mesure répétée a montré que, globalement, elles sont significativement différentes l'une de l'autre ($F(1,140)=10.557 ; p=0.001$).

Ensuite, nous avons examiné si les différences entre X_Lex et A_Lex étaient significatives pour chaque année. Les tests intrasujets ont plutôt montré que l'interaction entre l'année et le type de vocabulaire était non significative ($F(3,140)=1.661 ; p=0.178$). La combinaison de cette interaction non-significative et de l'effet significatif du type de vocabulaire signifie que la différence entre la TVO et la TVPh est significative, peu importe l'année. Cela mène donc à la conclusion que, de secondaire 1 à 5, la TVO est toujours plus élevée que la TVPh, tel qu'illustré par la figure 6.



Selon l'hypothèse de départ, plus un apprenant est exposé au système écrit de la langue, plus sa TVO est importante en comparaison à sa TVPh. Or, cette hypothèse ne peut se confirmer avec les résultats obtenus dans cette recherche, puisque l'écart entre la TVO et la TVPh ne croît pas dans le temps. L'écart est même très variable. La différence entre la TVO et la TVPh est de 345 familles de mots en secondaire 1, de 64 familles de mots en secondaire 2, 481 familles de mots en secondaire 4 et de 95 familles de mots en secondaire 5. Alors, pour terminer, il est possible d'affirmer que la TVO est toujours plus élevée que la TVPh chez les participants à cette étude, peu importe l'année scolaire.

Dans ce chapitre, nous avons rapporté les résultats obtenus suite à l'analyse des données recueillies afin de répondre aux trois questions de recherches formulées dans cette étude. Il s'agissait tout d'abord de mesurer la TV globale et par bande de fréquence des apprenants du français langue seconde de secondaire 1 à 5 pour voir s'il y avait une évolution d'une année à l'autre, puis d'un cycle à l'autre. Ensuite, la troisième analyse était de comparer la TVO et la TVPh chez ces apprenants. Dans le chapitre suivant, nous pourrons donc interpréter ces résultats.

CHAPITRE V

DISCUSSION

Dans ce chapitre, il sera avant tout question d'interpréter les résultats aux trois questions de recherche afin de répondre aux deux objectifs de cette étude, d'en exposer les limites, de relever les implications didactiques, pour finalement proposer des pistes de réflexion pour les recherches futures.

5.1 La TV globale

Il est essentiel, tout d'abord, de rappeler la première question de recherche : « Quelle est la taille du vocabulaire chez des apprenants du français langue seconde de secondaire 1 à 5 au Québec et comment évolue-t-elle ? » L'hypothèse de base était qu'il y aurait croissance de la taille du vocabulaire (TV) d'une année à l'autre, mais aussi que, selon les estimations de Milton et Meara (1998), les participants devraient avoir, en secondaire 5, une taille de vocabulaire se situant au-delà de 4 500 familles de mots.

Premièrement, les premiers calculs de la TV ont montré qu'il n'y a pas eu de progression constante entre les années, de secondaire 1 à 5. En effet, nous avons constaté que la TV moyenne des participants de première secondaire était significativement inférieure à celle de quatrième secondaire. De plus, la TV moyenne des apprenants de deuxième secondaire était significativement inférieure à celle de quatrième et cinquième secondaire. Par contre, la TV entre le secondaire un et deux et entre le secondaire quatre et cinq ne présentaient pas de différence significative. Le manque de progression d'année en année semble d'abord s'expliquer par les

différences individuelles. En effet, les écarts types sont grands : en secondaire 1, il était de 749, en secondaire 2, de 1 159, en secondaire 4, de 638 et en secondaire 5, de 869, alors que le score maximal possible était de 5 000, tel que montré dans le tableau 3 (page 38). De plus, les TV minimum et maximum de chaque année sont révélatrices des grandes différences à l'intérieur de chaque groupe de participants. Rappelons que Milton (2006) a avancé l'hypothèse que la progression irrégulière de ses participants était due au fait que les apprenants passaient beaucoup plus de temps à l'étude du français dans les deux dernières années de l'école secondaire, ce qui résultait en une augmentation notable de la TV en fin de parcours scolaire. Toutefois, cette interprétation n'est pas valable pour notre échantillon puisque le nombre d'heures de FLS diminue dans les deux dernières années du secondaire. En secondaire 1, 2 et 3, les élèves étudient 150 heures de français par année et 100 heures en secondaire 4 et 5 (Régime pédagogique de l'éducation préscolaire, de l'enseignement primaire et de l'enseignement secondaire, 2017).

D'un point de vue méthodologique, et ce malgré le fait qu'il n'y ait pas de progression entre les secondaires 1 et 2 puis 4 et 5, l'absence de données en secondaire 3 semble causer un bris dans la continuité que l'on voulait préalablement analyser. Cela pourrait aussi être dû à la comparaison entre des échantillons trop petits pour pouvoir détecter une différence (David, 2008). C'est pourquoi une analyse entre les cycles d'apprentissages au secondaire semblait essentielle.

Le système scolaire québécois se divise par cycle, tant au primaire qu'au secondaire. Rappelons qu'il y en a deux au secondaire : les secondaires un et deux forment le premier et les secondaires trois à cinq forment le deuxième. En effet, la structure par cycle d'apprentissage viendrait, selon le MELS : « soutenir la continuité des apprentissages. Il permet de mieux assurer la progression différenciée des apprentissages (...) » (2007, p. 11). Dans la présente recherche, une différence significative entre le premier et le deuxième cycle a été obtenue. Au premier cycle

(secondaire 1 et 2), la TV moyenne se situe sous les 2 500 familles de mots maîtrisées, tandis qu'au deuxième cycle (secondaire 4 et 5), les participants se rapprochent ou atteignent une TV moyenne de 3K familles de mots. Cette différence significative montre bien que l'interprétation de David (2008) se défend et il serait intéressant, pour des recherches futures qui désireraient analyser la croissance du vocabulaire d'une année à l'autre, d'avoir un échantillon plus important pour chaque année étudiée. De plus, si les différences individuelles entre les participants de chaque groupe viennent compliquer les analyses entre les années, l'analyse par cycle montre tout de même une progression significative.

Nous avons supposé que la TV des participants de secondaire 5 doit atteindre (ou même dépasser) les 4 500 mots, mais notre hypothèse ne s'est pas confirmée. La TV maximum a atteint 4 200 en secondaire 5. Il semble difficile d'expliquer une si faible performance, surtout si l'on considère que le FLS est appris depuis la première année du primaire au Québec, soit depuis 11 ans. À Montréal, les 75 écoles primaires des commissions scolaires *English Montreal School Board* (2015) et *Lester B. Pearson School Board* (2017) offrent des programmes d'immersion où de 30 à 100 % des matières scolaires sont enseignées en français. À l'école secondaire seulement, l'étude obligatoire du FLS totalise 650 heures (Régime pédagogique de l'éducation préscolaire, de l'enseignement primaire et de l'enseignement secondaire, 2017). Les programmes d'immersion sont aussi présents au secondaire où des matières comme les sciences sociales et les arts sont enseignées en français. Bref, les heures consacrées à l'étude du FLS sont nombreuses et il semble ardu de justifier une TV si peu élevée chez les participants de l'étude. De plus, le contexte dans lequel le français est appris au Québec devrait être un grand avantage pour les apprenants : comme il s'agit officiellement d'une province unilingue francophone, les occasions d'interagir en français à l'extérieur du contexte scolaire sont omniprésentes.

Les résultats obtenus dans cette recherche nous poussent à remettre en question, d'une part, l'enseignement et l'apprentissage du français dans le système scolaire et, d'autre part, les perceptions et les croyances que les apprenants anglophones ont de la langue française et de son apprentissage.

Comme les données manquent au sujet de l'enseignement et de l'apprentissage du vocabulaire en FLS dans le système québécois, ce sont celles en ALS qui peuvent potentiellement expliquer pourquoi il y aurait cette « pauvreté lexicale » à la fin des études secondaires (quoique ces données manquent aussi). Oliveira (2012) a analysé 3 manuels scolaires et cahiers d'exercices utilisés au deuxième cycle du secondaire en ALS. La chercheuse a montré que seuls les mots fréquents du 1K étaient complètement présents dans les manuels et cahiers étudiés et que ce vocabulaire était suffisamment recyclé. Le recyclage du vocabulaire est primordial pour son apprentissage, puisqu'un mot devait être rencontré au moins dix fois afin d'être considéré comme pleinement maîtrisé (Webb, 2007). Les mots des 2K et 3K étaient présents et recyclés, mais pas suffisamment pour l'apprentissage. Quant au vocabulaire à fréquence intermédiaire des tranches de 4K à 9K, il est peu représenté et peu recyclé. Il n'y a aucune raison de croire que la situation soit meilleure dans les manuels scolaires en FLS. Donc, il se peut que les élèves n'aient tout simplement pas assez d'occasions de lire ces mots à fréquence intermédiaire, pour qu'ils fassent partie de leurs connaissances lexicales.

Jean et Simard (2011) ont étudié les perceptions des apprenants et des enseignants envers l'apprentissage et l'enseignement de la grammaire en FLS et ALS, et elles ont rapporté qu'elle était perçue comme un facteur de démotivation important dans l'apprentissage de la L2. L'enseignement du vocabulaire n'était pas abordé à proprement dit, mais il est essentiel de mentionner que la motivation à apprendre une

langue seconde, ou même, plus précisément, d'apprendre du nouveau vocabulaire dans la langue seconde, peut être inhibée ou altérée par différents facteurs. Il pourrait être intéressant de s'interroger sur la perception des apprenants et des enseignants face à un enseignement explicite du lexique, tout comme sur les pratiques des enseignants. Par exemple, est-ce qu'apprendre des listes de mots est motivant pour des jeunes de 12 à 17 ans ? Est-ce que l'apprentissage fortuit du vocabulaire est suffisant ? Une chose est claire, pour apprendre, la motivation est un facteur clé en langue seconde (Laufer & Hulstijn, 2001).

En résumé, les données recueillies dans cette étude n'ont pas permis de confirmer que la TV augmente d'une année à l'autre, mais ont montré une progression significative du premier au deuxième cycle. De plus, il n'a pas été possible de démontrer que les participants de secondaire 5 avaient une TV se situant au-delà des 4 500 familles de mots. Il semble pertinent d'intégrer différentes méthodologies afin de mesurer la TV et ainsi vérifier si le vocabulaire plus fréquent est maîtrisé en profondeur. Le fait de l'analyser plus spécifiquement par bande de fréquence peut toutefois ajouter une dimension non négligeable pour les enseignants et les apprenants.

5.2 La TV selon les bandes de fréquence

Afin de mesurer la TV selon les bandes de fréquence, une deuxième question de recherche a été formulée : « Pour chaque année scolaire, quelle est la taille du vocabulaire des apprenants de secondaire 1 à 5 dans chacune des cinq bandes de fréquence testées et comment évolue-t-elle ? » L'hypothèse de base était que les mots les plus fréquents (1K-2K) en français sont davantage connus par les participants que les moins fréquents.

Les résultats montrent que le profil lexical standard de l'apprenant de Meara (1992) représente bien les participants de cette étude, c'est-à-dire qu'ils connaissent davantage de mots dans le 1K que dans le 2K, en connaissent plus dans le 2K que dans le 3K, et ainsi de suite. Cette tendance est observée de secondaire 1 à 5. Nos résultats rejoignent ceux de Milton (2006) et de Richard *et al.* (2008) qui précisait que la fréquence joue un rôle déterminant dans l'apprentissage d'une langue. De plus, il est encourageant que les mots les plus fréquents soient connus en premier, surtout qu'en français, les premiers 2K composent 85 % d'un texte oral de base comme une conversation de même que 85 % d'un texte académique écrit (Nation & Meara, 2002 ; Cobb & Horst, 2004).

Le vocabulaire fréquent est maîtrisé chez tous les participants de cette étude, et cela est encourageant. Mais, pourquoi le vocabulaire moins fréquent (3K et plus) ne l'est pas après tant d'années d'étude ? Il serait possible d'affirmer que, puisque les mots moins fréquents ne sont pas à proprement dit enseignés, mais plutôt rencontrés « au gré de leurs [les élèves] interactions et de leurs lectures » (Progression des apprentissages, MELS, p. 10) la rétention n'a possiblement pas lieu. Dans cette étude, ce n'est pas la rétention du vocabulaire que nous avons mesuré, par contre la reconnaissance du vocabulaire réceptif que mesurent X_Lex et A_Lex est la première étape vers une rétention durable (Laufer & Goldstein, 2004). De plus, les premiers 1K et 2K mots sont hautement polysémiques (Tidball & Treffers-Daller, 2008) et il est ainsi possible d'exprimer beaucoup sans à avoir recours à du vocabulaire moins fréquent (Crossley *et al.*, 2013). Notamment, le verbe « peser » fait partie de la tranche de 2K, a été testé dans A_Lex et comprend 4 différents sens : « mesurer le poids », « évaluer attentivement », « appuyer » et « être difficile à supporter ». De plus, ces définitions ne comprennent pas d'expressions figées telles que « peser lourd dans la balance » qui signifie « avoir une grande influence ». Au cours des années à l'école secondaire, l'apprentissage du français langue seconde vise peut-être à

approfondir les connaissances du vocabulaire fréquent en maîtrisant davantage les cooccurrences, la polysémie, les expressions figées d'un verbe par exemple. La progression des apprentissages prévue au programme de FLS de base précise d'ailleurs que la compréhension et la production d'expressions figées devraient être comprises et utilisées en secondaire 4, de même que des relations sémantiques comme la polysémie et la collocation en cinquième secondaire. Par contre, le document du MELS ne précise pas la nature des mots enseignés, encore moins leur fréquence, et ne suggère pas non plus de moyens pour enseigner le vocabulaire explicitement. La TV aurait donc avantage à être mesurée tant en profondeur qu'en étendue afin de vérifier le degré de maîtrise du vocabulaire fréquent par année scolaire. Nation (2007) disait d'ailleurs que, pour mieux comprendre les connaissances des apprenants, plusieurs types de tests devaient être utilisés.

L'analyse des résultats par bande de fréquence permet aussi d'observer une progression que la TV globale n'avait pas pu démontrer. En effet, les tests *post hoc* de l'ANOVA à mesures répétées ont montré des différences entre les années scolaires pour chacune des bandes de fréquence. Pour le 1K, il n'y a aucune différence entre les années scolaires, ce qui indique que les premiers mille mots sont connus de tous les secondaires, sans différence notable. Cela signifie que 75 % (Cobb & Horst, 2004) d'une conversation de base devrait être compréhensible par tous les participants. Le 2K présentait une différence significative entre le groupe de secondaire 1, qui maîtrisait significativement moins ce vocabulaire que les groupes de secondaire 4 et 5, les apprenants les plus expérimentés. Cela signifierait que 85 % des textes académiques seraient à la portée des participants de secondaire 4 et 5. Dans les bandes 3K et 4K, la progression est plus visible puisque les groupes du cycle 1 connaissent significativement moins ces familles de mot que les groupes de quatrième et de cinquième secondaire. En effet, le 3K est maîtrisé au premier cycle entre 74,5 % et 79,5 % et au deuxième cycle entre 83 % et 88 %. Pour le 4K, le premier cycle connaît entre 61,5 % et 66,5 % des mots, tandis que le deuxième cycle

se situe plutôt entre 74,5 % et 76,5 %. Et, finalement, pour le 5K, le groupe de secondaire 1 est différent de tous les autres. Ce dernier résultat ne concorde pas avec une progression claire dans la maîtrise du vocabulaire moins fréquent, puisque le groupe de secondaire 2 n'aurait pas dû avoir un score supérieur au groupe de secondaire 1 étant donné qu'il n'était pas significativement meilleur avec les mots de 4K. Cela pourrait s'expliquer par le fait que les données ne montrent pas toujours une uniformité, comme démontré par les écarts types élevés des TV moyennes.

Le plus stimulant dans ces résultats est le fait qu'une progression est observable de façon significative. Plus un apprenant avance dans son parcours scolaire, plus il maîtrise des mots moins fréquents, et ce, de façon significative en comparaison avec les plus jeunes. Les connaissances que cela apporte au domaine sont intéressantes puisqu'à l'aide de cette analyse, il devrait être possible de fixer des objectifs d'apprentissage et d'enseignement clairs. Les études précédentes qui avaient analysé les données de la TV par bande de fréquence montraient l'importance de maîtriser les 2K familles de mots les plus fréquentes, généralement acquises par leur enseignement ciblé et explicite (Nation, 2011 ; Schmitt, 2011 ; Staehr, 2008 ; Milton, 2006 ; Cobb & Horst, 2004 ; Meara, 1995). Alors, s'il est essentiel d'enseigner les mots les plus fréquents, qu'en est-il du reste ? Selon Schmitt et Schmitt (2014), le vocabulaire qu'ils nomment « à fréquence intermédiaire », de 3K à 9K, devrait aussi être enseigné puisqu'il serait dangereux de ne se fier qu'à l'apprentissage fortuit du nouveau vocabulaire, qui doit être rencontré plus de 10 fois afin d'en assurer la pleine maîtrise (Webb, 2007). Tout d'abord, les apprenants de L2 ne liraient pas suffisamment pour assurer une rétention du vocabulaire moins fréquent, car le contact avec ces nouveaux mots est évidemment moins fréquent (Cobb, 2007). De plus, les manuels scolaires n'assureraient pas un recyclage des mots de 3K et plus (Oliveira, 2012), ce qui rendrait difficile une exposition suffisante aux nouveaux mots appris en contexte scolaire. L'enseignement explicite et ciblé du vocabulaire reste donc une solution et, en effet, l'acquisition du vocabulaire serait avantagée lorsqu'appris dans des tâches

plutôt que par la lecture uniquement (Laufer & Hulstijn, 2001 ; Laufer, 2009). Il ne sera pas une mince tâche pour les enseignants d'assurer une récurrence dans l'exposition au vocabulaire introduit en classe, mais de savoir à quelle bande de fréquence l'enseignement du vocabulaire doit débiter semble un excellent point de départ.

En terminant, il est possible de confirmer l'hypothèse voulant que les participants de cette recherche maîtrisent davantage le vocabulaire le plus fréquent que le moins fréquent en français. La TV, de même que le détail de celle-ci par bande de fréquence, a été calculée à l'aide de deux tests combinés, X_Lex pour la TVO et A_Lex pour la TVPh. Les résultats à chacun de ces tests seront discutés dans la section suivante.

5.3 Le vocabulaire orthographique et phonologique

Pour comparer la TVO et la TVPh des participants, une troisième question de recherche a été posée : « Pour chaque année scolaire, comment la taille du vocabulaire orthographique et phonologique se comparent-elles chez ces apprenants ? » Rappelons que l'hypothèse de base était que, plus un apprenant avance dans le système scolaire, plus sa TVO prend d'importance en comparaison à sa TVPh (Milton, 2006).

Les résultats à la question de recherche ont été clairs : chez nos participants, la TVO est toujours plus élevée que la TVPh, peu importe l'année scolaire. Il y a peu d'études qui ont mesuré puis comparé la TVO et la TVPh. Les études existantes se sont intéressées à l'anglais langue étrangère chez des locuteurs de l'arabe et du grec. Toutefois, il est tout de même possible de dégager une tendance des travaux de Milton (Milton & Hopkins 2006; Milton, Wade & Hopkins, 2010) qui serait similaire

à nos résultats. Dans chaque étude, la TVPh était moins élevée que la TVO, sauf chez les participants de l'école élémentaire qui avaient une TVPh plus élevée que la TVO, possiblement parce que l'expérience des jeunes avec le système écrit de la langue était moins importante que chez les adultes. Une TVO plus grande s'expliquait aussi par le fait que c'est à l'écrit que le vocabulaire moins fréquent est utilisé, et non à l'oral. Finalement, la différence entre la TVO et la TVPh montrait surtout que les participants étaient exposés à différents types d'enseignements, que les différentes L1 jouaient un rôle important tout comme les différences individuelles.

Alors, il est important de considérer différentes raisons qui expliqueraient pourquoi la TVO est toujours plus élevée que la TVPh. Tout d'abord, lorsque l'on observe l'orthographe de certains mots en anglais et en français, plusieurs sont dits « transparents » pour les apprenants, c'est-à-dire facilement reconnaissables grâce à leur orthographe. Ce sont des « congénères » s'ils partagent le même sens dans les deux langues, mais de « faux amis » s'ils ne partagent que l'orthographe et non le sens (Séguin & Tréville, 1992). Une attention particulière a été portée aux congénères dans la liste de mots testés afin de, bien sûr, les éviter. Toutefois, la proximité des systèmes orthographiques du français et de l'anglais aurait pu favoriser la connaissance du système orthographique du français, donc la TVO. La même réflexion s'applique difficilement pour les systèmes phonologiques de l'anglais et du français. Par exemple, la prononciation d'un mot qui aurait une orthographe transparente dans les deux langues peut s'avérer grandement différente d'une langue à l'autre. La prononciation du mot /teɪbl/ en anglais se prononce plutôt /tabl/ en français. Un mot fréquent comme /ɪm'pɔ:tənt/ se prononce /ɛ̃pɔ̃tɑ̃/ en français, et pourrait ne pas avoir été reconnu par un participant. Pourtant, ces deux mots s'écrivent exactement de la même façon dans les deux langues et la liste d'exemples qui suit ce modèle est assez longue. Bref, la proximité des systèmes orthographiques de l'anglais et du français a pu augmenter le nombre de mots reconnus à l'écrit et favoriser la TVO au détriment de la TVPh.

Le fait que la TVPh soit moins élevée que la TVO montre peut-être un reflet des besoins lexicaux pour la communication orale en opposition à la communication écrite. En effet, il faut savoir que moins de mots sont requis pour comprendre une conversation à l'oral (VanZeeland & Schmitt, 2012). Ces chercheurs ont conclu que seulement 2-3K étaient nécessaires pour une compréhension suffisante. Ces 2-3K correspondent à 95 % de couverture d'un texte oral, ce qui semble suffisant en comparaison au 98 % de couverture requis ou aux 6-7K familles de mots auparavant établi par Nation (2006). Le système non verbal et le contexte viendraient pallier aux incompréhensions en situation authentique (Milton, Wade & Hopkins, 2010). Bref, les recherches fournissent la meilleure excuse pour nos apprenants: ils n'ont pas besoin d'une taille de vocabulaire élevée puisque 2-3K familles de mots leur permet de fonctionner adéquatement à l'oral. Ce serait « la loi du moindre effort » (*traduction libre*, Crossley *et al.*, 2013, p.966) en quelque sorte. Il est nécessaire ici de spécifier que ces données concernent l'anglais, et que peu de recherches ont montré le pourcentage de couverture nécessaire en français pour une compréhension optimale, à l'écrit ou à l'oral. Nous savons déjà qu'avec 2K familles de mot en français, nous obtenons 85 % de couverture, tant d'un texte oral que d'un texte académique (Cobb & Horst, 2004). Mais, il serait essentiel de connaître davantage de données pour le français afin de mieux cibler le vocabulaire à enseigner explicitement et de fournir un input compréhensible en salle de classe (Cobb, 2007).

Pour terminer, une TVO plus élevée qu'une TVPh pourrait être expliquée par le type d'enseignement reçu. Il peut sembler contradictoire que la TVO soit plus élevée étant donné que le programme d'enseignement du FLS est plutôt communicatif, mais Simard et Jean (2011) ont montré que plusieurs études soulignaient le fait que les apprenants du FLS recevaient un enseignement plus traditionnel. Dans cette étude, les chercheuses ont relevé que 43 % des interventions des enseignants du FLS au

secondaire concernaient la forme de la langue dont 38 % concernaient le vocabulaire. Bref, l'attention portée au vocabulaire en classe est notable, mais encore faut-il savoir quoi, quand et comment l'enseigner, et que ça le soit tant à l'oral qu'à l'écrit.

5.4 Les limites

Comme la plupart des recherches, il y a des limites qui doivent être mises en lumière. La première serait que les données du groupe de secondaire trois ont dû être éliminées. En effet, comme les objectifs de cette étude étaient directement liés à la progression de la TV dans le système scolaire, le fait d'enlever tout un groupe affecte nos résultats et nos interprétations, puisqu'il n'est pas possible d'analyser chaque année du secondaire.

La deuxième limite observée est la taille réduite de l'échantillon. À cause de réponses aléatoires aux tests, nous avons dû rejeter des participants dans chaque groupe scolaire, ce qui a réduit de façon notable la taille de l'échantillon. De plus, des problèmes techniques nous ont amenés à retrancher davantage de données puisque, lorsque les participants ont commencé à faire les tests en ligne, le système n'était pas suffisamment stable pour enregistrer tant de résultats en même temps. Malgré le fait que la taille totale de l'échantillon final ($n=144$) semble suffisante, le fait de comparer de petits groupes de chaque secondaire entre eux n'a pas permis de montrer de différence significative, tout comme dans l'étude de David (2008). Il serait important de voir à ce que la technique fonctionne parfaitement, mais aussi que les participants comprennent davantage le but de l'étude. Enfin, un échantillon plus grand aurait pu permettre d'établir davantage de liens entre chaque année du secondaire et de possiblement corrélérer le nombre d'années d'étude à la TV.

5.5 Les implications didactiques

Les résultats de la présente recherche permettent de faire quelques propositions afin d'améliorer l'apprentissage et l'enseignement du vocabulaire en français langue seconde, et ce, malgré les limites identifiées précédemment.

Pour commencer, le fait qu'il manque tant de données dans la mesure du vocabulaire en français langue seconde devrait être soulevée au niveau ministériel. En effet, si le Programme de formation de l'école québécoise n'a pas fixé de vocabulaire à apprendre en termes de quantité, il a tout de même décrit de façon très précise les attentes en fin de deuxième cycle au secondaire, soit la conclusion des études obligatoires au Québec (MELS, 2007). Ces attentes nous semblent élevées, du moins surtout lorsqu'on ne sait pas à combien de mots ce genre de compétence langagière s'applique et que, selon les résultats obtenus dans cette recherche, les participants semblent loin de la description faite dans le programme de formation puisque la TV moyenne en secondaire cinq se situe sous les 3 000 familles de mots. Cela correspond en anglais (Nation, 2006) à une TV suffisante pour comprendre une conversation de base seulement, et nous pouvons supposer que ces données soient comparables en français, quoique toujours inconnues. De plus, même si 2K équivaut à 85% de couverture de texte en français, cela ne permet pas une compréhension suffisante des textes puisque 95 à 98 % des mots doivent être connus pour avoir une compréhension suffisante (Nation, 2006). Rappelons que 95 % des mots d'un texte correspond à un mot sur 30 qui demeure inconnu. Alors, dans le souci de fournir aux enseignants et aux apprenants des objectifs spécifiques à atteindre, le ministère aurait avantage à préciser ses attentes en terme d'étendue du vocabulaire. De plus, le MELS a conçu, en français langue maternelle, une liste de mots à orthographier correctement pour les élèves à l'école primaire basée sur la fréquence (MELS, 2014).

Compléter les données manquantes sur la fréquence et la couverture de texte pourrait éventuellement permettre de créer une liste semblable en FLS puisqu'en ALS, cette tâche a été entamée par Oliveira (2012). Cette chercheuse précisait d'ailleurs qu'une telle liste permettrait aux enseignants de faire des ajustements au niveau lexical dans le développement de matériel pédagogique qui soit davantage adapté au niveau des élèves et qui ciblerait mieux les besoins lexicaux.

Pour les enseignants, le fait de connaître où se situe la connaissance lexicale des apprenants en salle de classe pourrait leur permettre de sélectionner un matériel plus adéquat. En effet, à l'aide d'un outil comme *VocabProfile* (www.lexutor.ca, Cobb, 2017), il est possible d'analyser la richesse lexicale d'un texte et d'ensuite identifier le vocabulaire qui pourrait poser problème aux apprenants, d'autant plus si la TV des apprenants est connue. L'enseignant peut ensuite traiter ces difficultés de différentes façons et modifier le texte, notamment, en remplaçant le vocabulaire moins fréquent qui est inconnu, en s'assurant d'inclure certains mots moins fréquents préalablement enseignés, en intégrant un glossaire en fin de texte pour définir ou même traduire ces mots, etc. L'idée est ici de cibler les mots qui sont vraiment inconnus aux élèves et de les enseigner de façon explicite. *VocabProfile* pourrait aussi être utilisé pour analyser leurs productions écrites et orales. En effet, les évaluations de ces dernières comprennent toujours une partie vocabulaire. Ce sont généralement des descriptions qui permettent d'évaluer si le vocabulaire est « précis et varié ». Un tel outil permettrait aux enseignants de mesurer la richesse lexicale des textes produits par les apprenants et d'ainsi évaluer si ceux-ci utilisent un vocabulaire de plus en plus riche au courant de l'année.

De plus, si les enseignants connaissent la TV des apprenants, il est possible d'utiliser des documents adaptés aux différents niveaux de la salle de classe. En effet, les

niveaux de compétence en langue seconde dans un même groupe scolaire peuvent être multiples, comme le démontrent les importants écarts types de cette étude, et cela amène, bien sûr, un défi supplémentaire pour les enseignants. La TV serait une précieuse information si un enseignant voulait, par exemple, construire des documents adaptés à ces multiples niveaux. C'est le cas d'une collection comme GB+, publiée chez Chenelière et utilisée à l'école primaire. Dans les niveaux les plus faciles, les livres utilisent un vocabulaire plus fréquent, une syntaxe simplifiée et répétitive, puis, graduellement, le nombre et la richesse des mots augmentent, tout comme la structure syntaxique se complexifie. Cette collection offre aussi une trousse d'évaluation afin de déterminer à quel niveau un élève peut entamer sa lecture. Une telle collection pourrait être un outil intéressant au secondaire. Par ailleurs, comme les sujets abordés en classe peuvent varier grandement d'une école à l'autre ainsi que d'une année scolaire à l'autre, les enseignants au secondaire pourraient, toujours à l'aide de *Vocabprofile*, créer des documents adaptés au niveau des élèves et les faire progresser à leur rythme. Cela éviterait d'utiliser des documents peu adaptés à l'âge, l'intérêt ou le niveau de compétence langagière, et affecterait aussi le niveau de motivation des élèves.

Pour terminer, il faut considérer que l'acquisition de la langue ne se fait pas que par la lecture et les interactions. Cet avis était répandu dans les cercles scientifiques dans les années 1970 et 1980 (O'Dell, 1997) et c'est peut-être pour cette raison que cette idée est encore véhiculée dans les programmes scolaires (MELS, 2007). En réalité, depuis presque 30 ans, la recherche en enseignement et en apprentissage du vocabulaire en langues secondes démontre l'importance des stratégies différentes pour enseigner le vocabulaire (Sökmen, 1997) Certes, l'apprentissage fortuit est très important, et suffisant pour certains items lexicaux (Laufer & Rozovski-Roitblat, 2011). Mais, dans d'autres cas, comme l'enseignement des mots non fréquents, des mots polysémiques, des mots avec une écriture/prononciation complexe, c'est un

enseignement planifié qui serait de rigueur (Schmitt & Schmitt, 2014) Celui-ci peut être fait d'une façon implicite ou explicite (Loewen, 2013), et les recherches en didactique des langues secondes produisent de plus en plus de données et d'analyses permettant de comprendre le meilleur type d'enseignement pour des mots, phrases et/ou structures différentes. Ce qui semble mis de l'avant dans cette recherche est le fait que la TV peut augmenter avec des enseignements explicites (Sökmen, 1997). Présentement, la TV des apprenants au secondaire ne semble pas correspondre à un élève qui : « lit et comprend avec facilité une diversité de textes de forme orale, écrite, visuelle ou mixte susceptibles de répondre aux besoins scolaires, personnels, sociaux ou professionnels d'un jeune de son âge » (MELS, 2007, p. 23), mais nous ne savons pas non plus à combien de mots s'évalue une telle description. Bref, il reste beaucoup à faire pour faire concorder les attentes de fin du secondaire et les réalités des apprenants.

5.6 Les perspectives de recherche

La mesure du vocabulaire a été étudiée de façon extensive depuis 25 ans, mais peu de tests sont disponibles en français. La présente recherche a su intégrer une version française de A_Lex pour mesurer la TVPh, une première auprès d'un échantillon comme le nôtre. Mais, il reste beaucoup de chemin à parcourir afin de mesurer la TV chez les apprenants du français et plusieurs recherches pourraient être réalisées. Nous ferons ici quelques suggestions.

Tout d'abord, l'étude de Jean et Simard (2011) a montré que les représentations et les croyances des apprenants et des enseignants au sujet de la grammaire pouvaient être résumées par l'expression « un mal nécessaire ». Cela pouvait être un facteur qui affectait négativement la motivation des élèves dans leur apprentissage de la langue

seconde. Il pourrait être intéressant de savoir comment la motivation des apprenants peut être affectée, positivement ou négativement, devant la quantité de mots à apprendre dans la langue seconde. Pour ce qui est des enseignants, il serait désirable de connaître plus précisément les pratiques d'enseignement du vocabulaire, de même que les croyances face à ces pratiques, qui contribueraient ou non à l'augmentation de la TV chez les apprenants.

Ensuite, pour ce qui est de l'enseignement du vocabulaire, il serait important de voir l'impact d'un enseignement explicite basé sur une analyse des besoins lexicaux des apprenants. Cette pratique pourrait être intégrée en salle de classe et la TV mesurée en pré- et posttest. Suite au prétest, des listes de fréquences pourraient être utilisées par les enseignants pour sélectionner le lexique à enseigner selon le niveau des apprenants. La mesure de la TV en posttest permettrait de vérifier, de façon quantitative et précise, la portée d'un enseignement explicite. Une mesure du vocabulaire orthographique et phonologique pourrait être appliquée pour vérifier si les nouveaux mots ont été intégrés tant à l'oral qu'à l'écrit.

De plus, des données qui semblaient manquer lors de cette recherche concernent le nombre de mots nécessaires pour assurer une compétence en français. Si une TV se situant entre 2K et 3K permet une communication de base en ALS ou qu'une TV entre 8K et 9K permet de lire des textes authentiques sur des sujets variés en ALS, qu'en est-il du français ? La seule donnée que nous avons pour le moment est qu'un apprenant possédant une taille de vocabulaire d'environ 2K en français peut arriver à comprendre 85 % d'un texte oral ou écrit, alors qu'il est connu, en ALS, qu'une maîtrise entre 95 et 98 % du vocabulaire d'un texte oral ou écrit est nécessaire afin de le comprendre de façon adéquate. Il y a un manque flagrant de données pour le français, tant à l'oral qu'à l'écrit.

Pour terminer, dans le but de mieux connaître les pratiques en salle de classe ainsi que le matériel mis à la disposition des enseignants de FLS, une recension des manuels utilisés pourrait être faite comme Oliveira (2012) a fait pour l'ALS. Cela permettrait d'analyser la richesse lexicale des manuels afin d'identifier de potentiels problèmes, comme l'utilisation d'un vocabulaire hors de portée pour les apprenants ou, au contraire, un vocabulaire trop simple qui ne servirait pas aux apprenants pour faire évoluer leurs connaissances lexicales. Cela permettrait aussi de calculer le recyclage (Schmitt, 2013a) fait du nouveau vocabulaire enseigné dans le but d'atteindre suffisamment de répétitions (plus de 10) pour favoriser la rétention des nouveaux mots.

Ce chapitre de discussion a permis d'interpréter les résultats des trois questions de recherche, d'exposer les limites de l'étude pour arriver à en dégager les impacts pédagogiques. Des pistes de recherches sont suggérées afin de continuer de faire avancer le domaine de la mesure du vocabulaire en langue seconde, et plus particulièrement en français.

CONCLUSION

Dans ce mémoire, nous nous sommes intéressés aux connaissances lexicales des apprenants du FLS au secondaire au Québec. La TV de 144 participants de secondaire 1 à 5 a été mesurée. Pour ce faire, nous nous sommes d'abord questionnés sur ce qu'était de connaître un mot, sur la façon dont la TV pouvait montrer la compétence langagière, sur la couverture de texte et la fréquence des mots français, sur les différents types de mesures du vocabulaire réceptif et sur la présence de ces mesures dans le système scolaire. Mais, aucun calcul de la TV (globale, orthographique ou phonologique) n'avait été fait auprès d'apprenants au secondaire, et il a alors semblé essentiel de dresser un portrait des connaissances lexicales et de comparer la TVO et la TVPh afin de déterminer s'il y avait une évolution au fil des années scolaires. C'est à l'aide d'un test orthographique, X_Lex (Meara & Milton, 2003), et d'un test phonologique, A_Lex (Milton & Hopkins, 2005), que la TV des participants a été mesurée.

Malgré le fait que nous ayons dû éliminer tout le groupe scolaire de secondaire 3, qu'il y ait eu certains problèmes techniques et que des réponses aléatoires des apprenants aient affecté la taille de l'échantillon, il a premièrement été possible de démontrer qu'il n'y avait pas de progression significative d'une année à l'autre, mais que, en comparant les résultats par cycle, les TV présentaient des différences significatives. Le fait que le groupe de secondaire 3 ait été éliminé de la recherche a affecté l'analyse d'une année à l'autre. Deuxièmement, la TV globale des apprenants de secondaire 5 n'a pas atteint le niveau supposé de 4 500 familles de mots, puisque la TV moyenne était en fait de 2 796 mots. Nous avons interprété ces résultats selon l'étude de Simard et Jean (2011) qui nous laisse croire que l'étude formelle du

lexique soit un facteur démotivationnel (comme Simard et Jean (2011) l'ont montré avec la grammaire). En outre, nous avons suggéré de s'interroger sur la perception des apprenants et des enseignants face à un enseignement explicite du vocabulaire, de même que sur le fait que celui-ci ne soit pas suffisamment recyclé dans les manuels scolaires, ce qui n'avantage pas la croissance de la TV. Troisièmement, les participants maîtrisaient mieux le 1K que le 2K, le 2K que le 3K, et ainsi de suite. Tous les participants maîtrisaient les mots les plus fréquents du français et, en effet, il est possible d'exprimer beaucoup à l'aide de 2K mots puisqu'ils sont hautement polysémiques. De plus, le faible recyclage des unités lexicales dans les outils d'apprentissage de la langue n'encourage pas l'apprentissage des mots à fréquence intermédiaire (Oliveira, 2012). Fait encourageant : plus un apprenant avance dans son parcours scolaire, plus il maîtrise des mots moins fréquents, et ce, de façon significative en comparaison avec les plus jeunes. Finalement, la TVO était toujours plus élevée que la TVPh, possiblement à cause de la proximité des systèmes orthographiques du français et de l'anglais qui favoriseraient le développement de la TVO. De plus, une TVPh moins élevée reflèterait de moindres besoins lexicaux pour la communication orale (VanZeeland & Schmitt, 2012).

Pour terminer, nous avons proposé de développer les connaissances de la TV, de la couverture de texte et de la fréquence en français, ainsi que de fixer des objectifs quantifiables quant au vocabulaire à connaître pour assurer une pleine compétence en FLS. Nous avons aussi proposé l'utilisation de *VocabProfile* qui pourrait permettre une sélection du matériel d'apprentissage qui soit davantage adapté aux besoins des apprenants puisque cet outil analyse et fournit le détail de la fréquence des mots d'un texte, ce qui favorise un enseignement planifié d'un vocabulaire ciblé.

En conclusion, il semble important de poursuivre la recherche sur la TV, en particulier pour arriver à déterminer combien de mots sont nécessaires pour atteindre 95 % à 98 % de couverture de texte en français, selon les différents besoins lexicaux

des apprenants, dans le but d'améliorer l'apprentissage et l'enseignement du vocabulaire en FLS.

ANNEXE A

LETTRE DE CONSENTEMENT

IDENTIFICATION

Responsable du projet : Mme Jennifer Vachon

Programme d'enseignement : Didactique des langues, français langue seconde

Adresse courriel : jennifervachon@hotmail.com

Téléphone : 514-608-6243

BUT GÉNÉRAL DU PROJET ET DIRECTION

Votre enfant est invité à prendre part à ce projet visant à mesurer la quantité de vocabulaire connu en français. Ce projet de recherche est réalisé dans le cadre d'un mémoire de maîtrise sous la direction de Philippa Bell, professeure du département de Didactique des langues de la Faculté des Sciences de l'éducation. Elle peut être joint au (514) 987-3000 poste 5501 ou par courriel à l'adresse :bell.philippa@uqam.ca.

La direction de l'école de votre enfant ainsi que son professeur(e),
_____ ont également donné leur accord à ce projet.

La contribution de votre enfant favorisera l'avancement des connaissances dans le domaine de l'apprentissage scolaire.

PROCÉDURE(S) OU TÂCHES DEMANDÉES À VOTRE ENFANT

Avec votre permission et l'accord de votre enfant, il sera invité à répondre en classe à deux tests qui mesureront le nombre de mots connus en français. Suite aux tests, les résultats ne seront pas dévoilés, mais utilisés à des fins d'analyse. Les tests se dérouleront dans le local informatique de l'école.

AVANTAGES ET RISQUES D'INCONFORT

Il n'y a pas de risque associé à la participation de votre enfant à ce projet. Néanmoins, soyez assuré que le responsable du projet demeurera attentif à toute manifestation d'inconfort chez votre enfant durant sa participation et prendra toutes les mesures afin que les informations demeurent confidentielles et sécurisées.

ANONYMAT ET CONFIDENTIALITÉ

Il est entendu que les renseignements recueillis auprès de votre enfant sont confidentiels et que seules, Mme Vachon et sa directrice de recherche, auront accès à ses résultats aux tests. L'ensemble du matériel de recherche sera conservé et protégé par un cryptage, ce qui empêche tout accès à d'autres personnes que Mme Vachon.

PARTICIPATION VOLONTAIRE

La participation de votre enfant à ce projet est volontaire. Cela signifie que, même si vous consentez aujourd'hui à ce que votre enfant participe à cette recherche, il demeure entièrement libre de ne pas participer ou de mettre fin à sa participation en tout temps sans justification ni pénalité. Vous pouvez également retirer votre enfant du projet en tout temps.

Votre accord à participer implique également que vous acceptez que le responsable du projet puisse utiliser, aux fins de la présente recherche (articles, mémoire, essai ou thèse, conférences et communications scientifiques), les renseignements recueillis à la condition qu'aucune information permettant d'identifier votre enfant ne soit divulguée publiquement, à moins d'un consentement explicite de votre part et de l'accord de votre enfant.

COMPENSATION

Votre enfant ne sera pas compensé directement. Sa contribution est offerte gratuitement. Sur demande, un résumé des résultats de recherche vous sera transmis au terme du projet.

DES QUESTIONS SUR LE PROJET OU SUR VOS DROITS?

Vous pouvez contacter Mme Vachon pour des questions additionnelles sur le déroulement du projet. Vous pouvez également discuter avec la directrice de recherche, Mme Bell, des conditions dans lesquelles se déroulera la participation de votre enfant et de ses droits en tant que participant de recherche.

Le projet auquel vous allez participer a été approuvé au plan de l'éthique de la recherche avec des êtres humains. Pour toute question ne pouvant être adressée au directeur de recherche ou pour formuler une plainte ou des commentaires, vous pouvez contacter le Président du Comité d'éthique de la recherche pour étudiants (CÉRPÉ), par l'intermédiaire de son secrétariat au numéro (514)-987-3000 # 1646 ou par courriel à : (savard.josee@uqam.ca).

REMERCIEMENTS

Votre collaboration et celle de votre enfant sont importantes à la réalisation de ce projet et nous tenons à vous en remercier.

AUTORISATION PARENTALE

En tant que parent ou tuteur légal de _____ (nom de l'enfant), je reconnais avoir lu le présent formulaire de consentement et consens volontairement à ce que mon enfant participe à ce projet de recherche. Je reconnais aussi que le responsable du projet a répondu à mes questions de manière satisfaisante, et que j'ai disposé suffisamment de temps pour discuter avec mon enfant de la nature et des implications de sa participation. Je comprends que sa participation à cette recherche est totalement

volontaire et qu'il peut y mettre fin en tout temps, sans pénalité d'aucune forme, ni justification à donner. Il lui suffit d'en informer la responsable. Je peux également décider, pour des motifs que je n'ai pas à justifier, de retirer mon enfant du projet.

J'autorise mon enfant à répondre en classe aux deux tests : OUI NON

Signature de l'enfant : _____

Date :

Signature du parent/tuteur légal : _____

Date :

Nom (lettres moulées) et coordonnées : _____

Je déclare avoir expliqué le but, la nature, les avantages, les risques du projet et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature du responsable du projet : _____

Date :

Nom (lettres moulées) et coordonnées : _____

ANNEXE B

QUESTIONNAIRE SOCIODÉMOGRAPHIQUE

1. What grade are you in?

- Secondary 1
- Secondary 2
- Secondary 3
- Secondary 4
- Secondary 5

2. What program are you registered in?

- Regular program
- Immersion program
- IBO program

3. For how many years have you been learning French? _____years

4. With your mother, you speak:

French: never sometimes often always

English: never sometimes often always

Other (please specify: _____)

never sometimes often always

5. With your father, you speak:

French: never sometimes often always

English: never sometimes often always

Other (please specify: _____)

never sometimes often always

6. With your siblings, you speak:

French: never sometimes often always

English: never sometimes often always

Other (please specify: _____)

never sometimes often always

7. How often do you speak another language than English with your friends?

French: never sometimes often always

Other (please specify: _____)

never sometimes often always

ANNEXE C

LISTE DES MOTS TESTÉS X_LEX

1K	2K	3K	4K	5K	Pseudomots
je	ouest	inspirer	détenir	concurrent	rojette
temps	transformer	paquet	accélérer	mouiller	tellène
chaque	formule	mélange	suprême	bavarder	sureux
fin	sien	proclamer	élimination	raser	touceul
revenir	faveur	distinction	avenue	divorce	brossin
fil	agent	débarrasser	purement	pardon	dour
fort	chasse	perce	gant	ascenseur	froise
anglais	observation	creuser	remporter	cravate	gillais
âge	mériter	fleuve	insuffisant	lucide	leinte
lever	empire	prudence	border	tecture	morler
bientôt	fameux	menace	amiral	hautement	overge
faible	passé	miel	étiquette	pneu	pentée
actuellement	immeuble	substance	teneur	célébrité	poulard
sécurité	courant	logique	chaire	différemment	ralive
prévoir	fonctionner	barbe	fureur	fragment	tentre
entrée	angoisse	décoration	mentir	inconcevable	champile
charge	debout	procédure	récolter	pochoir	gélard
revoir	intéressé	coude	taverne	ruelle	houroux
presse	extrêmement	psychologie	ultime	succession	introis
mise	liste	taureau	vernis	tumeur	loruste

ANNEXE D

LISTE DES MOTS TESTÉS A_LEX

1K	2K	3K	4K	5K	Pseudomots
que	clair	dessus	brouillard	buffle	abattir
dont	octobre	grouper	tante	innocent	aige
devenir	complexe	élu	proximité	animation	bletter
ville	bouche	toit	paire	habileté	brottin
nuit	inconnu	classer	habiller	vieillesse	champile
oui	centaine	négliger	déchirer	éther	clabrer
docteur	spécialiser	progresser	vol	régir	culon
époque	pareil	silhouette	metteur	brûlé	jupaire
malgré	contemporain	avancé	disparition	habillé	diroir
marché	accomplir	muscle	montage	originaire	détrame
aussitôt	défaut	collaborateur	terrestre	épanouir	fontre
connu	causer	formuler	vigoureux	fronce	fréquir
air	revue	tasse	malin	sauvegarder	gélard
lendemain	odeur	auditoire	véritablement	comtesse	introis
ton	peser	précipiter	contribuable	élaboration	lifrer
soi	plage	fatiguer	intégral	grillage	manchir
admettre	rendement	judiciaire	pastoral	modéré	outrir
méto	réduit	signal	requête	radio	piédeur
rare	bataille	accrocher	retrait	remise	rubir
saison	réduction	caché	vernis	adjudant	vusier

RÉFÉRENCES

- Adolphs, S., Schmitt, N. (2003) Lexical coverage of spoken discourse, *Applied Linguistics*, 24(4), pp. 425-438.
- Armand, F. (2012). Enseigner en milieu pluriethnique et plurilingue : place aux pratiques innovantes!. *Québec français*, (167), 48-50.
- Astika, G. G. (1993). Analytical assessments of foreign students' writing. *RELC journal*, 24(1), 61-70.
- Batista, R. (2014) *A Receptive Vocabulary Knowledge Test for French L2 Learners With Academic Reading Goals* (Mémoire de maîtrise). Concordia University.
- Baudot, J. (1992). *Les fréquences d'utilisation des mots en français écrit contemporain*. Montréal: Presses de l'Université de Montréal.
- Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M. and Van de Velde, H. (2001) Examining the Yes/No vocabulary test: some methodological issues in theory and practice, *Language Testing*, 18, pp. 235-274
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language testing*, 16(2), pp. 131-162.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.
- Blais, I. (2014). La taille du vocabulaire et les habiletés réceptives des étudiants adultes de niveau 6 en francisation. (Mémoire de maîtrise inédit). Université du Québec à Montréal.
- Bulté, B., Housen, A., Pierrard, M., & Van Daele, S. (2008). Investigating lexical proficiency development over time—the case of Dutch-speaking learners of French in Brussels. *Journal of French language studies*, 18(03), 277-298.
- Coady, J. (1997). L2 vocabulary acquisition: A synthesis of the research. *Second language vocabulary acquisition*, 273-290.

- Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French. *Vocabulary in a second language*, pp. 15-38.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38-63.
- Cobb, T. (2017) *Web Vocabprofile* [<http://www.lex tutor.ca/vp/>], an adaptation of Heatley, Nation & Coxhead's (2002) *Range*.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1), 204-256.
- Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*, 41(4), 965-981.
- Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. *Handbook of psycholinguistics*, 2, 249-284.
- Daller, H., & Phelan, D. (2007). What is in a teacher's mind? Teacher ratings of EFL essays and different aspects of lexical richness. *Modelling and assessing vocabulary knowledge*, 234-244.
- David, A. (2008). Vocabulary breadth in French L2 learners. *Language Learning Journal*, 36(2), pp. 167-180.
- Dictionnaire Larousse en ligne. Repéré à <http://www.larousse.fr/dictionnaires/francais/menterie/50515>
- English Montreal School Board (2015), *Elementary school registration*. Repéré à http://www.emsb.qc.ca/emsb_en/pdf_en/schools_en/Elem_Reg_eng-en_2015_web.pdf
- Ellis, N. C., & Robinson, P. (2008). An introduction to cognitive linguistics, second language acquisition, and language instruction. *Handbook of cognitive linguistics and second language acquisition*, 3-24.
- Ellis, N. & Wulff, S. (2014) In VanPatten, B., & Williams, J. *Theories in second language acquisition: An introduction*. Routledge. p. 75-93
- Eyckmans, J. (2004). Measuring receptive vocabulary size. *Utrecht: LOT*.

- Gaudreau, L. (2011). *Guide pratique pour créer et évaluer une recherche scientifique en éducation*. Guérin.
- Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). RANGE and FREQUENCY programs. Repéré à <http://www.victoria.ac.nz/lals/staff/paul-nation.aspx> .
- Horst, M., Cobb, T., & Nicolae, I. (2005). Expanding academic vocabulary with an interactive on-line database. *Language learning & technology*, 9(2), 90-110.
- Horst, M., & Collins, L. (2006). From faible to strong: How does their vocabulary grow?. *Canadian modern language review*, 63(1), 83-106.
- Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: necessary, but boring?. *Foreign Language Annals*, 44(3), 467-494.
- Kamimoto, T. (2014). Local item dependence on the Vocabulary Levels Test revisited. *Vocabulary Learning and Instruction*, 3(2), 56-68.
- Larson-Hall, J., & Herrington, R. (2010). Improving data analysis in second language acquisition by utilizing modern developments in applied statistics. *Applied Linguistics*, 31(3), 368-390.
- Laufer, B. (1992). Reading in a foreign language: how does L2 lexical knowledge interact with the reader's general academic ability?. *Journal of Research in Reading*, 15(2), 95-103.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16(3), pp. 307-322.
- Laufer, B. 1997. «What's in word that makes it hard or easy: some intralexical factors that affect the learning of words». Dans Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press. p. 140-155.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics*, 22(1), 1-26.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B. (2009). Second language vocabulary acquisition from language input and from form-focused activities. *Language Teaching*, 42(03), 341-354.

- Laufer, B., & Ravenhorst-Kalovski, G.C. (2010) Lexical Threshold Revisited : Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language* 22(1), pp. 15-30.
- Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research*, 15(4), 391-411.
- Lester B. Pearson (2017). *Elementary school programs*. Repéré à <http://www.lbpsb.qc.ca/eng/schools/admissions/admissionsV2.asp>
- Loewen, S. (2013). Instructed second language acquisition. *The encyclopedia of applied linguistics*.
- Meara, P. Buxton, B. (1987) An alternative to multiple choice vocabulary tests, *Language testing*, 4, pp. 142-154
- Meara, P. (1992). *EFL Vocabulary Tests*. Swansea: CALS University of Wales Swansea.
- Meara, P. (1995). The importance of an early emphasis on L2 vocabulary. *LANGUAGE TEACHER-KYOTO-JALT-*, 19, pp. 8-11.
- Meara, P. (1996). The dimensions of lexical competence, In Brown, G., Malmkjaer, K. & Williams, J., *Performance and competence in second language acquisition*, Cambridge University Press, pp. 35-53
- Meara, P. & Milton, J. (2003). *X_Lex, The Swansea Levels Test*. Newbury : Express.
- Ministère de l'Éducation, du Loisir et du Sport (2007). Programme de formation de l'école québécoise. Repéré à <http://www1.education.gouv.qc.ca/sections/programmeFormation/index.asp>
- Ministère de l'Éducation, du Loisir et du Sport (2007). Progression des apprentissages au secondaire. Repéré à <http://www1.education.gouv.qc.ca/progressionSecondaire/>
- Ministère de l'Éducation, du Loisir et du Sport (2014). Liste orthographique à l'usage des enseignantes et des enseignants. Repéré à http://www.education.gouv.qc.ca/fileadmin/site_web/documents/dpse/formation_jeunes/ListeOrthographique_Primaire.pdf

- Milton, J. and Meara, P. (1998). Are the British really bad at learning foreign languages? *Language Learning Journal*, 18, pp. 68–76.
- Milton, J. and Hopkins, N. (2005) *Aural Lex: Aural Vocabulary Size Test*. University of Wales Swansea; Swansea
- Milton, J., Hopkins, N. (2006) Comparing phonological and orthographic vocabulary size : do vocabulary test underestimate the knowledge of some learners ?, *Canadian Modern Language Review/ La Revue Canadienne des langues vivantes*, 63 (1), pp. 127-147
- Milton, J. (2006). Language Lite? Learning French Vocabulary in School. *Journal of French Language Studies*, 16, pp 187-205
- Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Multilingual Matters.
- Milton J., Wade, J. & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, . Torreblanca-López (Eds.), *Insights into non-native vocabulary teaching and learning* (pp. 83-98). Bristol : Multilingual Matters.
- Milton, J. (2013) Measuring the contribution of vocabulary knowledge to proficiency in the four skills, *EUROSLA monographs series 2, L2 vocabulary acquisition, knowledge and use*, pp. 57-78
- Mochida, K., Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge , *Language Testing*, 23 (1), pp. 73-98
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston : Heinle & Heinle.
- Nation, I.S.P. (1997) Bringing today's vocabulary research into tomorrow's classrooms. In G.M. Jacobs (ed.) *Language Classrooms of Tomorrow: Issues and Responses* RELC Anthology Series No. 38, pp. 170-182.
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, 14, 6-19.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge, UK: Cambridge University Press.
- Nation, P. and P. Meara. (2002). *Vocabulary* in N. Schmitt (ed.): *An Introduction to*

Applied Linguistics. London : Arnold.

- Nation, I.S.P. (2006) How Large a Vocabulary Is Needed For Reading and Listening? *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63, 1, pp. 59-82
- Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, 1(1), pp. 2-13.
- Nation, I. S. P. (2011). Research into practice: Vocabulary. *Language Teaching*, 44(4), 529.
- O'Dell, F. (1997) In Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press. P. 258-278
- Oliveira Pisani Martini, J. (2012). *High Frequency Vocabulary in a Secondary Quebec ESL Textbook Corpus* (Doctoral dissertation, Concordia University).
- Pellicer-Sánchez, A., Schmitt, N. (2012) Scoring Yes–No vocabulary tests : Reaction time vs. nonword approaches. *Language Testing*, (29 (4), pp. 489-509
- Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, 11, 67-86.
- Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian modern language review*, 56(2), 282-308.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Régime pédagogique de l'éducation préscolaire, de l'enseignement primaire et de l'enseignement secondaire (2017). Repéré à <http://legisquebec.gouv.qc.ca/fr/showdoc/cr/I-13.3,%20r.%208>
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 77-89.
- Richards, B. and Malvern, D. (2008). Validity and threats to the validity of vocabulary measurement. In : H. Daller, J. Milton, and J. Treffers-Daller, *Modelling and Assessing Vocabulary Knowledge*. Cambridge University Press.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in second language acquisition*, 19(01), 17-36.

- Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy* (Vol. 2035). Cambridge: Cambridge University Press.
- Schmitt, N. 2000. *Vocabulary in Language Teaching*. Cambridge : Cambridge University Press.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language testing*, 18(1), pp. 55-88.
- Schmitt, N. (2010). *Researching vocabulary*. Basingstoke, England: Palgrave Macmillan.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26-43.
- Schmitt, N. (2013a, October). *Research-based principles of vocabulary teaching*. Paper presented at the conference Classroom-oriented Research: Reconciling Theory and Practice, Konin, Poland.
- Schmitt, N. (2013b, December). *Size and Depth of Vocabulary: What Does the Research Show?*. Paper presented at the conference Vocab@Vic, Wellington, New Zealand.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(04), 484-503.
- Schonell, F. J., I. G. Meddleton, and B. A. Shaw. 1956. *A Study of the Oral Vocabulary of Adults*. Brisbane: University of Queensland Press.
- Sokmen, A. 1997. «Current trends in teaching second language vocabulary». In Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press. p. 237-257.
- Stæhr, L. S. (2008) Vocabulary size and the skills of listening, reading and writing, *The Language Learning Journal*, 36 (2), pp. 139-152
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(04), pp. 577-607.
- Séguin, H., & Tréville, M.-C. (1992). « Les congénères interlinguaux : un atout pour accélérer l'acquisition du vocabulaire et faciliter la compréhension des

textes ? », dans Robert J. Courchêne, Jennifer I. Glidden, J. St-John et Christiane Thérien, *L'enseignement des langues secondes axé sur la compréhension*. Ottawa : Presses de l'Université d'Ottawa, 1992.

Thouin, M. (2014). *Réaliser une recherche en didactique*. Éditions MultiMondes,.

Tidball, F., & Treffers-Daller, J. (2008). Analysing lexical richness in French learner language: What frequency lists and teacher judgements can tell us about basic and advanced words. *Journal of French Language Studies*, 18(03), pp. 299-313.

Van der Maren, J.-M. (1995). *Méthodes de recherche pour l'éducation*. Montréal: Presses de l'université de Montréal.

Van Zeeland, H., & Schmitt, N. (2012). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?. *Applied Linguistics*, ams074.

Van Zeeland, H. (2013). L2 vocabulary knowledge in and out of context. *Australian Review of Applied Linguistics*, 36(1), 52-70.

Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics*, 28 (1), 46-65.

Wilkins, D. A. (1972) *Linguistics in Language Teaching*. London : Arnold.

Zimmerman, K. J. (2004). *The role of vocabulary size in assessing second language proficiency* (Unpublished master's thesis). Brigham Young University, USA.