

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA TAILLE DU VOCABULAIRE ET LES HABILITÉS RÉCEPTIVES DES
ÉTUDIANTS ADULTES DE NIVEAU 6 EN FRANCISATION

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

PAR
ISABELLE BLAIS

JUIN 2014

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.01-2006). 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 permis que ce mémoire voie le jour et je souhaite les en remercier.

Tout d'abord, je tiens à exprimer toute ma gratitude à mon directeur, Tom Cobb, qui m'a guidé tout au long de ce mémoire. J'ai apprécié son enthousiasme communicatif pour la recherche en vocabulaire, sa disponibilité, et ses précieux conseils.

Ensuite, je veux remercier mes deux lecteurs, Dominic Anctil et Gladys Jean, d'avoir manifesté leur intérêt pour mon projet et d'avoir accepté de lire mon mémoire.

Je souhaite aussi dire merci à l'enseignante qui m'a accueillie dans sa classe et aux étudiants qui ont participé à ma recherche. Sans leur participation, ce projet n'aurait pas pu exister.

Je veux adresser mes remerciements à Jill Vandermeerschen, statisticienne au SCAD, pour son aide et sa patience lors de l'analyse statistique.

Finalement, je souhaite remercier de tout coeur mon conjoint et ma famille, qui n'ont jamais cessé de croire en moi et qui m'ont encouragée à atteindre mes objectifs.

TABLE DES MATIÈRES

LISTE DES TABLEAUX	iii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	iv
RÉSUMÉ.....	v
INTRODUCTION.....	1
CHAPITRE I PROBLÉMATIQUE.....	3
1.1 Introduction	3
1.2 Le vocabulaire et l'apprentissage d'une langue seconde	5
1.3 La taille du vocabulaire	6
1.4 Le vocabulaire et les habiletés langagières	7
1.5 Pertinence de cette étude	9
1.6 Questions et objectifs de recherche	10
1.7 Conclusion.....	11
CHAPITRE II CADRE THÉORIQUE.....	12
2.1 Introduction	12
2.2 Qu'est-ce qu'un mot?	12
2.3 Qu'est-ce que connaître un mot?.....	16
2.4 La taille du vocabulaire	18
2.5 Les facteurs influençant l'acquisition du vocabulaire.....	20
2.6 La fréquence des mots	22
2.7 Le vocabulaire et la lecture.....	23
2.8 Le vocabulaire et la compréhension orale.....	25
2.9 L'enseignement du vocabulaire.....	27
2.10 Conclusion	29
CHAPITRE III MÉTHODOLOGIE	30
3.1 Introduction	30
3.2 Population.....	30
3.3 Description des instruments de mesure	32
3.3.1 Questionnaire sociodémographique.....	32

3.3.2	Test de vocabulaire	35
3.3.3	Examens de compréhension orale et de compréhension écrite.....	37
3.4	Déroulement de la collecte de données	39
3.5	Traitement des données	39
3.6	Conclusion.....	40
CHAPITRE IV ANALYSE DES RÉSULTATS		41
4.1	Introduction	41
4.2	Test de vocabulaire.....	41
4.3	Taille du vocabulaire et habiletés réceptives.....	44
4.4	Questionnaire sociodémographique	49
4.5	Conclusion.....	53
CHAPITRE V DISCUSSION.....		54
5.1	Introduction	54
5.2	Questions et objectifs de recherche	54
5.2.1	Taille du vocabulaire	55
5.2.2	Taille du vocabulaire et habiletés réceptives	57
5.2.3	Rôle des facteurs sociodémographiques	60
5.3	Limites.....	63
5.4	Implications pédagogiques	64
5.5	Perspectives de recherches	67
5.6	Conclusion.....	68
CONCLUSION		69
ANNEXE A : LETTRE DE CONSENTEMENT		71
ANNEXE B : QUESTIONNAIRE SOCIOLOGIQUE.....		74
ANNEXE C : TEST DE VOCABULAIRE (3 NIVEAUX)		75
BIBLIOGRAPHIE		78

LISTE DES TABLEAUX

Tableau 4-1 Résultats au test de vocabulaire	43
Tableau 4-2 Taille du vocabulaire (en nombre de mots)	43
Tableau 4-3 Résultats aux examens de compréhension orale et de compréhension écrite.....	44
Tableau 4-4 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k, 2k et 3k) et le résultat à l'examen de compréhension orale	45
Tableau 4-5 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k et 3k) et le résultat à l'examen de compréhension orale	46
Tableau 4-6 Régression multiple pour le résultat au test de vocabulaire (niveau 1k) et le résultat à l'examen de compréhension orale	47
Tableau 4-7 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k, 2k et 3k) et le résultat à l'examen de compréhension écrite	47
Tableau 4-8 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k et 3k) et le résultat à l'examen de compréhension écrite	48
Tableau 4-9 Régression multiple pour le résultat au test de vocabulaire (niveau 1k) et le résultat à l'examen de compréhension écrite	49
Tableau 4-10 Tests T sur la langue maternelle	50
Tableau 4-11 Tests T sur la scolarité	52
Tableau 4-12 Corrélations entre le contact avec la langue française et la taille du vocabulaire, la compréhension orale et la compréhension écrite.....	53

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

MELS	Ministère de l'Éducation, du Loisir et du Sport
1k	premiers 1000 mots les plus fréquents en français
2k	deuxièmes 1000 mots les plus fréquents en français
3k	troisièmes 1000 mots les plus fréquents en français

RÉSUMÉ

Les immigrants adultes qui terminent le programme de francisation du MELS souhaitent avoir un niveau de maîtrise suffisant de la langue pour s'intégrer à la société québécoise. Cette maîtrise s'exprime, entre autres, par une connaissance adéquate du vocabulaire fréquent, laquelle n'a pas encore été clairement définie dans le domaine de la francisation. Pourtant, les études antérieures démontrent que la taille du vocabulaire réceptif est un bon indicateur des habiletés linguistiques générales (David, 2008), notamment les habiletés réceptives que sont la lecture et la compréhension orale (Staeher, 2008).

Cette recherche vise à mesurer la taille du vocabulaire réceptif des étudiants adultes qui terminent le programme de francisation du MELS. Elle vise également à évaluer le rôle joué par la taille du vocabulaire dans les habiletés réceptives. Les 17 participants de notre étude sont des étudiants adultes de niveau 6 suivant le programme du MELS. La taille de leur vocabulaire a été mesurée à l'aide d'un test couvrant les 3000 premiers mots les plus fréquents, inspiré du X-Lex de Meara. Leurs habiletés réceptives ont été évaluées par le biais des examens ministériels de compréhension orale et de compréhension écrite. Les participants ont également complété un questionnaire afin d'identifier si certains facteurs sociodémographiques (sexe, âge, scolarité, contact avec la langue française) jouent un rôle dans la taille du vocabulaire et les habiletés réceptives.

Les principaux résultats démontrent que les étudiants ont, en moyenne, un vocabulaire de 2228 mots (écart-type=370). Plusieurs étudiants n'ont pas une maîtrise suffisante des mots fréquents. Une corrélation significative existe entre la taille du vocabulaire et les résultats aux examens mesurant les habiletés réceptives, en particulier pour celui de compréhension orale. L'analyse démontre le rôle très important des mots du premier niveau de 1000. Ces résultats prouvent qu'il est important pour le MELS de fixer des objectifs précis en matière d'apprentissage du lexique fréquent et de donner aux enseignants et aux apprenants les outils nécessaires pour les atteindre.

Mots clés : taille du vocabulaire, francisation, adultes, compréhension orale, compréhension écrite

INTRODUCTION

Au Québec, la maîtrise du français chez les immigrants adultes est un sujet d'intérêt évident. Notre expérience en tant qu'enseignante de francisation nous a appelée à nous interroger sur la qualité de la formation offerte aux étudiants. Nous nous sommes demandé si, au terme de leur formation, les immigrants possédaient les outils linguistiques nécessaires pour atteindre leurs objectifs personnels et professionnels.

Certes, le niveau de maîtrise de la langue peut être considéré de plusieurs manières. Parmi tous les angles d'études possibles, notre choix s'est arrêté sur le vocabulaire. Cette décision se justifie par les nombreuses recherches démontrant que la taille du vocabulaire est un des meilleurs prédicateurs des habiletés langagières générales, notamment les habiletés réceptives que sont la compréhension écrite et la compréhension orale.

Notre mémoire se divise en cinq chapitres. Le premier chapitre présentera la problématique, les questions et les objectifs de recherche. Le deuxième chapitre, basé sur une revue de la littérature, décrira le cadre théorique dans lequel s'insère notre recherche. Nous y définirons les concepts-clés dans le domaine du vocabulaire et nous établirons les liens entre le lexique et les habiletés langagières réceptives. Le troisième chapitre sera consacré à la méthodologie. Nous y présenterons en détails les instruments de mesure utilisés dans notre recherche et nous justifierons nos choix. Au

quatrième chapitre, nous procéderons à l'analyse statistique des données recueillies. Finalement, dans le cinquième chapitre, nous discuterons des résultats obtenus afin de répondre à nos questions de recherche. Puis, nous exposerons les limites de notre étude et les implications pédagogiques de nos découvertes. Finalement, nous offrirons aux chercheurs quelques pistes de recherches futures dans le même domaine. Après ce dernier chapitre, nous serons en mesure de tirer des conclusions générales.

CHAPITRE I

PROBLÉMATIQUE

1.1 Introduction

L'immigration joue un rôle important dans le développement de la société québécoise. Chaque année, environ 50 000 immigrants s'établissent au Québec et plusieurs d'entre eux parlent peu ou pas du tout le français. Effectivement, des données provenant du Ministère de l'Immigration et des communautés culturelles (2011) nous apprennent qu'en 2010, un total de 53 985 personnes ont immigré au Québec et que 65% d'entre elles connaissaient le français. Cela signifie que 18 895 nouveaux arrivants ne connaissaient pas la langue officielle. Or, la maîtrise du français est au coeur d'une intégration réussie à la communauté francophone. Dans un document intitulé *Apprendre le Québec*, rédigé à l'intention des futurs immigrants, le Gouvernement du Québec tient ces propos :

La connaissance du français est indispensable pour participer à la vie culturelle, civique et sociale, pour vivre au quotidien, pour mettre pleinement à profit vos connaissances professionnelles et pour intégrer plus rapidement le marché du travail. (Québec, 2010, p. 30)

Pour être fonctionnels dans la société d'accueil majoritaire, beaucoup d'immigrants adultes doivent donc suivre des cours de francisation. Dans cette optique, le Gouvernement du Québec investit d'importantes sommes d'argent afin

que les immigrants bénéficient de cours de francisation. Les cours de francisation visent à ce que les étudiants puissent, à la fin de leur formation, s'intégrer à la société québécoise, se trouver un emploi ou poursuivre des études en français.

Le Ministère de l'Éducation, du Loisir et du Sport (MELS) est l'un des organismes, avec le Ministère de l'Immigration et des communautés culturelles (MICC), qui assure la francisation des immigrants. Au sein des commissions scolaires, les centres de formation générale des adultes offrent une formation en francisation suivant le Programme d'études du français langue seconde. Ce programme, datant de 1994, sera bientôt remplacé par un nouveau programme qui compte douze niveaux et qui vise à uniformiser la formation de francisation entre le MELS et le MICC. Cependant, comme ce nouveau programme n'est pas encore implanté dans les centres de francisation au moment des présentes, nous ferons référence au programme de 1994. Ce dernier comporte un tronc commun de six niveaux, d'une durée d'environ huit semaines chacun, visant à développer les quatre habiletés langagières, c'est-à-dire la compréhension orale, la production orale, la compréhension écrite et la production écrite. La maîtrise de ces habiletés est mesurée à la fin de chaque niveau au moyen de quatre examens élaborés par le MELS. L'objectif général du programme est le suivant :

Au terme du tronc commun, l'élève doit être en mesure de comprendre des messages oraux et écrits dans des situations variées, liées à ses besoins et à ses intérêts [et] d'intervenir oralement et par écrit dans des situations variées, liées à ses besoins et à ses intérêts. (Gouvernement du Québec, 1994, p. 34)

Les situations dont parle le programme appartiennent à six domaines différents : commercial, culturel, professionnel, public, scolaire et social. À titre d'exemples, voici quelques situations qui sont au programme : *À la banque; Recherche d'un logement; Consultation au sujet de la santé; et Recherche d'un emploi.* Il s'agit donc d'une approche thématique, inspirée de situations de la vie quotidienne.

1.2 Le vocabulaire et l'apprentissage d'une langue seconde

Dans le programme de francisation, on préconise l'approche communicative et la connaissance du vocabulaire n'y est pas considérée comme une habileté en soi. Le lexique se retrouve au second plan, au service des activités communicatives orales et écrites. En effet, il n'existe pas de liste de lexique spécifique à maîtriser à chacun des niveaux. Le programme indique seulement que l'étudiant doit connaître le lexique propre aux notions et fonctions énumérées. À titre d'exemple, au niveau 6, on aborde la question de l'entrevue d'emploi. Le programme présente la fonction *Parler de soi* et la notion *poste* : *études, compétences, objectifs de carrière, disponibilité*. Ces notions et fonctions laissent beaucoup de place aux variations individuelles. Dans chaque classe, on risque de présenter un vocabulaire différent suivant les besoins et les intérêts des étudiants.

Puisqu'il n'y a pas d'objectif précis quant à la maîtrise du vocabulaire, le lexique ne fait pas l'objet d'une évaluation spécifique : il est plutôt intégré aux autres évaluations. Par exemple, la richesse et la variété du vocabulaire font partie des critères à évaluer par l'enseignant dans les examens de production orale et de production écrite. De plus, la compréhension du vocabulaire est importante pour répondre aux questions des examens de compréhension orale et écrite. Le programme du MELS n'offre en revanche aucun instrument de mesure objectif pour évaluer le vocabulaire des étudiants.

Pourtant, le lexique est loin d'être un détail dans l'apprentissage d'une langue seconde. Un grand nombre de chercheurs ont mis en évidence toute l'importance du vocabulaire. Read (2000) compare les mots à des briques, car ils sont des matériaux essentiels et porteurs de sens permettant de construire des phrases. En se basant sur de nombreuses années de recherche en matière de vocabulaire, Vermeer (cité par Laufer,

1997), arrive à cette conclusion : « Knowing words is the key to understanding and being understood. The bulk of learning a new language consists of learning new words. Grammatical knowledge does not make for great proficiency in a language. » (p.140)

Étant donné le caractère essentiel de l'apprentissage du vocabulaire, l'idée d'en mesurer la connaissance chez les étudiants de langue seconde est tout à fait logique. Les chercheurs s'intéressent principalement à deux aspects du vocabulaire : sa taille, aussi appelée étendue, et sa profondeur. La première réfère au nombre de mots connus par un apprenant et la deuxième correspond plutôt au degré de connaissance de ces mots. Notre recherche se penchera sur la taille du vocabulaire, car cet aspect est plus facile à mesurer, en plus d'être un bon indicateur des habiletés langagières générales des étudiants (David, 2008). D'ailleurs la majorité des recherches dans le domaine du lexique ont exploré la taille du vocabulaire.

1.3 La taille du vocabulaire

Les apprenants d'une langue seconde doivent connaître un certain nombre de mots afin d'être en mesure de communiquer et de comprendre. Il ne s'agit pas, comme le soulignent Nation et Waring (1997), de connaître tous les mots contenus dans un dictionnaire ou d'atteindre la taille de vocabulaire d'un locuteur natif. Cela serait un objectif aussi irréaliste qu'inutile. En effet, bien que toute langue compte un nombre très élevé de mots, ceux-ci n'ont pas tous la même utilité. Une façon de déterminer à quel point un mot est utile dans une langue est d'en connaître la fréquence. Milton (2009) explique qu'une petite quantité de mots sont beaucoup plus fréquemment utilisés que les autres, ce qui fait qu'un apprenant est plus susceptible de les rencontrer et de les apprendre. Les recherches démontrent que les mots les plus fréquents ont tendance à être appris en premier. Par exemple, Biemiller (2001) a

étudié l'acquisition du vocabulaire en langue maternelle chez les enfants et a conclu que les mots sont acquis dans un ordre relativement fixe. Le développement du lexique n'est donc pas aléatoire, mais bien organisé et prévisible, dans une certaine mesure. Des analyses de fréquence effectuées sur de larges corpus démontrent qu'une petite quantité de mots couvrent une grande partie des textes écrits et des conversations orales. Nation et Waring (1997) disent qu'en anglais, la connaissance des 2000 premières familles de mots les plus fréquentes permet de couvrir environ 80% d'un texte écrit. Selon Adolphs et Schmitt (2003), la même quantité de mots offre une couverture encore plus grande pour un texte oral, soit un peu plus de 90%. Pour bénéficier d'une bonne base en langue seconde, Nation et Waring concluent que les étudiants devraient maîtriser les 3000 mots les plus fréquents. Par la suite, ils peuvent utiliser diverses stratégies pour acquérir le nouveau vocabulaire: l'utilisation du contexte, par exemple.

1.4 Le vocabulaire et les habiletés langagières

La maîtrise du vocabulaire est importante chez les étudiants, car elle influence toutes leurs habiletés langagières. Comme le mentionne David (2008), des études antérieures ont prouvé que la taille du vocabulaire d'un apprenant de langue seconde est un bon prédicateur de son habileté langagière générale. À titre d'exemple, l'auteure cite les chercheurs Meara et Jones (1990), qui ont observé une forte corrélation ($r = 0,7$) entre la taille du lexique et les habiletés en production écrite, en compréhension écrite et en grammaire.

La majorité des recherches ayant étudié le lien entre le vocabulaire et les habiletés langagières portent sur la lecture. Ces études démontrent qu'il existe une corrélation très forte entre la compréhension en lecture et la taille du vocabulaire réceptif. À ce sujet, Laufer (2010) affirme: « Regarding the relationship between

vocabulary and reading, most researchers agree that vocabulary is a good predictor of reading, if not the best. » (p. 16) Plus un étudiant connaît de mots dans un texte, plus il est susceptible de comprendre ce texte. Le pourcentage de mots connus du lecteur dans un texte donné correspond à ce que les chercheurs appellent *coverage*, terme que nous traduirons en français par *couverture*. Bien sûr, cela ne signifie pas que la connaissance de tous les mots d'un texte permet d'atteindre sa compréhension totale, car un texte est davantage qu'une somme de mots. Malgré cela, Milton (2009) soutient que, « ... there is more than an element of truth to the idea that coverage is important to comprehension and that knowledge of the most important words, which contributes so much to coverage, is also important. » (p. 48) Dans un même ordre d'idées, Schmitt, Jiang et Grabe (2011) ont découvert que la couverture et la compréhension étaient étroitement liées. En effet, plus le pourcentage de mots connus par le lecteur est élevé, plus la compréhension augmente, et ce, d'une façon linéaire.

Bien que moins nombreuses que celles sur la lecture, d'autres recherches ont étudié le rôle du vocabulaire dans la compréhension orale en langue seconde. Kelly (1991) a identifié la méconnaissance du vocabulaire comme étant l'obstacle principal à la compréhension orale en langue étrangère chez les étudiants de niveau avancé. Bonk (2000) a établi qu'il existait une corrélation significative, mais modeste, entre la connaissance du lexique d'un document oral et la qualité de sa compréhension. Staehr (2008), de son côté, a observé une corrélation significative ($r = 0,69$) entre la taille du vocabulaire réceptif et le résultat à un test de compréhension orale. Finalement, une étude menée par Van Zeeland et Schmitt (2013) a démontré que le niveau de compréhension orale des apprenants était systématiquement et proportionnellement affecté par le remplacement d'une proportion de mots par des non mots dans un document.

1.5 Pertinence de cette étude

La revue de littérature nous démontre que la très grande majorité des recherches sur le vocabulaire ont porté sur l'anglais langue seconde. Très peu d'études ont été menées auprès d'étudiants de français langue seconde. L'une d'entre elles est celle de Milton (2006), qui a étudié le développement du vocabulaire réceptif d'apprenants de français de niveau secondaire en Angleterre. Milton a conclu qu'après cinq ans d'étude du français, la taille du vocabulaire moyen était de 852 mots (écart-type = 440). Après sept années d'étude, les finissants possédaient en moyenne un vocabulaire de 1930 mots (écart-type = 475). L'auteur juge que ces résultats sont faibles lorsqu'ils sont comparés à des données historiques ou à ce qui est atteint dans l'apprentissage d'autres langues secondes. Bien que cette étude soit intéressante, elle diffère grandement de la nôtre à cause de son contexte. D'abord, les étudiants de l'étude de Milton étaient des adolescents et non des adultes. Ensuite, ces jeunes ont étudié le français pendant plusieurs années dans un contexte non intensif, et non pendant seulement quelques mois de manière intensive. Finalement, en Angleterre, le français est une langue étrangère et non la langue officielle, comme c'est le cas au Québec, ce qui signifie que les étudiants ne se trouvaient pas dans un contexte immersif.

La deuxième étude concernant le vocabulaire en français langue seconde a été menée par David (2008). Cette dernière s'est inspirée de la recherche de Milton (2006), mais l'a poussée un peu plus loin en s'intéressant aux étudiants universitaires, en plus des étudiants du secondaire. Elle a mesuré la taille du vocabulaire des étudiants et elle a également vérifié s'il existait une corrélation entre la taille du lexique et le résultat aux examens finaux mesurant les compétences langagières générales. Sa recherche n'a pas été concluante sur ce dernier point, étant donné qu'aucune corrélation significative n'a été découverte. Encore une fois, cette étude s'est effectuée dans un contexte différent quant à l'âge des participants, le nombre de

temps passé à étudier le français et le caractère non immersif de l'apprentissage. Jusqu'à maintenant, aucune recherche sur le vocabulaire n'a étudié la population des étudiants adultes en francisation au Québec. Il s'agit d'un contexte unique en son genre et notre recherche sera une pionnière dans le domaine.

L'autre aspect innovateur de notre recherche est qu'elle se penche sur le lien entre le vocabulaire et la compréhension orale. Or, la revue de la littérature a démontré que ce lien n'a été que très peu exploré jusqu'à maintenant, même en anglais langue seconde. Van Zeeland et Schmitt (2013) déplorent la tendance des chercheurs à utiliser les découvertes faites dans les études sur le vocabulaire et la lecture pour les transférer à la compréhension orale, sans tenir compte de la spécificité de chacune des habiletés. Par la présente étude, nous tenterons de remédier à cette lacune.

Dans cette recherche, nous nous intéresserons particulièrement aux étudiants du niveau 6 qui est, selon le programme du MELS, le dernier niveau du tronc commun. Les étudiants qui le complètent visent habituellement l'un des trois objectifs suivants : poursuivre leurs cours de francisation au niveau écrit, s'inscrire dans un collège ou une université francophone ou se trouver un emploi. Il s'agit donc d'un niveau charnière au terme duquel nous jugeons important de mesurer les acquis des étudiants en terme de vocabulaire. Le fait de connaître la taille du vocabulaire des apprenants et son impact sur la performance aux examens de compréhension orale et de compréhension écrite pourrait aussi avoir des répercussions pédagogiques.

1.6 Questions et objectifs de recherche

Jusqu'à présent, nous avons mis en lumière toute l'importance de la taille du vocabulaire chez les apprenants de langue seconde et le rôle majeur que joue le vocabulaire dans les différentes habiletés langagières. Nous avons aussi souligné

l'importance du niveau 6 dans le cheminement des étudiants qui suivent des cours de francisation. Cet état de fait nous amène à nous poser les deux questions de recherche suivantes :

- Quelle est la taille du vocabulaire réceptif des étudiants adultes de francisation de niveau 6?
- Quelle est la corrélation entre la taille du vocabulaire réceptif des étudiants et leurs résultats aux examens ministériels en compréhension orale et en compréhension écrite?

Notre recherche vise principalement les trois objectifs suivants :

- Déterminer quelle proportion des 3000 premiers mots les plus fréquents est connue par les étudiants adultes qui terminent le tronc commun de la formation en francisation.
- Déterminer le rôle que joue la connaissance ou le manque de connaissance de vocabulaire des étudiants dans leur compréhension orale et leur compréhension écrite.
- Déterminer si certains facteurs sociodémographiques ont une influence sur la taille de vocabulaire et les résultats aux tests mesurant les habiletés réceptives.

1.7 Conclusion

Dans ce premier chapitre, nous avons établi le contexte général dans lequel s'insère notre recherche pour ensuite présenter nos questions et nos objectifs de recherche. Le prochain chapitre sera consacré au cadre théorique, dans lequel nous définirons les concepts de notre étude grâce à une revue de la littérature sur le sujet.

CHAPITRE II

CADRE THÉORIQUE

2.1 Introduction

Dans ce chapitre, nous définirons les concepts-clés de la recherche en vocabulaire : la définition de *mot*, le sens de *connaître un mot*, la taille du vocabulaire, les facteurs qui influencent l'apprentissage et la fréquence des mots, la nature des liens qui unissent le vocabulaire et les habiletés linguistiques réceptives et l'enseignement du vocabulaire.

2.2 Qu'est-ce qu'un mot?

Le vocabulaire, comme nous l'avons démontré précédemment, joue un rôle central dans l'apprentissage d'une langue seconde. Comme le dit David (2008): « Vocabulary is one of the building blocks of language, one of the basics of communication and a necessary component in learners' development. » (p. 167)

Lorsque nous parlons de la taille du vocabulaire, nous faisons référence au nombre de mots connus par un étudiant. La question qui se pose alors est : qu'est-ce qu'un mot? Milton (2009) présente quatre façons différentes de définir les mots.

D'abord, on peut envisager les mots en terme de *tokens*, lesquels correspondent au nombre total de mots présents dans un texte, peu importe qu'ils se répètent ou non. Cette conception n'est pas pertinente pour l'évaluation de la taille du vocabulaire, car elle n'indique pas le nombre de mots différents que l'étudiant connaît.

Une autre conception du mot est les types, qui tiennent seulement compte des mots de formes différentes. Cela signifie que, si un mot se répète dans un texte, il ne sera compté qu'une seule fois. Cette définition est déjà plus appropriée, mais elle peut être simplifiée. En effet, avec les types, les mots *enfant* au singulier et *enfants* au pluriel seraient considérés comme deux mots différents. Or, une fois apprise, la règle de l'ajout de la lettre *s* pour former le pluriel permet facilement de reconnaître une grande quantité de mots au pluriel sans qu'ils aient besoin d'être considérés comme des mots différents de leur forme au singulier. À ce sujet, Milton (2009) dit: « it makes sense, to assume, for most learners that if one form of a word is known, then other, very common derivations and inflections will also be known. » (p. 9) En conséquence, il est possible de concevoir les mots en terme de lemmes, lesquels comprennent un mot de base et ses formes fléchies les plus fréquentes : marques du féminin, du pluriel, ou forme conjuguée dans le cas d'un verbe. Cependant, tous les mots qui forment un lemme doivent être de même nature, par exemple seulement des noms, des adjectifs ou des verbes. Ainsi, le lemme du mot *grand* comprendrait donc les adjectifs *grands*, *grande* et *grandes*. Il faut noter que, pour que les différents mots faisant partie d'un même lemme soient reconnus, la base du mot doit être aisément identifiable. Par exemple, il n'est pas évident qu'un apprenant fera le lien entre les adjectifs *vieux* et *vieille*, même s'ils font partie du même lemme. Il en va de même pour des verbes très irréguliers dont la base se transforme. Un étudiant pourrait connaître le verbe *être*, mais ne pas pouvoir l'identifier dans sa forme au futur simple : *je serai*. Malgré ces bémols, les lemmes sont souvent utilisés comme unité de mesure par les chercheurs pour évaluer la taille du vocabulaire des apprenants de langue

seconde. C'est le cas de Milton et Meara dans la construction du X-Lex, le test de vocabulaire qui sera utilisé dans notre recherche et que nous décrirons dans le chapitre portant sur la méthodologie.

Si on choisit d'inclure, en plus des formes fléchies, les dérivations d'un mot, on parle alors d'une famille de mots (Bauer et Nation, 1993). Contrairement aux lemmes, les familles de mots peuvent inclure des mots de nature différente. Par exemple, si nous constituons une famille à partir du mot *société*, nous incluons le mot *sociétés*, mais également *social*, *socialement*, *socialiser*, *socialisme*, *sociologie*, *asocial*. D'après Milton (2009), l'utilisation des familles de mots dans les tests qui mesurent la taille de vocabulaire réceptive convient seulement aux apprenants avancés ou aux locuteurs natifs. Il est vrai que l'apprentissage des dérivations à partir d'un mot de base est un processus à long terme. Suivant une étude de Schmitt et Zimmerman (2002), ce n'est pas parce qu'un étudiant connaît un membre d'une famille de mots qu'il connaît automatiquement tous les autres. Les règles régissant les dérivations sont irrégulières et le sens des mots dérivés n'est pas toujours transparent. Il existe aussi des préfixes et des suffixes plus rares et moins productifs. Dans certains cas, il est nécessaire d'avoir des connaissances approfondies sur la langue et sur sa morphologie pour déduire le sens de certains mots d'une même famille. Par exemple, peu d'apprenants pourront voir un lien entre le nom commun *chien* et l'adjectif *canin*. Évidemment, dans certains cas, les mots d'une même famille peuvent être facilement reconnus. Même un étudiant débutant pourra faire le lien entre le nom *danse* et le verbe *danser*. Dans le cadre de cette recherche, nous utiliserons les lemmes plutôt que les familles, car les tests de vocabulaire destinés aux apprenants de langue seconde utilisent cette unité. De plus, nous croyons que les étudiants ne maîtrisent pas suffisamment la langue pour avoir une connaissance exhaustive des familles de mots.

Maintenant, il faut faire la distinction entre les mots de fonction et les mots de contenu. Les mots de fonction comprennent les déterminants, les prépositions ou les pronoms et sont habituellement très utilisés. Par contre, leur connaissance n'affecte pas énormément la compréhension globale d'un texte. Même si le mot *et* se répète de nombreuses fois dans un texte courant, il contribue très peu à la construction du sens. C'est pourquoi on considère souvent les mots de fonction dans une classe à part, que Milton définit comme le niveau 0. Dans le calcul de la taille du vocabulaire, les mots de contenu sont plus pertinents. Les noms, les verbes, les adjectifs et les adverbes véhiculent beaucoup de sens, même s'ils sont présentés de façon isolée. En conséquence, notre recherche s'intéressera seulement aux mots de contenu.

Un autre point important à aborder lorsque l'on parle du sens des mots est la question des sens multiples. Comme Laufer (1997) l'explique, un même mot peut avoir plusieurs significations. Lorsque les différents sens sont reliés, on parle de polysémie. Par exemple, le mot *artère* peut être, au sens propre, une partie du système circulatoire ou, au sens figuré, une voie de communication. En revanche, si les sens n'ont pas de lien entre eux, il s'agit d'homonymie. Ainsi, le mot *grève* a deux sens différents, non reliés et correspondant à deux entrées dans le dictionnaire. On peut *marcher sur la grève* ou *faire la grève*. Évidemment, la distinction n'est pas toujours aussi simple, même pour les spécialistes. Laufer (1997) le confirme: « it's hard to distinguish which meanings are related and which are not. » (p. 152) Ceci dit, notre étude ne demande pas de faire une telle distinction. Nous avons choisi de mesurer le vocabulaire à l'aide du X-Lex, un test présentant une série de mots isolés de leur contexte. Ce type de test présente l'avantage de pouvoir tester beaucoup de mots, mais comporte aussi des limites. Une de ces limites concerne les mots à sens multiples. Si certains mots du test ont plusieurs significations possibles, il est impossible de savoir lesquelles sont familières à l'apprenant.

Lorsqu'il est question des *mots* en tant qu'unité porteuse de sens, nous devons souligner l'existence des unités à plusieurs mots, ce que la littérature anglophone désigne comme *multi-word items*. Les unités à plusieurs mots sont définies par Moon (1997) comme étant une séquence de deux mots ou plus, formant, sémantiquement, psychologiquement, ou syntaxiquement, un tout significatif dont les éléments sont inséparables (p. 43). Par exemple, en français, *pomme de terre* est composé de trois mots, mais désigne une seule réalité. La connaissance de chacun des mots de cette unité lexicale ne permet pas de comprendre sa signification. Il en va de même pour des unités à plusieurs mots comme *bleu marine*, *mise à jour*. Même si ce phénomène fait partie de la langue française et représente un intérêt, les unités à plusieurs mots ne seront pas considérées dans cette étude. Nous nous intéressons aux mots fréquents et la majorité des unités à plusieurs mots ne font pas partie de cette catégorie. Après avoir étudié un large corpus, Moon rapporte qu'en anglais, plus de 70% des unités à plusieurs mots apparaissent moins d'une fois par 1 million de mots. Il y a lieu de croire que cette statistique est valable pour les autres langues et peut donc s'appliquer au français. Le choix de ne pas considérer les unités à plusieurs mots est aussi justifié par le fait que les tests qui mesurent l'étendue du vocabulaire, dont le X-Lex, n'incluent pas ce type de mots.

2.3 Qu'est-ce que connaître un mot?

Maintenant que nous avons défini le terme *mot* par rapport à notre recherche, nous souhaitons aborder la question de la connaissance d'un mot. D'après Milton (2009), une des manières de définir la connaissance d'un mot est de la diviser d'une part, en connaissance réceptive ou passive et, d'autre part, en connaissance productive ou active. La connaissance réceptive correspond plutôt à une reconnaissance, car elle implique de reconnaître un mot lorsqu'il est lu ou entendu. La connaissance productive, quant à elle, signifie qu'une personne est capable

d'utiliser un mot de façon appropriée dans un discours oral ou écrit. Nous sommes consciente, comme le souligne Melka (1997), que la ligne qui sépare le réceptif et le productif n'est pas claire et nette et qu'il existe différents degrés de connaissance, qui vont d'une connaissance minimale à une connaissance complète. Ceci dit, pour les besoins de notre recherche, nous maintiendrons la distinction entre l'actif et le passif. La taille de ces deux types de vocabulaire, réceptif et productif, peut être mesurée à l'aide d'instruments objectifs. Les recherches se concentrent davantage sur le vocabulaire réceptif car il est, selon David (2008), plus facile à mesurer et à analyser, spécialement dans le contexte des études portant sur l'étendue ou la taille du vocabulaire. Effectivement, les tests servant à mesurer le vocabulaire réceptif sont faciles à administrer et à corriger, en plus d'être de bons prédicateurs des compétences langagières générales. Pour toutes ces raisons, notre recherche sera axée sur la taille du vocabulaire réceptif.

Lorsqu'il est question de la connaissance lexicale, il faut également établir la différence entre la taille du vocabulaire et la profondeur du vocabulaire. La taille fait référence au nombre de mots connus par une personne. Dans ce cas, il est possible que la connaissance soit obscure ou partielle. En revanche, la profondeur correspond à ce qu'un étudiant connaît à propos des mots. Reconnaît-il simplement la forme du mot ou peut-il en donner une définition précise? Peut-il l'utiliser adéquatement dans un contexte approprié? Bien que fort intéressantes, ces questions liées à la profondeur du vocabulaire ne seront pas traitées dans notre recherche. Les tests qui mesurent la profondeur du vocabulaire exigent des questions plus détaillées et permettent de tester moins de mots. Étant donné que nous préférons évaluer un grand nombre de mots et ainsi obtenir un portrait plus précis de la situation, nous avons choisi de nous intéresser seulement à la taille du vocabulaire.

2.4 La taille du vocabulaire

D'après Nation et Waring (1997), un locuteur natif de l'anglais de niveau universitaire a un lexique réceptif approximatif de 20 000 familles de mots. À son entrée à l'école, un enfant de cinq ans possède un vocabulaire d'environ 4000 à 5000 familles de mots et, chaque année, il ajoute à son lexique environ 1000 familles. Le portrait est différent lorsqu'il est question de la langue seconde. Nation et Waring soulignent que le fossé est habituellement grand entre le vocabulaire des apprenants adultes d'anglais langue seconde et celui des locuteurs natifs. En se basant sur leur vaste expérience, les auteurs mentionnent que plusieurs étudiants présentent un vocabulaire de moins de 5000 mots, malgré plusieurs années d'apprentissage de la langue. « Although the goal of a native speaker vocabulary size is a possible goal, it is very ambitious for most learners of English as a second language. » (Nation et Waring, 1997, p. 8)

Un étudiant de niveau 6 qui aurait suivi toute la formation en francisation depuis le niveau 1 aurait un an d'étude intensif du français à son actif. Chaque niveau dure en moyenne huit semaines, à raison de 4 à 6 heures de cours par jour, selon que l'étudiant reçoit un financement gouvernemental ou non. Par conséquent, chaque étudiant reçoit de 160 à 240 heures d'enseignement pour chaque niveau complété. Par exemple, un étudiant ayant étudié six heures par jour pour les niveaux 1 à 5 aurait, une fois arrivé au niveau 6, reçu 1200 heures d'enseignement du français. Il s'agit d'une somme de temps considérable pour un adulte qui se consacre à l'étude d'une langue seconde. Cependant, est-ce suffisant pour acquérir un niveau de vocabulaire permettant d'être fonctionnel? Par ses recherches, Milton (2009) a évalué que des étudiants de langue seconde apprenaient environ 4 mots par heure de classe. Il s'agit, selon l'auteur, d'une moyenne très générale pouvant varier selon les individus et le

type d'enseignement. Si nous appliquions ce principe à notre étude, un étudiant de niveau 6 ayant complété les niveaux 1 à 5 aurait un vocabulaire d'environ 4800 mots. Il reste à déterminer si cette évaluation est réaliste, considérant que les évaluations de Milton se basent sur des apprenants enfants et adolescents qui étudient le français pendant plusieurs années.

Une autre étude portant sur la croissance du vocabulaire a été menée par Milton et Meara (1995), cette fois sur de jeunes adultes dans un contexte d'immersion. Les auteurs se sont intéressés à l'augmentation de la taille du lexique de 53 universitaires européens participants à un échange étudiant en Angleterre. Ils ont découvert que le vocabulaire de ces étudiants avait augmenté en moyenne de 1325 mots en six mois, avec beaucoup de variations individuelles cependant. Si ce rythme d'apprentissage se poursuivait, il serait donc possible d'apprendre 2600 nouveaux mots en une année. La question de l'immersion nous intéresse, car les participants de notre étude vivent au Québec, donc dans un milieu majoritairement francophone, ce qu'on peut certainement considérer comme un contexte d'immersion. En effet, les adultes en francisation ne sont pas en contact avec le français strictement durant leurs cours. Ils entendent le français dans les rues, dans les commerces, à la radio et à la télévision et le voit sur les affiches, dans les journaux ou dans les communications écrites. Un tel milieu, s'il est exploité par l'étudiant, favorise certainement l'apprentissage plus rapide de la langue.

Peu d'études ont été menées sur la taille du lexique des étudiants de français langue seconde. Comme nous l'avons vu au chapitre précédent, Milton (2006) a publié une des rares recherches à ce sujet, dans laquelle il a évalué le vocabulaire réceptif des étudiants de français langue seconde de niveau secondaire en Angleterre. Il a découvert qu'un étudiant de dix-huit ans à la fin du secondaire ayant étudié le français pendant sept ans possède un lexique d'environ 2000 mots, résultat considéré

comme très modeste par le chercheur. En effet, comme nous l'avons vu un peu plus tôt, l'objectif à atteindre pour posséder une bonne base de vocabulaire serait plutôt de 3000 mots (Nation et Waring, 1997).

2.5 Les facteurs influençant l'acquisition du vocabulaire

L'analyse faite dans la section précédente nous indique que le nombre de mots de vocabulaire que les étudiants parviennent à acquérir reste à déterminer. Comme nous l'avons vu précédemment, la règle générale veut que les mots les plus fréquents soient appris en premier. Il ne s'agit cependant pas d'une règle absolue, car d'autres facteurs influencent l'acquisition du vocabulaire. Certains mots, même s'ils sont fréquents, peuvent être plus difficiles à apprendre que d'autres. Laufer (1997) mentionne plusieurs éléments pouvant rendre un mot plus ardu à mémoriser. D'abord, si le mot contient des lettres ou des sons inhabituels pour l'étudiant, ce dernier aura de la difficulté à le retenir. Ensuite, il est évident qu'un nouveau mot sera moins facilement retenu lorsqu'il prend une forme très différente de sa traduction dans la langue maternelle de l'apprenant. À l'inverse, les congénères, c'est-à-dire les mots qui sont similaires dans la langue cible et la langue première, seront rapidement acquis. Pour un étudiant de français dont la langue première est l'anglais, il est plus facile d'apprendre le mot *aéroport* que le mot *gare*. Une autre difficulté citée par Laufer est la question des mots qui ont une forme très similaire dans la langue cible. Si un nouveau mot ressemble trop à un autre déjà appris, cela pourrait occasionner de la confusion. Par exemple, un apprenant de français pourrait confondre *manteau* et *menton* ou encore *jeune* et *jaune*. En outre, les études démontrent que le vocabulaire référant à des réalités concrètes sera mieux appris que le vocabulaire abstrait. Finalement, il semble que les mots plus courts causent moins de problèmes aux étudiants que les mots plus longs.

En plus des considérations liées aux mots eux-mêmes, des facteurs externes jouent également un rôle important. Évidemment, la question du temps est primordiale. En effet, les adultes suivant la formation du niveau 1 à 6 n'ont habituellement que douze mois à leur disposition pour construire une solide base de vocabulaire. O'Dell (1997) affirme que la répétition et la révision tiennent un rôle important dans l'apprentissage des mots. L'auteure rappelle la recherche de Kachroo (1962), qui conclut que les mots doivent être répétés 7 fois ou plus dans le manuel utilisé par les étudiants pour être appris. Or, les programmes du MELS sont très chargés et seule une petite portion du matériel écrit et audio traite du même thème. Il est loin d'être assuré que le nouveau vocabulaire sera suffisamment recyclé pour être acquis.

Dans un même ordre d'idées, O'Dell écrit qu'un nouveau mot est mieux appris s'il est revu plusieurs fois, en espaçant de plus en plus les intervalles de temps. Elle suggère de revoir le mot 5 à 10 minutes après le cours, 24 heures après, une semaine après, un mois après et, finalement, six mois après. Ce principe s'applique très difficilement dans le cadre actuel des cours de francisation. Effectivement, le temps dont dispose les enseignants pour enseigner le programme d'un niveau est plutôt limité et ne leur permettent pas toujours d'effectuer une révision systématique de tous les nouveaux mots appris. De plus, d'un niveau à l'autre, c'est-à-dire à tous les deux mois, les étudiants changent de classe et de professeur, ce qui implique que le suivi du vocabulaire ne peut pas être fait systématiquement.

Un autre facteur qui pourrait nuire à l'acquisition des mots fréquents est la forme du programme privilégié par le MELS. En effet, le Ministère privilégie une approche thématique (topic-based). Chaque niveau aborde quelques thématiques précises liées à des situations de vie concrètes : *Je magasine; Je cherche un logement; Je m'occupe de ma santé; Je m'informe...* Or, selon O'Dell (1997), « While topics certainly are an

important aspect of vocabulary work, vocabulary teaching that is only topic-based can lead to the exclusion of many important items. » (p. 270) David (2008) abonde en ce sens en disant que les thèmes étudiés en classe peuvent amener les étudiants de langue seconde à apprendre beaucoup de mots qui, sans être complètement inutiles, ne sont pas fréquents dans la langue cible. Par exemple, en traitant le thème de l'alimentation, les étudiants pourraient apprendre les mots *brocoli* ou *aubergine*, lesquels ne sont pas utilisés fréquemment en français par les locuteurs natifs. Il est donc possible que les étudiants de francisation présentent certaines lacunes lexicales dans les niveaux les plus fréquents.

2.6 La fréquence des mots

La langue contient une grande quantité de mots, mais tous ne sont pas utilisés à la même fréquence. Des chercheurs ont constitué des listes de fréquence à partir de très larges corpus comptant des millions de mots et constitués de différents types de textes. Ces listes considèrent la fréquence, mais également la portée (*range*) dans le corpus. Cela veut dire que, pour faire partie d'une liste de fréquence, un mot doit non seulement être fréquent, mais également l'être dans un grand nombre de types de textes. Ces listes ont permis aux chercheurs de constater qu'un petit nombre de mots sont utilisés de manière plus fréquente. Le fait de connaître ces mots de haute fréquence permet de couvrir une grande partie d'un texte et de le comprendre. Nous avons dit dans le chapitre précédent qu'un lexique de 2000 familles de mots permettait de couvrir 80% d'un texte authentique en anglais (Nation et Waring, 1997). Ce pourcentage, bien qu'important, s'avère être insuffisant pour comprendre adéquatement un texte, car un mot sur cinq demeure inconnu. Dans le cas de la compréhension orale, les 2000 premiers mots de la langue offrent une couverture encore plus grande qu'à l'écrit, soit environ 90% (Adolphs et Schmitt, 2003). Ces pourcentages peuvent-ils s'appliquer à la langue française? La distribution lexicale du

français est semblable à celle de l'anglais, puisque ces deux langues utilisent beaucoup les mots les plus fréquents du niveau 1000 et 2000 (Cobb et Horst, 2004). Cependant, dans le cas de textes écrits, il semble que le français utilise encore davantage ses 2000 premières familles de mots les plus fréquentes. Celles-ci permettraient à l'apprenant d'atteindre une couverture de 90%. Cobb et Horst le confirment: « In almost all of the cases we examined, it appears that the goal of achieving 90% text coverage can be met by mastering the common vocabulary of French. » (p. 25) Lorsque les auteurs parlent de *vocabulaire commun*, ils font référence aux 2000 premières familles de mots les plus fréquentes.

Nation et Waring (1997) s'interrogent sur un élément important pour la didactique des langues, en l'occurrence le nombre de mots que les étudiants de langue seconde devraient apprendre. Les deux chercheurs considèrent comme prioritaire l'apprentissage des 3000 mots les plus fréquents de la langue. Les mots de basse fréquence ne permettent pas d'accroître de façon significative la compréhension, alors il n'est pas rentable de les enseigner en classe. Pour apprendre les mots après le niveau 3000, les auteurs considèrent plus importants d'enseigner des stratégies comme l'utilisation du contexte, l'utilisation de trucs mnémotechniques ou l'utilisation de cartes de vocabulaire associant le mot en langue première et le mot en langue seconde. Ces stratégies permettent aux étudiants d'apprendre de nouveaux mots de façon plus autonome et d'augmenter la taille de leur vocabulaire.

2.7 Le vocabulaire et la lecture

La plupart des études portant sur la relation existante entre la taille du vocabulaire et la compétence langagière en langue seconde portent sur la lecture. Staehr (2008) et Van Zeeland et Schmitt (2013) rappellent qu'un grand nombre de recherches antérieures démontrent une corrélation élevée entre le vocabulaire réceptif et la

compréhension écrite ($r = 0,40$ à $0,85$). Milton cite une recherche de Laufer et Sim (1985), lesquels se sont intéressés aux connaissances requises par les étudiants pour comprendre adéquatement un texte de niveau académique dans l'examen du First Certificate in English de l'Université Cambridge : « The study reached the conclusion that vocabulary knowledge is the most important area of knowledge required for comprehension; more important than knowledge of the subject and more important than syntactic knowledge. » (Milton, 2006, p. 50)

À l'instar des autres chercheurs, Staehr a conclu que la compréhension écrite était l'habileté la plus dépendante de la taille du vocabulaire. La corrélation entre le résultat au test de vocabulaire et le résultat au test de lecture est très élevée ($r = 0,83$). De plus, l'analyse statistique démontre que 72% de l'habileté à obtenir un résultat au-dessus de la moyenne dans le test de lecture pouvait s'expliquer par la performance au test de vocabulaire. Il suggère également que la connaissance de 2000 mots constitue une sorte de seuil qui prédit une performance adéquate au test de compréhension écrite.

Plusieurs chercheurs dans le domaine du vocabulaire ont porté leur attention sur l'étendue, c'est-à-dire le pourcentage de mots devant être connu pour assurer la compréhension d'un texte authentique. Laufer (1992) a évalué que 95% des mots devaient être connus pour que l'étudiant démontre une compréhension raisonnable, laquelle correspond à une note minimale de 55% à un examen de compréhension de lecture. Pour atteindre cette étendue de 95%, un apprenant devrait connaître 3000 familles de mots. Hu et Nation (2000), quant à eux, croient qu'une couverture de 98% est nécessaire pour être en mesure de lire pour le plaisir. Une telle étendue impliquerait un vocabulaire beaucoup plus vaste, soit de 8000 à 9000 familles de mots dans le cas de l'anglais. Les chercheurs Schmitt, Jiang et Grabe ont conclu pour leur part qu'il n'existe pas de seuil de compréhension. D'une étendue de 90% à une

étendue de 100%, la compréhension augmente de façon linéaire sans faire de bond spectaculaire ou franchir un seuil à un moment ou un autre. Malgré cela, les auteurs considèrent que pour obtenir une compréhension jugée adéquate, c'est-à-dire supérieure à 60%, la connaissance de 98% des mots d'un texte était nécessaire. Ces questions ne constituent pas l'objet de ce mémoire, mais demeurent très importantes pour décrire le contexte de notre recherche.

2.8 Le vocabulaire et la compréhension orale

Les recherches sur le rôle du vocabulaire dans la compréhension orale sont beaucoup moins nombreuses que celles concernant la lecture. Les résultats de ces études démontrent qu'un lien existe entre le lexique et la compréhension orale, mais qu'il est moins important que dans la compréhension écrite. Notre expérience en tant qu'enseignante et aussi apprenante d'une langue seconde nous permet d'affirmer que le rôle du vocabulaire diffère à l'oral et à l'écrit. Un mot écrit est plutôt stable dans sa forme et peut habituellement être reconnu d'un texte à l'autre. En revanche, à l'oral, il est possible qu'un apprenant ne reconnaisse pas certains mots en raison d'un débit un peu trop rapide, d'une prononciation moins claire ou d'un accent différent de ce qu'il a l'habitude d'entendre. Le discours oral pourrait aussi dépendre moins du vocabulaire car il contient davantage d'indices contextuels que l'écrit.

Une des recherches menée sur le vocabulaire et la compréhension orale est celle de Bonk (2000). Ce dernier s'est intéressé à la compréhension orale d'apprenants d'anglais langue seconde de niveau universitaire. Ses analyses démontrent qu'il existe une corrélation significative ($r = 0,45$) entre la proportion de mots connus dans un passage oral et la compréhension de ce passage. Cependant, cette corrélation n'est pas élevée : certains étudiants maîtrisant 100% du vocabulaire ne parviennent pas à obtenir une bonne compréhension, alors que d'autres connaissant 75% du lexique

arrivent à comprendre adéquatement. L'auteur explique ce phénomène par le fait que l'écoute ne se limite pas à la compréhension de mots isolés, mais à des groupes de mots connectés les uns aux autres.

Staeher (2008) pour sa part, a observé une corrélation significative, mais modérée ($r = 0,69$) entre la taille du vocabulaire des étudiants et le résultat à un test de compréhension orale. Tous les participants à la recherche dont la taille du vocabulaire était d'au moins deux mille mots ont obtenu une note au-dessus de la moyenne au test d'écoute. Cependant, l'auteur note également que 65% des étudiants dont la taille du vocabulaire était inférieure à 2000 mots ont réussi à obtenir un résultat égal ou supérieur à la moyenne du groupe. Staeher explique ce phénomène par plusieurs raisons. D'abord, il a mesuré la taille du vocabulaire par un test de reconnaissance écrite, alors que l'écoute demande une reconnaissance de la forme orale des mots. De plus, la qualité de la compréhension orale ne dépend pas seulement de la connaissance des mots, mais également à la capacité de l'apprenant de traiter rapidement les informations. Cette capacité, aussi connue sous le terme anglais de *word recognition*, implique un accès rapide et automatique au vocabulaire. De plus, plusieurs autres facteurs, en plus du vocabulaire, jouent un rôle dans la qualité de la compréhension orale des étudiants. L'auteur mentionne entre autre la compétence stratégique comme un facteur important de la performance au test d'écoute. Bref, le vocabulaire n'est pas le seul facteur responsable de la variance en compréhension orale. Cependant, il s'agit du facteur le plus facilement identifiable et du plus enseignable.

Finalement, Van Zeeland et Schmitt (2013) se sont intéressés au pourcentage de mots que les étudiants devraient connaître pour atteindre un bon niveau de compréhension orale. Ils ont choisi quatre courtes histoires présentant le même niveau de difficulté. Ensuite, ils ont remplacé un certain nombre de mots par des non

mots pour simuler une couverture de 100% (tous les mots sont connus), 98%, 95% et 90%. Après l'écoute des quatre passages, 28 étudiants d'anglais langue seconde, pour la plupart avancés, ont effectué un test de compréhension orale de type objectif. Pour les questions concernant le passage dans lequel 100% des mots étaient connus, la majorité des étudiants ont obtenu une note parfaite ou très élevée. Avec une couverture de 98%, le résultat moyen au test était de 8 sur 10. En ce qui concerne les histoires présentant une couverture de 95% et 90%, les notes moyennes au test de compréhension étaient respectivement de 7,65 sur 10 et de 7,35 sur 10. Ces deux moyennes sont très similaires, mais les chercheurs ont remarqué plus de variation dans la compréhension de l'histoire présentant une couverture de 90% (écart-type = 2,13) que dans celle de 95% (écart-type = 1,29). À cause de cette variation, Van Zeeland et Schmitt considère qu'une couverture de 95% constitue le seuil pour assurer la compréhension orale. Cette étude démontre que la connaissance du vocabulaire a effectivement un impact sur le niveau de compréhension orale des étudiants. Le fait de connaître une plus grande proportion de vocabulaire augmente la compréhension, avec des variations selon les individus. Les auteurs remarquent que les participants ont été affectés différemment par les non mots : certaines personnes sont plus sensibles que d'autres à la présence de mots inconnus.

2.9 L'enseignement du vocabulaire

Jusqu'à maintenant, les recherches démontrent qu'en général, plus on connaît de mots, plus la compréhension des textes écrits et oraux est facilitée. Il reste donc à savoir comment l'enseignant peut arriver à faire connaître davantage de mots aux étudiants. L'approche communicative en vogue dans les années 1970 préconisait l'approche implicite, c'est-à-dire d'acquérir le vocabulaire en devinant les mots grâce au contexte. De plus en plus de recherches démontrent l'inefficacité d'un enseignement strictement implicite du vocabulaire. Sökmen (1997) souligne plusieurs

problèmes avec une telle méthode : la lenteur du processus, le risque d'erreur et le manque de rétention à long terme. Bien entendu, il ne s'agit pas d'éliminer toute forme d'apprentissage implicite, mais d'y ajouter une part d'enseignement explicite.

Sökmen mentionne plusieurs recherches qui suggèrent qu'il est utile d'ajouter un enseignement explicite du vocabulaire. L'auteure recommande aux enseignants de langue seconde de porter une attention particulière aux 2000 premiers mots les plus fréquents, car ils sont très productifs. En revanche, un cours de langue ne peut pas seulement s'organiser autour des mots fréquents. Il est important de s'attarder aussi aux mots plus difficiles que les étudiants évitent d'utiliser, ainsi qu'aux mots correspondant aux besoins et aux intérêts spécifiques des apprenants.

Pour faciliter l'acquisition du vocabulaire, Sökmen présente plusieurs principes documentés par diverses recherches. Elle cite Nation (1982), qui a prouvé que l'utilisation de listes de mots peut être une manière d'apprendre rapidement une bonne quantité de vocabulaire, à condition que les mots soient ensuite réutilisés dans un contexte plus authentique. Cette méthode serait même particulièrement appropriée pour les débutants. Un autre principe important devant guider l'enseignant est de faciliter l'intégration des nouveaux mots aux anciens, par l'utilisation de cartes sémantiques par exemple. De plus, le professeur doit s'assurer que les nouveaux mots sont présentés plusieurs fois dans des contextes différents. La répétition, jumelée aux légères variations de sens que prennent les mots dans les contextes différents, sont des clés de l'apprentissage. Il est également suggéré de rendre le vocabulaire le plus concret possible, par l'utilisation d'images ou de diagrammes. Finalement, le fait d'utiliser une variété de techniques semble la façon la plus efficace de favoriser l'acquisition du nouveau vocabulaire.

Évidemment, il n'est pas possible qu'un étudiant apprenne tout le vocabulaire dont il a besoin en classe. Il est donc important qu'il soit impliqué dans son apprentissage et l'enseignant peut l'aider en ce sens en l'aidant à développer des stratégies. Schmitt (1997) a recensé des recherches démontrant que certaines stratégies augmente la rétention des mots: la formation d'associations entre les mots, la *Keyword method* et même la mémorisation par cœur. Les stratégies évoluent également avec le niveau des étudiants : celles de surface conviendraient davantage aux débutants, alors que les stratégies en profondeur seraient plus appropriées aux étudiants intermédiaires et avancés, lesquels ont la capacité d'utiliser le contexte.

2.10 Conclusion

Le chapitre qui s'achève nous a permis de définir les concepts importants de notre cadre théorique, en particulier celui de la taille du vocabulaire et de ses relations avec les habiletés réceptives. Dans le chapitre suivant, nous présenterons la méthodologie utilisée dans notre recherche.

CHAPITRE III

MÉTHODOLOGIE

3.1 Introduction

Dans ce chapitre, nous décrirons la population de notre recherche et les instruments de mesure choisis, en l'occurrence le questionnaire sociodémographique, le test de vocabulaire et les examens ministériels en compréhension orale et en compréhension écrite. De plus, nous expliquerons le déroulement de la collecte de données et la technique de traitement des données

3.2 Population

Les participants à notre étude sont tous des adultes étudiants en francisation au niveau 6 au moment de la recherche. Les étudiants fréquentent un établissement d'éducation aux adultes de Montréal et suivent le programme de francisation du Ministère de l'Éducation et des Loisirs (MELS). L'échantillon compte 17 étudiants, 14 femmes et 3 hommes, tous inscrits dans la même classe. Les participants sont âgés entre 28 et 46 ans, pour une moyenne de 35 ans. La langue maternelle déclarée est de la même famille que le français pour 10 personnes (espagnol et roumain). Les autres langues maternelles sont le russe (cinq participants), l'arménien (deux participants),

l'ukrainien (un participant), l'albanais (un participant) et le vietnamien (un participant). Le niveau de scolarité varie également d'un individu à l'autre, mais 10 personnes ont complété des études universitaires. Deux personnes ont fréquenté l'université, mais n'ont pas complété leurs études, tandis que quatre étudiants ont complété une formation professionnelle ou des études collégiales. Un seul participant a déclaré avoir un diplôme d'études secondaires ou moins. Chaque étudiant ayant son propre parcours, il est également difficile d'estimer combien d'heures d'étude du français les étudiants ont à leur actif. Les étudiants qui sont admis au programme de francisation passent un test de classement afin de déterminer leur niveau. Ceux qui ont peu ou pas de connaissances de la langue débutent au niveau 1, mais les autres peuvent être classés dans des niveaux supérieurs. Par la suite, les apprenants étudient le français cinq jours par semaine, six heures par jour, s'ils bénéficient d'un financement (MICC ou Emploi-Québec). Les étudiants qui ne sont pas éligibles au financement assistent normalement à quatre heures de cours par jour. Une session dure en moyenne huit semaines, ce qui représente de 160 à 240 heures d'enseignement pour chaque niveau complété. Si, par exemple, un étudiant avait étudié six heures par jour pour les niveaux 1 à 5, il aurait, une fois arrivé au niveau 6, reçu 1200 heures d'enseignement du français.

En ce qui concerne le contact avec la langue française à l'extérieur de l'école, il varie également d'une personne à l'autre. L'écoute est l'habileté linguistique la plus pratiquée : 13 personnes écoutent du français hors de la salle de classe et 4 ne le font que parfois, rarement ou jamais. La lecture est également pratiquée par la majorité : 11 participants lisent très souvent ou souvent. Les résultats sont plus partagés lorsqu'il est question de la pratique de la conversation. Neuf personnes affirment parler français très souvent ou souvent, alors que huit ne le parlent que parfois, rarement ou jamais. L'écriture est l'habileté la moins pratiquée : sept étudiants déclarent écrire très souvent ou souvent.

3.3 Description des instruments de mesure

Pour les fins de notre recherche, nous avons choisi trois instruments de mesure : un questionnaire sociodémographique, un test écrit de vocabulaire et les examens finaux de niveau 6 du MELS en compréhension orale et en compréhension écrite. Nous présenterons et expliquerons chacun de ces instruments dans cette section.

3.3.1 Questionnaire sociodémographique

D'abord, nous avons élaboré un court questionnaire permettant d'obtenir des informations complémentaires au sujet des participants (voir annexe B). Les informations recueillies permettront de remplir notre troisième objectif de recherche, en explorant les liens qui existent entre les données sociodémographiques et la taille de vocabulaire, la performance à l'examen de compréhension orale et celle à l'examen de compréhension écrite. Ces analyses seront effectuées au chapitre suivant. Avant tout, nous allons expliquer et justifier les informations demandées aux participants dans le questionnaire, qui comprend cinq parties : sexe, âge, langue maternelle, scolarité complétée et contact avec la langue française.

Comme notre échantillon est très petit, nous sommes consciente que les résultats des analyses concernant les facteurs sociodémographiques ne seront pas représentatifs de la population des étudiants de niveau 6 en général. Cependant, nous croyons que le fait de recueillir de telles informations pourra être très utile pour d'éventuels prolongements ou des reproductions de cette étude. Ces données seraient aussi d'une grande importance dans le cadre d'une méta-analyse, jumelées avec d'autres provenant d'études sur le même sujet.

Premièrement, le sexe des participants est une variable intéressante à explorer dans toute recherche en éducation. Le lien entre le sexe et la réussite scolaire a déjà fait l'objet de nombreuses études. Par exemple, le programme PISA (Programme for International Student Assessment) a mené une étude en 2009 dans les pays membres de l'OCDE. Dans l'examen de lecture, les filles ont obtenu des résultats supérieurs aux garçons dans tous les pays, avec 39 points de plus en moyenne. Comme la lecture est une des habiletés clés dans l'apprentissage d'une langue, les femmes se trouvent nettement avantagées. Dans le cadre de la même étude, on apprend que les femmes persévèrent plus longtemps que les hommes dans leurs études. Dans les pays de l'OCDE, on estime que 66% des jeunes femmes fréquenteront l'université, en comparaison avec 52% des jeunes hommes (OCDE, 2012). Il serait donc intéressant de vérifier si l'avance des femmes se confirme également chez les étudiants de francisation.

Deuxièmement, l'âge des étudiants présente un intérêt. Les chercheurs se sont souvent penchés sur les différences entre adultes et enfants dans l'apprentissage d'une langue seconde. Mais qu'en est-il de la différence entre les jeunes adultes et les adultes plus âgés? Notre expérience en enseignement nous démontre que les étudiants plus jeunes ont tendance à apprendre une nouvelle langue plus rapidement et plus facilement. Ils possèdent généralement une meilleure mémoire, en plus de prendre plus de risques et de faire preuve d'une plus grande ouverture. Ces caractéristiques constituent des avantages dans l'acquisition d'une langue.

Troisièmement, nous sommes d'avis que la langue maternelle des étudiants a un rôle à jouer dans l'apprentissage du français. Étant donné que le français est une langue latine, elle a beaucoup de vocabulaire en commun avec les autres langues latines, c'est-à-dire l'espagnol, le portugais, le roumain et l'italien. Les locuteurs d'une langue latine possèdent donc un grand avantage lorsqu'ils apprennent le

français : ils reconnaissent spontanément une grande quantité de mots sans avoir à les apprendre. Ils progressent généralement plus rapidement que les étudiants dont la langue maternelle est plus éloignée. En conséquence, nous croyons que les résultats des participants de langue maternelle latine seront supérieurs aux autres.

Quatrièmement, le niveau de scolarité atteint par l'étudiant nous semble pertinent dans le processus d'apprentissage d'une langue. Plusieurs facteurs font que les personnes plus scolarisées ont tendance à mieux réussir dans l'acquisition de la langue. Les individus scolarisés ont en général une connaissance plus approfondie de leur langue maternelle. Ils sont ainsi en mesure de faire des transferts positifs lorsqu'ils apprennent une langue. De plus, le fait d'avoir fréquenté l'école plus longtemps permet aux gens de développer un plus grand nombre de stratégies d'apprentissage, par exemple des stratégies d'écoute, de lecture, d'étude ou de prise de notes. Ces stratégies sont transférables dans les cours de francisation et favorisent la réussite.

Cinquièmement, nous avons ajouté une partie concernant le contact avec le français hors de l'école. Les étudiants doivent indiquer à quel point ils parlent, écoutent, lisent et écrivent en français : très souvent, souvent, parfois, rarement/jamais. La littérature révèle que l'apprentissage indirect par la lecture ou l'écoute hors du contexte scolaire peut être une manière d'apprendre de grandes quantités de vocabulaire. Cependant, ce type d'apprentissage bénéficie en majorité aux apprenants de niveau plus avancé. Si les étudiants ne possèdent pas suffisamment de vocabulaire, ils ne seront pas en mesure d'utiliser le contexte pour apprendre de nouveaux mots (Nation et Waring, 1997). Il s'agit donc de vérifier si les étudiants de niveau 6 sont assez avancés pour tirer profit du contact avec la langue française. Les individus en contact fréquent avec le français ont-ils une taille de vocabulaire plus

grande ou des résultats plus élevés que les autres? C'est la dernière question à laquelle nous souhaitons répondre pour remplir notre troisième de recherche.

3.3.2 Test de vocabulaire

Notre deuxième instrument de mesure est un test de vocabulaire prenant la forme d'une liste à cocher (voir annexe C). Nous avons décidé de mesurer la connaissance des mots écrits provenant des trois premières tranches de mille mots les plus fréquents en français (1k, 2k et 3k). Ces 3000 mots, comme nous l'avons déjà mentionné, sont primordiaux pour former une base de vocabulaire menant l'étudiant vers plus d'autonomie. De plus, compte tenu de la période de temps limitée dédiée à l'étude du français, il nous semble peu probable que les étudiants de niveau 6 possèdent un vocabulaire allant au-delà de 3000 mots. Pour construire notre test, nous nous sommes inspirée du X-Lex de Meara (2005), qui mesure la taille de vocabulaire. Pour développer la version française du X-Lex, Meara a utilisé le corpus de Baudot (1992). Ce corpus contient 22 000 lemmes provenant d'environ 800 échantillons de textes : revues et magazines (42%), livres et manuels (25%), journaux (24%), bulletins et rapports (7%) et brochures et circulaires (2%). Les textes datent des années 1966-1967 et proviennent de la France et du Canada.

Nous avons téléchargé le X-Lex français, mesurant la connaissance des 5000 premiers mots les plus fréquents, disponible en ligne sur le site Lognostics (<http://www.lognostics.co.uk/tools/index.htm>). Nous avons utilisé les listes de mots fournies pour les mots de fréquence 1000, 2000 et 3000. Pour chacune des trois bandes de fréquence, nous avons retenu 40 mots de façon aléatoire. À ces 40 mots, nous avons ajouté 20 non mots plausibles, c'est-à-dire des mots ressemblant à des mots français mais qui n'existent pas. Ces non mots ont également été puisés dans la liste du X-Lex de Meara : par exemple, *crétable* ou *manchir*. Pour chaque niveau de

1000, le test présente donc 60 mots, pour un total de 180 mots. La tâche de l'étudiant est simple : il doit cocher les mots qu'il connaît sur la liste qui lui est présentée. Bien entendu, les étudiants sont avisés que le test contient des mots inventés, ce qui les incite à faire le test avec attention. La présence de non-mots est une manière d'ajuster le résultat des étudiants qui surestiment leurs connaissances ou qui tentent de deviner des mots.

Notre choix s'est arrêté sur la liste à cocher, car elle comporte plusieurs avantages. En plus d'être de toute évidence peu coûteuse, Milton et Read s'entendent pour dire qu'il s'agit de la manière la plus simple de tester le vocabulaire, tant pour le chercheur que pour l'étudiant. Selon Milton (2009), la liste à cocher est relativement rapide et facile à construire et, de surcroît, elle permet de tester rapidement un plus grand nombre de mots que les autres types de tests. L'auteur explique que ce test permet aussi d'obtenir de bons résultats:

The results are likely to be more reliable both because a larger sample size is always likely to give better results than a smaller sample, and because the test can be relatively brief and there is less opportunity for learners to become bored and lose concentration. (Milton, p. 72)

Par contre, Milton reconnaît que le test présente un désavantage : il n'est pas possible de savoir à quel point l'étudiant connaît le mot. Un apprenant qui n'est pas certain de connaître un mot pourrait l'identifier comme connu, et alors surestimer la taille de son vocabulaire. À l'inverse, il pourrait aussi l'identifier comme étant inconnu et cela entraînerait une sous-estimation de la taille du lexique. Une manière de remédier à la surestimation est l'ajout de non mots, lesquels servent à ajuster à la baisse le résultat des participants ayant tendance à cocher des mots qu'ils ne sont pas convaincus de connaître.

Il y a lieu de se demander si l'utilisation d'un test de vocabulaire écrit s'avère pertinent pour prédire les habiletés orales. Un test de vocabulaire oral aurait-il été plus approprié? Nous devons souligner qu'en anglais, il existe une version orale du test de vocabulaire X-Lex : l'Aural Lex (Milton et Hopkins, 2005). L'Aural Lex suit le même principe que le X-Lex : on présente à l'étudiant des mots choisis aléatoirement parmi les 5000 premiers mots les plus fréquents (20 mots par tranche de 1000) et on y ajoute des non mots plausibles. Cependant, plutôt que de présenter les mots par écrit, on les présente oralement et le tout est informatisé. En 2006, Milton et Hopkins ont fait une étude visant à comparer la performance aux deux tests (X-Lex et Aural Lex) afin de vérifier si le test écrit sous-estimait la taille du vocabulaire des étudiants arabes, moins performants en orthographe. La recherche démontre que la taille du vocabulaire écrit et celle du vocabulaire oral sont fortement corrélées. Chez les étudiants grecs, la taille du vocabulaire écrit est supérieure à la taille du vocabulaire oral. Aucune différence significative n'est cependant décelée chez les étudiants arabes. Nous pouvons donc conclure que le X-Lex est un test pouvant mesurer de façon fiable la taille du vocabulaire des apprenants et qu'il est fortement lié à leur connaissance du vocabulaire à l'oral, même si certaines différences liées à la langue maternelle peuvent se présenter.

3.3.3 Examens de compréhension orale et de compréhension écrite

En plus du test de vocabulaire, notre recherche s'intéresse aux résultats des étudiants aux examens finaux de niveau 6 pour les deux compétences réceptives, c'est-à-dire la: compréhension orale et la compréhension écrite. Ces examens ont été conçus par le MELS : ils sont les mêmes à travers tous les centres de francisation pour adultes suivant le programme du Ministère. Tous les examens comprennent quelques versions qui sont utilisées en rotation par les centres de francisation. Pour

des raisons de sécurité, aucune copie d'examen ne doit circuler, ce qui fait que nous ne sommes pas autorisée à présenter les examens dans ce mémoire. Les deux examens mesurant la compréhension sont objectifs, donc composés de questions à choix multiples. Cela signifie que la correction ne laisse place à aucune interprétation subjective de la part de l'enseignant. Il s'agit d'un avantage de taille, car l'objectivité de l'instrument de mesure est ainsi assurée.

Outre leur objectivité, plusieurs raisons justifient le fait d'utiliser les examens ministériels pour notre étude. D'abord, ces examens émanent d'un organisme crédible : ils ont été élaborés par des professionnels et éprouvés durant de nombreuses années. Ils ont été conçus pour mesurer la maîtrise des éléments au programme de chaque niveau, ce qui assure leur fiabilité. De plus, les enseignants et les étudiants ont l'habitude de ces examens, ce qui fait que les résultats ne seront pas influencés par un effet de nouveauté. Un autre élément n'est pas à négliger : le fait d'utiliser un instrument de mesure déjà existant représente une grande économie de temps, tant pour le chercheur que pour les participants. En effet, le chercheur n'a pas à élaborer deux examens, ce qui impliquerait du temps de recherche, la consultation de professionnels et la validation des outils auprès de la population ciblée. Les participants à l'étude, quant à eux, n'ont pas à subir des examens supplémentaires, lesquels rendraient le processus beaucoup plus laborieux. Il y aurait probablement un nombre plus élevé d'abandons. De surcroît, il est évident que les étudiants feront leurs évaluations finales avec sérieux, car elles déterminent leur réussite au cours. Si d'autres examens étaient créés pour les fins de l'étude, il n'est pas certain qu'ils seraient pris aussi sérieusement par les participants. Les résultats des étudiants seraient moins élevés et l'étude entière en serait affectée.

3.4 Déroulement de la collecte de données

Nous avons contacté une enseignante de niveau 6 dans un centre de francisation de Montréal. Cette enseignante, avec l'accord de la direction du centre, a accepté de collaborer à l'étude et de nous recevoir. Le 6 juin 2013, nous nous sommes présentée dans la classe afin de présenter notre projet, de répondre aux interrogations des étudiants et de faire signer les lettres de consentement (voir annexe A). Tous les étudiants de la classe ont démontré de l'intérêt et ont accepté de participer à la recherche. La première collecte de données s'est déroulée le 12 juin 2013, au centre de francisation, directement dans la salle de classe des étudiants. Les personnes présentes étaient les 19 participants, nous-même, en tant que chercheure, et l'enseignante des étudiants. Le déroulement de la recherche a brièvement été rappelé. Les participants ont d'abord rempli le questionnaire sociodémographique. Ensuite, ils ont effectué le test de vocabulaire individuellement, en silence et sans aucun autre matériel (dictionnaire, manuels, téléphone intelligent). Il n'y avait aucune limite de temps pour le test de vocabulaire, mais les étudiants ont pris environ une quinzaine de minutes pour le compléter. La deuxième collecte de données a eu lieu deux semaines plus tard, lorsque les étudiants ont passé leurs examens ministériels en présence de leur enseignante régulière, selon la procédure habituelle. Après avoir corrigé les examens de compréhension orale et de compréhension écrite, l'enseignante nous a communiqué les résultats.

3.5 Traitement des données

Suite à la collecte de données du 12 juin, nous avons d'abord transformé les noms des participants en codes pour assurer l'anonymat. Pour ce faire, nous avons utilisé les trois premières lettres du nom de famille et les trois premières lettres du prénom de chaque étudiant. Ensuite, nous avons entré les données du questionnaire

sociodémographique dans un tableur Excel. Les données qualitatives ont été transformées en données quantitatives. Par exemple, pour la variable sexe, le masculin a été remplacé par 1 et le féminin par 2. Par la suite, nous avons corrigé les tests de vocabulaire et nous avons entré les résultats pour chaque niveau de 1000. Notre manière de corriger sera présentée dans le prochain chapitre. Nous avons calculé la taille du vocabulaire pour chacun des étudiants et ajouté cette variable dans le tableur. Puis, lorsque l'enseignante nous a communiqué les résultats des examens, nous les avons entrés dans notre tableur Excel. Une fois toutes les données entrées, nous avons fait appel au Service de Consultation en Analyse de Données (SCAD) où une statisticienne nous a aidée à effectuer nos analyses grâce au logiciel SPSS.

3.6 Conclusion

Maintenant que nous avons présenté les détails de la méthodologie utilisée dans notre étude, nous pourrions, au chapitre suivant, procéder au traitement de nos données et analyser les résultats obtenus.

CHAPITRE IV

ANALYSE DES RÉSULTATS

4.1 Introduction

Dans ce chapitre, nous procéderons à l'analyse des résultats obtenus lors de la cueillette de données. Nous analyserons donc les résultats des étudiants au test de vocabulaire, les résultats aux examens de compréhension orale et de compréhension écrite et les données recueillies grâce au questionnaire sociodémographique.

4.2 Test de vocabulaire

Lors de la correction du test de vocabulaire X-Lex, il nous a fallu identifier la meilleure manière de calculer les résultats. Même si le X-Lex est largement utilisé dans la recherche en vocabulaire, les chercheurs ne s'entendent pas sur la manière de le noter. Il ne s'agit pas simplement de calculer le nombre de bonnes réponses cochées par l'étudiant : la proportion de non-mots a également un grand rôle à jouer. Comme Beeckmans *et al* (2001) l'affirment, toutes les formules de calcul rencontrées dans la littérature suivent un principe commun : parmi les étudiants ayant coché le même nombre de bons mots, ceux avec une proportion de non mots plus élevé

obtiennent un résultat inférieur (p. 238). Il reste à déterminer à quel point les non mots doivent réduire le score d'un étudiant.

Nous avons d'abord utilisé une formule à laquelle Pellicer-Sanchez et Schmitt (2012) font référence en tant que *correction for guessing formula* (cfg). Cette formule peut être traduite en nos propres termes par la suivante : $cfg = \frac{P(BM) - P(NM)}{1 - P(NM)}$. On calcule le numérateur en calculant le pourcentage de bons mots (BM) identifiés moins le pourcentage de non mots (NM) déclarés comme étant connus. Pour calculer le dénominateur, on soustrait de la proportion de non mots cochés. Cependant, après avoir utilisé cette formule pour calculer les résultats des participants, nous avons observé qu'elle ne permettait pas de faire suffisamment de discrimination entre les étudiants. Schmitt rapporte également ce problème en affirmant que la formule met davantage l'emphase sur les bonnes réponses, appelés *hits*, que sur les non mots cochés, appelés *false alarms* (p. 3).

Suite à ce constat, nous avons choisi d'utiliser une formule qui donne plus de poids aux non mots. Nous avons attribué un point pour chaque bon mot correctement coché par l'étudiant. Puis, nous avons soustrait deux points pour chaque non mot coché. Le raisonnement à la base de cette méthode de calcul est qu'un étudiant qui cocherait tous les mots du test obtiendrait la note zéro ($40x1+20x-2=0$). Le résultat de chaque étudiant, d'abord noté sur 40, a été transformé en pourcentage. Ce pourcentage correspond à la proportion de mots connus pour un niveau de fréquence donné. Par exemple, un étudiant qui obtiendrait une note de 75% au test de 1k connaîtrait donc 750 mots du premier niveau de 1000.

Nous avons donc calculé les résultats au test de vocabulaire pour chaque participant, par niveau de fréquence. Ces résultats, présentés dans le tableau 4.1, permettent l'atteinte de notre premier objectif de recherche, soit celui de déterminer

quelle proportion des 3000 premiers mots les plus fréquents est connue par les étudiants adultes qui terminent le tronc commun de la formation en francisation. Nous remarquons que la moyenne la plus élevée est au niveau du premier mille. Elle diminue de 5% au niveau du deuxième mille et de 9% au troisième mille.

Tableau 4-1 Résultats au test de vocabulaire

Niveau	Moyenne (%)	Écart-type	Minimum (%)	Maximum (%)
1k	80,88	13,01	55	95
2k	75,88	11,14	55	90
3k	66,03	16,66	45	87,5

Par la suite, nous avons calculé la taille du vocabulaire en nombre de mots pour chacun des étudiants. Pour faire le calcul, nous avons transformé le pourcentage obtenu pour chacun des niveaux de 1000 (1k, 2k et 3k) en nombre de mots. Par exemple, un résultat de 90% au niveau 1k correspondrait à 900 mots ($90/100 = 900/1000$). Puis, nous avons additionné le nombre de mots connus pour les trois niveaux pour en arriver à un total correspondant à la taille du vocabulaire. Les statistiques de la taille du vocabulaire pour le groupe sont présentées dans le tableau 4.2.

Tableau 4-2 Taille du vocabulaire (en nombre de mots)

Moyenne	Écart-type	Minimum	Maximum
2228	370,16	1650	2675

4.3 Taille du vocabulaire et habiletés réceptives

Comme nous l'avons expliqué dans la méthodologie, c'est l'enseignante des étudiants qui a procédé à la correction des examens ministériels de compréhension orale et de compréhension écrite. Nous pouvons observer les résultats de ces examens dans le tableau 4.3. Nous remarquons que les moyennes sont faibles, particulièrement celle de l'examen de compréhension orale, qui se situe en dessous de la note de passage, c'est-à-dire 60%. Les raisons de ces faibles résultats seront analysées plus en profondeur dans notre discussion, au chapitre suivant.

Tableau 4-3 Résultats aux examens de compréhension orale et de compréhension écrite

Examen	Moyenne (%)	Écart-type	Minimum (%)	Maximum (%)
Compréhension orale	52,65	17,06	25	80
Compréhension écrite	60,29	20,88	25	90

Ensuite, nous avons fait les analyses nécessaires pour atteindre notre deuxième objectif de recherche, c'est-à-dire de déterminer le rôle que joue la connaissance ou le manque de connaissance de vocabulaire des étudiants dans leur compréhension orale et leur compréhension écrite. Pour atteindre cet objectif, nous avons d'abord fait, dans un premier temps, une corrélation de Pearson entre la taille du vocabulaire et le résultat en compréhension orale. Le résultat est qu'il existe une forte corrélation entre ces deux éléments car $r = 0,79$ et que cette corrélation est significative ($p < 0,001$). Dans un deuxième temps, nous avons fait une corrélation de Pearson, mais cette fois entre la taille du vocabulaire et le résultat en compréhension écrite. Cette analyse

démontre qu'il y a une corrélation entre les deux éléments car $r = 0,59$ et que cette corrélation est significative ($p < 0,05$).

En complément aux deux corrélations précédentes, nous avons fait deux régressions multiples. Notre première régression multiple a pour but de mesurer à quel point les résultats du test de vocabulaire aux niveaux 1k, 2k, 3k sont responsables de la variance de la variable dépendante, soit le résultat à l'examen de compréhension orale. Nous présentons les résultats de cette régression dans le tableau 4.4. Nous observons que $r^2 = 0,64$, ce qui est statistiquement significatif ($p < 0,05$). Cela veut dire que 64% de la variance du résultat en compréhension orale est due à la connaissance du vocabulaire, telle que représentée par le résultat global au test de vocabulaire. En regardant de plus près les coefficients pour chacun des niveaux de 1000, nous remarquons que les mots du niveau 1k semblent avoir un poids plus important, avec un coefficient de 0,73. Par contre, ce résultat, ainsi que les résultats pour les niveaux 2k et 3k, ne sont pas significatifs. Nous soupçonnons un manque de puissance lié à la petite taille de notre échantillon

Tableau 4-4 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k, 2k et 3k) et le résultat à l'examen de compréhension orale

<i>r</i>	<i>r</i> ²	valeur <i>p</i>
0,80	0,64	0,003
Niveaux	Coefficients	valeur <i>p</i>
1k	0,73	0,109
2k	-0,08	0,870
3k	0,37	0,147

Étant donné que notre régression multiple ne donnait pas de résultats significatifs pour chacun des niveaux, nous avons opté pour la méthode d'élimination descendante. Avec cette méthode, le logiciel SPSS retire la variable ayant la plus faible contribution au modèle si la variation du r^2 n'est pas significative en l'éliminant. Il répète la procédure jusqu'à ce que toutes les variables conservées contribuent significativement à l'amélioration du r^2 . Dans notre régression, le logiciel a d'abord retiré la variable 2k, considérant que cette dernière est celle qui contribue le moins au modèle. Nous pouvons voir les résultats de cette autre régression dans le tableau 4.5. Comme la valeur p pour la variable 3k n'est toujours pas significative dans ce modèle, le logiciel l'a retirée. Une nouvelle régression contenant seulement la variable indépendante 1k a été créée et est présentée dans le tableau 4.6. Nous observons que $r^2 = 0,576$ et que ce résultat est significatif puisque $p < 0,001$. Cela signifie que 58% de la variance du résultat en compréhension orale est due à la connaissance des mots du premier niveau de 1000 dans le test de vocabulaire. En d'autres mots, les apprenants doivent avoir une forte connaissance des mots très fréquents pour assurer leur compréhension.

Tableau 4-5 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k et 3k) et le résultat à l'examen de compréhension orale

r	r^2	valeur p
0,80	0,64	0,001
Niveaux	Coefficients	valeur p
1k	0,68	0,033
3k	0,36	0,128

Tableau 4-6 Régression multiple pour le résultat au test de vocabulaire (niveau 1k) et le résultat à l'examen de compréhension orale

<i>r</i>	r^2	valeur <i>p</i>
0,76	0,58	0,000
Niveau	Coefficient	valeur <i>p</i>
1k	1	0,0001

Après avoir obtenu les résultats pour la compréhension orale, nous avons fait une deuxième régression multiple pour la compréhension écrite. Le but visé était de mesurer à quel point les variables indépendantes des résultats du test de vocabulaire aux niveaux 1k, 2k, 3k étaient responsables de la variance de la variable dépendante, c'est-à-dire le résultat à l'examen de compréhension écrite. Les résultats de cette régression sont présentés dans le tableau 4.7. Nous obtenons ainsi un $r^2 = 0,38$, mais ce résultat n'est pas significatif car $p > 0,05$. Cependant, nous ne sommes pas loin du seuil car $p = 0,097$. Encore une fois, il est possible que la petite taille de l'échantillon ait entraîné un manque de puissance.

Tableau 4-7 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k, 2k et 3k) et le résultat à l'examen de compréhension écrite

<i>r</i>	r^2	valeur <i>p</i>
0,61	0,38	0,097
Niveaux	Coefficients	valeur <i>p</i>
1k	0,84	0,243
2k	-0,06	0,942
3k	0,19	0,629

Nous avons donc utilisé le même processus que dans la régression précédente, pour la compréhension orale. Le logiciel SPSS a procédé à une élimination descendante des variables, dont nous pouvons observer la première étape dans le tableau 4.8. En retirant la variable 2k, la régression complète est significative étant donné que $p < 0,05$. Cependant, les résultats pour chacune des variables ne sont pas significatifs. Le logiciel a donc procédé à la deuxième étape, soit l'élimination de la variable 3k. Dans ce nouveau modèle, présenté dans le tableau 4.9, $r^2 = 0,36$ est significatif car $p < 0,05$. En d'autres mots, la connaissance des mots du premier niveau de 1000 est responsable de 36% de la variance dans le résultat à l'examen de compréhension orale. Nous constatons que, en compréhension écrite tout comme en compréhension orale, le niveau 1k est le plus déterminant dans la variance.

Tableau 4-8 Régression multiple pour les résultats au test de vocabulaire (niveaux 1k et 3k) et le résultat à l'examen de compréhension écrite

<i>r</i>	<i>r</i> ²	valeur <i>p</i>
0,61	0,38	0,037
Niveaux	Coefficients	valeur <i>p</i>
1k	0,80	0,106
3k	0,19	0,616

Tableau 4-9 Régression multiple pour le résultat au test de vocabulaire (niveau 1k) et le résultat à l'examen de compréhension écrite

<i>r</i>	<i>r</i> ²	valeur <i>p</i>
0,60	0,36	0,010
Niveau	Coefficient	valeur <i>p</i>
1k	0,97	0,010

4.4 Questionnaire sociodémographique

Afin d'atteindre notre troisième objectif de recherche, nous avons analysé les données recueillies grâce au questionnaire sociodémographique complété par les participants. Notre objectif était de déterminer si les facteurs sociodémographiques avaient une influence sur la taille de vocabulaire et les résultats aux examens mesurant les habiletés réceptives.

Tout d'abord, nous devons mentionner que deux variables ont été éliminées de notre analyse : le sexe et l'âge. La variable du sexe ne peut être considérée, étant donné que notre échantillon compte 14 femmes et seulement 3 hommes. Le faible nombre d'individus de sexe masculin ne permet pas de faire des comparaisons valables entre les deux sexes. De la même manière, la variable de l'âge ne peut faire l'objet d'analyse. La classe qui constitue notre échantillon est assez homogène du point de vue de l'âge. La majorité du groupe, soit 14 individus sur 17, ont entre 30 et 40 ans. Il n'y a pas d'écart assez marqué pour effectuer des comparaisons entre les groupes d'âge.

Pour analyser les effets de la langue maternelle, nous avons créé deux groupes distincts dans notre échantillon. Le premier groupe se compose des étudiants ayant une langue latine comme langue maternelle ($n=10$). Le deuxième groupe est constitué de ceux ayant une langue non latine comme langue maternelle ($n=7$). Nous avons ensuite vérifié si la langue maternelle avait un impact sur la taille de vocabulaire, le résultat en compréhension orale et le résultat en compréhension écrite. Comme la langue maternelle est une variable non continue et que nous désirions la mettre en relation avec des variables continues, il était préférable de procéder à des tests T sur l'égalité des moyennes. Nous avons donc effectué trois tests T, dont les résultats sont présentés dans le tableau 4.10. Comme nous pouvons le constater, aucun des tests T n'atteint le seuil pour être statistiquement significatif. Cependant, la valeur p est très près du seuil de 0,05 pour la taille de vocabulaire et la compréhension écrite. Il se peut que le petit nombre de participants ait entraîné un manque de puissance et qu'en augmentant la taille de l'échantillon, nous atteindrions un résultat significatif. Il est donc possible que, comme les moyennes nous l'indiquent, les apprenants de langue maternelle latine connaissent un plus grand nombre de mots et obtiennent un meilleur résultat à l'examen de compréhension écrite.

Tableau 4-10 Tests T sur la langue maternelle

	Langue maternelle	N	Moyenne	Écart-type	t	Valeur p
Taille de vocabulaire	latine	10	2370 mots	316,62	2,08	0,055
	non latine	7	2025 mots	365,15		
Compréhension orale	latine	10	58,00%	12,74	1,62	0,128
	non latine	7	45,00%	20,41		
Compréhension écrite	latine	10	68,00%	20,71	2,05	0,058
	non latine	7	45,71%	23,88		

Notre exploration se poursuit avec la variable de la scolarité. Comme pour la variable précédente, nous avons créé deux groupes au sein de l'échantillon : ceux qui ont terminé leurs études universitaires ($n=10$) et les autres ($n=7$). Cette division nous a permis de vérifier si le fait d'avoir complété des études universitaires avait un effet sur la taille du vocabulaire, le résultat en compréhension orale et le résultat en compréhension écrite. Encore une fois, puisque nous mettions en relation une variable non continue et des variables continues, nous avons encore opté pour trois tests T. Le tableau 4.11 nous permet d'observer que les deux premiers tests ne sont pas significatifs et ne sont pas près du seuil. Dans le premier test, nous pouvons voir que les étudiants détenant un diplôme universitaire connaissent en moyenne moins de mots (2046 mots) que ceux qui n'ont pas de diplôme (2355 mots). Le deuxième test suggère qu'en compréhension orale, les étudiants qui n'ont pas complété leurs études universitaires ont en moyenne un résultat supérieur (57,5%) à ceux qui les ont complétées (45,71%). Il est possible que, puisque les études universitaires n'ont pas été faites en français, le fait de détenir un diplôme n'influence pas la taille du vocabulaire et les habiletés en compréhension orale. Nous ne pouvons tirer de conclusions étant donné que le test n'est pas significatif, mais il serait intéressant de creuser la question. Cependant, pour la compréhension écrite, le test détecte une différence significative entre les moyennes des deux groupes. Les individus ayant complété des études universitaires obtiennent en moyenne des résultats plus élevés dans l'examen de compréhension écrite que les autres. La différence entre les deux groupes est non seulement significative ($p < 0,05$), mais également très importante : la moyenne est de 71,5% dans le premier groupe, alors qu'elle est de 40,71% dans le deuxième.

Tableau 4-11 Tests T sur la scolarité

	Études universitaires	N	Moyenne	Écart-type	t	Valeur <i>p</i>
Taille de vocabulaire	Terminées	7	2046 mots	377 mots	1,81	0,091
	non terminées	10	2355 mots	325 mots		
Compréhension orale	Terminées	7	45,71%	13,97%	1,45	0,168
	non terminées	10	57,50%	17,99%		
Compréhension écrite	Terminées	7	40,71%	19,88%	3,29	0,005
	non terminées	10	71,50%	18,42%		

Les dernières variables étudiées dans cette section sont celles liées au contact avec la langue française. Nous nous rappelons que nous avons demandé aux participants d'évaluer leur contact avec le français à l'extérieur de la classe. Ils devaient déterminer à quelle fréquence ils parlaient, écoutaient, lisaient et écrivaient en français. Pour chacune des quatre habiletés, ils devaient choisir entre très souvent, souvent, parfois ou rarement/jamais. Nous avons converti les réponses des étudiants en échelle numérique dans laquelle 4 correspond à très souvent et 1 correspond à rarement/jamais. Avec ces informations, nous souhaitons observer si le fait d'avoir plus ou moins de contact avec le français avait un impact sur la taille de vocabulaire, le résultat en compréhension orale et le résultat en compréhension écrite. Pour cela, nous avons effectué des corrélations de Pearson dont les résultats sont présentés dans le tableau 4.12. En observant le tableau, nous constatons que les coefficients de corrélations sont très bas. De plus, aucune corrélation n'est significative, ou même près du seuil de significativité. Le fait de parler, écouter, lire ou écrire en français ne semble pas avoir un effet mesurable sur les résultats aux examens passés par les étudiants.

Tableau 4-12 Corrélations entre le contact avec la langue française et la taille du vocabulaire, la compréhension orale et la compréhension écrite

	Taille de voc	Compréhension orale	Compréhension écrite
Parler			
Corrélation	-0,04	-0,09	-0,29
Valeur <i>p</i>	0,877	0,735	0,256
Écouter			
Corrélation	0,13	0,00	0,08
Valeur <i>p</i>	0,628	1	0,776
Lire			
Corrélation	0,02	0,01	-0,15
Valeur <i>p</i>	0,936	0,964	0,580
Écrire			
Corrélation	0,26	0,19	0,03
Valeur <i>p</i>	0,305	0,466	0,922

4.5 Conclusion

Dans ce chapitre, nous avons analysé les données recueillies et nous avons présenté les résultats permettant d'atteindre nos trois objectifs de recherche. Ceci fait, nous pourrions donc répondre deux questions de recherche et discuter de nos résultats dans le chapitre suivant.

CHAPITRE V

DISCUSSION

5.1 Introduction

Dans cette partie de la recherche, nous allons explorer la signification des découvertes faites au chapitre précédent. Le chapitre se divise en quatre sections. Dans la première section, nous répondrons à nos deux questions de recherche et nous discuterons de nos résultats concernant notre troisième objectif de recherche. La deuxième section exposera les limites de notre recherche. Dans la troisième section, nous présenterons les implications pédagogiques. Finalement, la dernière section donnera des perspectives de recherches futures.

5.2 Questions et objectifs de recherche

Comme nous l'avons exposé au premier chapitre, notre recherche est orientée en fonction de deux questions et trois objectifs. Nos deux questions et nos deux premiers objectifs concernent la taille du vocabulaire et les habiletés réceptives. Notre troisième objectif, quant à lui, concerne le rôle des facteurs sociodémographiques.

5.2.1 Taille du vocabulaire

Nous vous rappelons que notre première question de recherche était la suivante : quelle est la taille du vocabulaire réceptif des étudiants adultes de francisation de niveau 6? Nous n'avions pas d'hypothèse précise quant au nombre de mots maîtrisés par ces étudiants, étant donné qu'aucune étude antérieure ne s'était penchée sur une population comparable à la nôtre. En revanche, plusieurs éléments nous laissaient croire que le vocabulaire des étudiants présenterait des lacunes : l'absence d'enseignement systématique du lexique, le programme organisé par thèmes et les changements fréquents de professeur.

Les résultats au test de vocabulaire nous démontrent que la taille du vocabulaire moyenne des étudiants est de 2228 mots. En analysant les résultats par niveau de fréquence, nous remarquons que la proportion de mots connus est plus élevée au premier niveau de 1000, pour diminuer ensuite aux niveaux 2000 et 3000. Cela confirme l'hypothèse selon laquelle les mots les plus fréquents ont tendance à être appris en premier. Il est donc normal que les étudiants connaissent davantage de mots provenant des mille premiers mots les plus fréquents. Ils maîtrisent moins les mots de la fréquence 2000 et encore moins ceux de la fréquence 3000, ce qui correspond au profil normal d'apprentissage du vocabulaire.

Nous allons maintenant regarder de plus près la proportion de mots connus par les étudiants pour chacun des niveaux de 1000. Au premier niveau de 1000, la moyenne est d'environ 81%, avec un écart-type de 13%. Cela signifie que plusieurs étudiants ont un résultat inférieur à 68%, c'est-à-dire qu'ils connaissent moins de 680 mots au niveau 1k. Nous pouvons observer que la limite inférieure est de 55%. De nombreux étudiants n'ont donc pas une maîtrise suffisante des premiers 1000 mots fréquents. Il faut rappeler que les mots de ce niveau représentent 80% d'un texte écrit courant. De la même manière, au deuxième niveau de 1000, la moyenne est de 76%

et l'écart-type est de 11%. Par conséquent, beaucoup d'apprenants possèdent un vocabulaire de 650 mots ou moins au niveau 2k. En français, les 2000 premiers mots, permettent de couvrir une étendue de 90% d'un texte écrit (Cobb et Horst, 2004). Donc, si un étudiant maîtrisait seulement 65% des mots aux niveaux 1k et 2k, il connaîtrait seulement en moyenne ($90\% \times 65\% =$) 59% des mots d'un texte courant, ou, autrement dit, il n'en connaîtrait pas 41%. Quatre mots inconnus sur dix sont ainsi inconnus, en incluant les mots de fonction (déterminants, pronoms, prépositions, conjonctions) qui n'ont pas été inclus dans le test de vocabulaire, mais qui ne seront pas d'un grand secours si les mots entre lesquels ils établissent les rapports sont inconnus. Un apprenant qui se trouve dans cette situation peut, dans le meilleur des cas, avoir une idée très générale du sujet, mais sa compréhension demeure très incomplète. Comme nous l'avons vu plus tôt, les études démontrent que, pour avoir une compréhension adéquate d'un texte, les lecteurs doivent connaître de 95% à 98% des mots (Laufer, 1995; Hu et Nation, 2000). Les résultats au test de vocabulaire, tel qu'expliqué précédemment, démontrent que les étudiants de niveau 6 n'atteignent pas une couverture si élevée. La taille de leur vocabulaire est donc insuffisante pour comprendre un texte authentique en français. .

Il y a lieu de se demander pour quelles raisons certains apprenants n'ont pas atteint une maîtrise suffisante des mots fréquents. Trois explications nous paraissent plausibles. Premièrement, suivant la théorie de O'Dell citée dans notre cadre théorique, il se peut que l'orientation thématique du programme fasse que certains mots fréquents soient délaissés pour d'autres, moins fréquents mais liés aux thèmes. Deuxièmement, le problème pourrait être lié à un manque de recyclage de vocabulaire. Nous avons également expliqué ce concept de Kachroo présenté par O'Dell (1997) au deuxième chapitre. Selon ce concept, les mots devraient être répétés en moyenne sept fois pour être appris. Dans le contexte de la francisation, il est fort possible que les mots ne soient pas revus suffisamment souvent pour être appris.

Effectivement, les thèmes vus à l'intérieur d'un niveau sont très différents et les thèmes varient d'un niveau à l'autre. De plus, les étudiants changent d'enseignant tous les deux mois. Il y a donc un risque que les mêmes mots ne soient pas revus fréquemment et que les étudiants ne les mémorisent jamais. Troisièmement, il faut considérer que le caractère très intensif des cours peut défavoriser les étudiants. Le fait d'étudier le français de quatre à six heures par jour est très exigeant au plan cognitif. Les adultes suivant ce type formation n'ont peut-être pas le temps d'assimiler toutes les connaissances en un si court laps de temps.

5.2.2 Taille du vocabulaire et habiletés réceptives

Avant de poursuivre la discussion avec notre seconde question de recherche, il est important d'aborder la question des résultats aux examens ministériels. En effet, notre analyse au chapitre précédent a révélé une moyenne sous la note de passage à l'examen de compréhension orale (52,65%) et une moyenne atteignant à peine la note de passage à l'examen de compréhension écrite (60,29%). Comment expliquer ces faibles résultats? Une première hypothèse est que la maîtrise insuffisante du vocabulaire des étudiants, telle que démontrée dans la section précédente, affecte leurs résultats. Un entretien avec l'enseignante nous a permis d'avancer une seconde hypothèse. Le groupe de participants est, aux dires de l'enseignante, un groupe moyen, mais non faible. Cependant, les étudiants ont tendance à sous performer lors des examens ministériels, en particulier celui de compréhension orale. Puisque le programme encore utilisé date de 1994, les examens qui y sont associés sont également anciens. L'examen oral, originalement sur cassette audio, a été transféré sur disque compact plutôt que d'être réenregistré. La qualité sonore laisse donc à désirer et certains passages sont difficiles à comprendre. De surcroît, la forme de l'examen, c'est-à-dire les questions à choix multiples, entraîne une certaine confusion chez plusieurs étudiants. Certaines questions ressemblent à des pièges : les choix de réponses se ressemblent beaucoup et cela sème le doute dans l'esprit des étudiants.

Par conséquent, un étudiant pourrait avoir bien compris le dialogue, mais ne pas réussir à donner la bonne réponse. En ce qui a trait à l'examen écrit, l'enseignante note moins de difficultés. Étant donné que l'évaluation porte sur des documents écrits, les difficultés techniques n'affectent pas l'examen. En revanche, on y retrouve aussi des questions à choix multiples qui peuvent être confondants. Il est donc possible qu'un étudiant comprenne le texte, mais ne donne pas la réponse attendue dans l'examen. Ces explications nous amènent à croire que les étudiants de l'échantillon ont obtenu un résultat inférieur à leur capacité réelle. Ce fait pourrait constituer une limite à notre recherche. Nous en mentionnerons d'autres dans la section suivante de ce chapitre.

Entre-temps, nous répondrons à notre deuxième question de recherche : quelle est la corrélation entre la taille du vocabulaire réceptif des étudiants et leurs résultats aux examens ministériels en compréhension écrite et en compréhension orale? Notre hypothèse, basée sur la littérature, était qu'il existe une corrélation forte entre la taille du vocabulaire et la compréhension écrite. Toujours selon la littérature, il y avait également une corrélation, mais cette fois modérée, entre la taille du vocabulaire et la compréhension orale.

Notre analyse, effectuée au chapitre précédent, nous démontre qu'il existe bel et bien une corrélation significative entre la taille du vocabulaire des étudiants et leurs résultats à l'examen de compréhension écrite. Cependant, cette corrélation est modérée, et non forte, avec $r = 0,59$. En considérant que $r^2 = 0,35$, nous pouvons affirmer que 35% de la variance en compréhension écrite est liée à la taille du vocabulaire. Il existe également une corrélation significative entre la taille du vocabulaire et la compréhension orale. Dans ce cas, la corrélation est forte et non modérée, avec $r = 0,79$. Cela correspond à un r^2 de 0,62, ce qui signifie que la taille du vocabulaire est responsable de 62% de la variance en compréhension orale.

Il y a lieu de se demander pourquoi notre étude arrive à des résultats différents de ceux de la littérature. Une étude comme celle de Staehr (2008) rapporte que le vocabulaire est davantage corrélé à la lecture qu'à l'écoute. Nous croyons que la structure du programme du MELS est en partie responsable de cette différence. Effectivement, parler et écouter sont les premières habiletés à développer en début de formation. On intègre la lecture et l'écriture au fur et à mesure que les étudiants progressent d'un niveau à l'autre. Globalement, ce sont tout de même les habiletés orales qui priment. La lecture n'est pas nécessairement une activité pratiquée régulièrement en classe. Il se peut donc que les étudiants finissants en francisation aient développé davantage d'habiletés orales qu'écrites. Pourtant, nous nous rappelons que le questionnaire sociodémographique avait indiqué qu'une grande partie de l'échantillon lisait très souvent ou souvent en français à l'extérieur de la classe. Nous ne savons pas quels types de lecture les étudiants font et quel est leur degré de compréhension. Il va sans dire que le type de textes que l'on retrouve dans un examen de compréhension écrite peut différer grandement de la lecture libre des étudiants. De plus, si le vocabulaire de base des apprenants est déficient, ceux-ci ne seront pas en mesure d'utiliser le contexte pour apprendre de nouveaux mots et ne pourront pas bénéficier de la lecture comme moyen d'accroître leur vocabulaire.

Un élément très intéressant révélé par l'analyse des résultats est la grande importance de la connaissance des 1000 premiers mots les plus fréquents. Il est normal que ces premiers 1000 mots jouent un rôle majeur dans le test de compréhension orale. En effet, le discours oral comprend généralement une moins grande diversité de vocabulaire et les mots les plus fréquents y sont encore plus souvent utilisés qu'à l'écrit. Malgré le caractère primordial de la connaissance des mots du premier niveau de 1000, nous avons vu plus tôt qu'un grand nombre d'apprenants n'en possèdent pas une connaissance suffisante. L'enseignement

thématique, l'absence d'objectif mesurable en terme de vocabulaire et le changement fréquent de professeur empêchent beaucoup d'étudiants d'atteindre la connaissance lexicale dont ils ont besoin.

Cependant, il est plus étonnant de constater que les 1000 mots les plus fréquents ont aussi une grande importance dans la compréhension écrite. De manière générale, le discours écrit est plus diversifié et comprend des mots provenant de différents niveaux de fréquence. Ce n'est peut-être pas le cas des textes présentés aux étudiants pour l'examen de compréhension orale. Nous aurions souhaité analyser le niveau de fréquence des mots contenus dans les textes en questions mais nous n'avons pas été autorisée à accéder à l'examen. En revanche, nous pouvons faire une hypothèse au sujet des textes de l'examen de compréhension écrite. Puisqu'il s'agit de textes spécialement rédigés pour l'occasion et non de textes authentiques, il est fort possible qu'ils aient été simplifiés. La distribution normale ayant été modifiée, les mots du premier niveau de 1000 pourraient s'y retrouver en plus grande proportion.

5.2.3 Rôle des facteurs sociodémographiques

Au chapitre précédent, nous avons exploré les données obtenues au moyen du questionnaire sociodémographique : le sexe, l'âge, la langue maternelle, la scolarité et le contact avec la langue française. Nous souhaitons ainsi atteindre notre troisième objectif de recherche, soit de déterminer l'effet de ces facteurs sur la taille du vocabulaire, la compréhension orale et la compréhension écrite.

Il ne nous a pas été possible de faire des comparaisons entre les sexes, la majorité de l'échantillon étant de sexe féminin. Nous ne sommes pas en mesure de vérifier si les femmes obtiennent de meilleurs résultats que les hommes dans le test de vocabulaire ou les examens ministériels. Cependant, nous sommes d'avis que le fait

de compter seulement trois hommes dans une classe de dix-sept étudiants parle de lui-même. Cela démontre que les femmes sont plus nombreuses à étudier le français à un niveau avancé. Ce fait pourrait être interprété de différentes manières. Il se pourrait que les femmes présentent un plus grand intérêt pour l'étude de la langue. Il est aussi possible qu'elles réussissent mieux que les hommes en francisation, et que cette meilleure réussite leur permettant de cheminer à travers les niveaux sans décrocher. Quelles qu'en soient les raisons, la présence dominante des femmes dans cette classe de francisation suit la tendance observée dans les collèges et les universités.

Nous n'avons pas pu analyser l'influence de l'âge sur la taille de vocabulaire ou la réussite aux examens. L'absence d'écart marqué entre les âges des participants nous empêche de faire des comparaisons entre les étudiants plus jeunes et les plus âgés. Nous ne pourrions donc pas vérifier l'intuition selon laquelle les individus plus jeunes ont un vocabulaire plus vaste et obtiennent de meilleurs résultats que les autres. Nous croyons qu'il s'agit tout de même d'une question intéressante qui mériterait d'être explorée.

La langue maternelle des étudiants nous a donné l'occasion de faire des observations intéressantes. Le test T ne nous a pas permis de conclure que la langue maternelle avait un effet significatif sur la taille de vocabulaire, le résultat en compréhension orale et le résultat en compréhension écrite. En revanche, la valeur p est très près d'être significative pour la taille du vocabulaire et la compréhension écrite. Nous nous permettons donc de faire certaines hypothèses. Premièrement, le fait d'avoir une langue latine comme langue maternelle pourrait avoir un effet sur le nombre de mots français connus par l'apprenant. Cela nous semble logique, étant donné que les langues latines ont un grand nombre de mots en commun et qu'il est facile de reconnaître ces mots, même sans les avoir rencontrés auparavant. Deuxièmement, les participants ayant une langue latine comme langue maternelle

pourraient avoir plus de succès en compréhension écrite. Cela apparaît également logique, étant donné les observations faites précédemment. Un locuteur de langue latine reconnaîtra spontanément un grand nombre de mots rencontrés dans un document écrit en français. Par le fait même, il accédera plus aisément au sens du texte. En revanche, il est intéressant de se questionner sur le résultat loin d'être significatif en ce qui concerne la compréhension orale. Il se pourrait que le fait de parler une langue latine ne facilite pas autant l'écoute que la lecture. Même en rencontrant des mots connus dans le discours oral, il est possible que le locuteur de langue latine ne les reconnaisse pas en raison de la prononciation et de la vitesse d'élocution.

Nous avons également fait des observations en lien avec le niveau de scolarité des participants. Le test T nous a révélé que le fait d'avoir complété des études universitaires n'avait pas d'effet significatif sur la taille du vocabulaire et sur le résultat en compréhension orale. En revanche, l'effet est significatif pour le résultat en compréhension écrite. Les apprenants qui détiennent un diplôme universitaire réussissent mieux en lecture. Cette supériorité est sans doute liée aux connaissances générales acquises et aux stratégies de lecture développées par ces personnes lors de leurs études.

En dernier lieu, nous n'avons trouvé aucune corrélation entre le contact avec la langue française hors de l'école et la taille de vocabulaire, la compréhension orale ou la compréhension écrite. Le fait qu'un étudiant parle, écoute, lise ou écrive en français à l'extérieur de la salle de classe ne semble avoir aucune influence sur sa performance aux différents examens. Comme nous l'avons mentionné dans le chapitre portant sur la méthodologie, pour bénéficier de l'apprentissage indirect, les apprenants doivent posséder une base de vocabulaire suffisante. Il apparaît que les

étudiants de niveau 6 ne possèdent pas cette base et se trouvent dans l'impossibilité d'utiliser le contexte pour enrichir leur lexique.

5.3 Limites

Comme la plupart des autres recherches, la nôtre comprend certaines limites. La première limite, qui est sans contredit la plus importante, est la très petite taille de notre échantillon. Ce dernier est constitué de seulement 17 participants, soit l'équivalent d'une seule classe. Ce petit nombre nous empêche de généraliser les résultats à l'ensemble des étudiants de niveau 6. Nous justifions cet échantillon restreint par le fait que, depuis le début de ce projet, il y a eu de grands bouleversements dans le domaine de la francisation des adultes. Les coupures gouvernementales ont engendré la fermeture de plusieurs classes et de nombreuses pertes d'emploi. Une réorientation de carrière et la perte de contacts dans le milieu ont rendu le recrutement des participants beaucoup plus ardu. Ceci dit, un échantillon plus grand nous aurait probablement permis d'observer les effets du sexe, de l'âge et du contact avec la langue française.

La deuxième limite de la recherche concerne le test de vocabulaire et son lien avec la compréhension orale. Le test de vocabulaire que nous effectuons mesure la reconnaissance de la forme écrite des mots, alors que la compréhension orale implique la reconnaissance orale des mots. Il est donc possible qu'un étudiant connaisse certains mots à l'oral, mais ne soit pas en mesure de les reconnaître à l'écrit. À l'inverse, une personne peut identifier la forme écrite d'un mot sans pouvoir identifier sa forme orale. La fiabilité de la corrélation entre le test de vocabulaire et la compréhension pourrait être plus élevée si nous faisons un test de vocabulaire supplémentaire, soit un test de reconnaissance orale, tel l'Aural Lex de Milton. Certes, nous avons rapporté une étude de Milton (2006) qui démontre une forte corrélation

entre le X-Lex et l'Aural Lex. Il aurait tout de même pu être intéressant de l'observer nous-même auprès de notre population cible. Cependant, cela aurait demandé une somme de temps supplémentaire afin de créer une version française du test et, pour la passation du test, nous aurions eu besoin d'un accès à un laboratoire informatique.

La troisième et dernière limite identifiée est également liée au test de vocabulaire. Celui-ci couvre seulement les 3000 mots les plus fréquents en français. Il est possible que certains étudiants connaissent des mots qui se situent à d'autres niveaux de fréquence. Normalement, la proportion de mots connus décroît à chaque niveau de 1000. Cependant, certains apprenants pourraient présenter des profils atypiques et connaître des mots aux niveaux 4000 ou 5000, même sans maîtriser complètement les niveaux 1000, 2000 et 3000.

5.4 Implications pédagogiques

En dépit de ses limites, nous croyons que cette recherche peut mener à des implications pédagogiques concrètes. Rappelons-nous que les étudiants de niveau 6 arrivent à la fin du tronc commun : certains étudient ensuite au niveau écrit, certains entreprennent des études collégiales ou universitaires en français, alors que d'autres se dirigent vers le marché du travail. Peu importe l'orientation choisie, ces adultes doivent être bien outillés au niveau langagier, particulièrement au plan du vocabulaire. Notre étude a confirmé le rôle significatif joué par le vocabulaire fréquent, car il existe une corrélation entre la taille du vocabulaire et la compréhension de textes oraux et écrits. Nous avons constaté que la maîtrise du vocabulaire fréquent est nécessaire, mais qu'elle n'est pas suffisante chez les étudiants de niveau 6 ayant participé à cette étude.

Même si un nouveau programme-cadre en francisation des adultes fera sous peu son entrée dans les salles de classe, nous ne croyons pas que l'apprentissage du vocabulaire fréquent s'en trouvera améliorée. Le vocabulaire mis au programme demeure le vocabulaire lié aux thèmes de la vie quotidienne ce qui, comme nous l'avons vu, peut entraîner des lacunes au niveau du vocabulaire fréquent. Il nous semble donc essentiel que le MELS ajoute à son programme des objectifs précis et mesurables en matière de lexique. Il doit également développer des moyens de s'assurer que les étudiants intègrent progressivement le vocabulaire fréquent d'un niveau à l'autre.

Il va sans dire qu'il est primordial de sensibiliser les enseignants de francisation à l'importance du vocabulaire. Le Ministère devrait organiser des conférences et des formations sur ce sujet dans les centres de francisation. Il faudrait également fournir aux professeurs des outils permettant de connaître le vocabulaire fréquent, afin de le présenter aux étudiants. Par exemple, il existe un dictionnaire de fréquence en français qui est tout à fait à jour et fiable. *A Frequency dictionary of French : Core Vocabulary for Learners*, par Lonsdale et Le Bras (2009), est basé sur un corpus de 25 millions de mots et présente les 5000 premiers mots les plus fréquents en français. Le dictionnaire classe ces 5000 mots par fréquence, selon l'ordre alphabétique et par champ sémantique. Un tel ouvrage constitue une référence précieuse pour les enseignants. Ceux-ci pourront ainsi intégrer davantage de vocabulaire fréquent à leurs leçons et s'assurer qu'il soit suffisamment répété pour être appris. Milton (2009) suggère aux enseignants de présenter huit à douze nouveaux mots par leçon.

On devrait également initier les enseignants à l'utilisation de *Lex Tutor*, un site Internet consacré au vocabulaire développé par Tom Cobb. Ce site offre de nombreux outils gratuits et intéressants, tant pour les professeurs que pour les apprenants de langue seconde. D'abord, on peut y trouver un test de vocabulaire des 5000 premiers

mots les plus fréquents en français. Les professeurs pourraient utiliser ce test pour connaître le niveau de leurs étudiants et éventuellement pallier leurs lacunes. Un autre outil pour les enseignants est le *VocabProfile*, lequel permet d'analyser facilement le niveau de fréquence des mots contenus dans les textes présentés aux étudiants. Au *VocabProfile* s'ajoute le *VP negative*, qui offre aux professeurs l'opportunité de voir quels mots fréquents ne sont pas présentés dans un manuel scolaire ou un roman ainsi que le nombre de fois que chacune des familles de mots est recyclée. Le *Vocabprofile* français a été récemment mis à jour et utilise dorénavant les nouvelles listes de fréquence de Lonsdale et Le Bras. En plus des outils dédiés aux enseignants, *Lex Tutor* offre également des activités en ligne pour que les étudiants pratiquent le vocabulaire fréquent.

Jusqu'à maintenant, nos suggestions ont été principalement adressées au MELS et aux enseignants. Ils sont bien entendu des acteurs majeurs dans le monde éducatif. En revanche, il ne faudrait pas oublier celui qui se trouve au cœur du processus d'apprentissage : l'apprenant lui-même. Nous croyons que les étudiants en francisation doivent être informés de l'importance de construire une solide base en vocabulaire. Si les apprenants savent que la connaissance du lexique fréquent leur permettra d'améliorer leur compréhension, ils seront certes plus motivés et réceptifs. Il est important qu'ils fournissent des efforts conscients et consacrent du temps pour acquérir le lexique nécessaire à une plus grande autonomie. Une fois la base acquise, les apprenants seront en mesure de profiter davantage du riche environnement linguistique à leur disposition. Puisque le Québec est un environnement majoritairement francophone, les occasions d'être en contact avec la langue française ne manquent pas : lecture, films, radio, chansons... Les adultes en francisation doivent saisir cette chance pour progresser plus rapidement. À ce sujet, Milton (2009) souligne : « A combination of good classroom practice and well-directed effort

outside class can begin to explain how learners acquire very large vocabularies of thousand of words, so they can achieve functional fluency. » (p. 250)

Bref, notre étude pourrait guider les différents acteurs de la francisation aux adultes, tant le MELS que le personnel enseignant et les étudiants eux-mêmes. Il s'agit, à notre avis, d'une des clés pour rehausser la qualité de l'enseignement et de l'apprentissage du français par les immigrants au Québec.

5.5 Perspectives de recherches

Cette étude est la première à se pencher sur le sujet du vocabulaire chez les étudiants adultes en francisation au Québec. Maintenant que la glace est brisée, il est important que les chercheurs s'intéressent à cette question et poursuivent le travail amorcé. Il va sans dire que de multiples questionnements demeurent et que les possibilités de nouvelles recherches sont nombreuses.

Comme nous l'avons mentionné dans les limites, il serait intéressant de mesurer le vocabulaire réceptif à l'oral, au moyen d'un test inspiré de l'Aural Lex. Cela donnerait l'occasion de vérifier s'il existe des différences entre la taille du vocabulaire oral et celle du vocabulaire écrit. Cela permettrait également d'observer si la corrélation avec la compréhension orale est plus élevée avec un test de vocabulaire oral qu'avec un test écrit.

Il serait très intéressant de mesurer le rythme d'apprentissage du vocabulaire des étudiants en faisant passer un test de vocabulaire aux étudiants de chaque niveau. Nous pourrions ainsi observer de quelle manière la taille du vocabulaire évolue :

augmente-t-elle graduellement d'un niveau à l'autre ou observe-t-on des effets de plateau?

Une autre étude, plus pratique, serait de demander à un enseignant d'intégrer un segment d'enseignement explicite du vocabulaire à chacune de ses leçons, à l'aide des listes de fréquence. L'accent pourrait être mis sur les 1000 premiers mots les plus fréquents, puisque notre étude a démontré qu'ils jouent un rôle prépondérant dans les habiletés réceptives. À la fin de la session, nous pourrions mesurer la taille du vocabulaire des étudiants ainsi que leurs résultats aux examens de compréhension orale et de compréhension écrite. La performance du groupe expérimental pourrait ensuite être comparée avec celle d'un groupe témoin.

Finalement, nous nous rappelons que la petite taille de l'échantillon ne nous a pas permis d'analyser l'effet de certains facteurs sociodémographiques tels que le sexe et l'âge. Une recherche future comptant davantage de participants pourrait se pencher sur le rôle de ces variables dans la taille du vocabulaire et les habiletés langagières.

5.6 Conclusion

Dans ce dernier chapitre consacré à la discussion, nous avons répondu à nos questions de recherche, en plus d'expliquer les limites de notre étude. Ceci fait, nous avons exposé les implications concrètes de nos découvertes sur l'enseignement du vocabulaire en classe de francisation. Pour conclure, nous avons proposé diverses avenues pour poursuivre les recherches dans ce domaine. Maintenant, nous sommes prête à conclure le mémoire.

CONCLUSION

Dans ce mémoire, nous nous sommes intéressée à la taille du vocabulaire et aux habiletés réceptives des adultes de niveau 6 qui terminent la formation en francisation. Le programme du MELS ne fixe pas d'objectif précis en terme de lexique à acquérir, pas plus qu'il ne mesure la connaissance du vocabulaire chez les étudiants. Or, nous jugeons important de vérifier si ces adultes possédaient une base de vocabulaire suffisante en français. La revue de la littérature a démontré la corrélation entre la taille du vocabulaire réceptif et la compréhension écrite. Elle a également souligné l'importance de la connaissance des 3000 premiers mots les plus fréquents, seuil à compter duquel les apprenants de langue seconde peuvent bénéficier de l'apprentissage par le contexte.

Notre étude indique que les apprenants de niveau 6 possèdent en moyenne un vocabulaire de 2228 mots. L'analyse par niveau de fréquence démontre que certains étudiants ne possèdent pas une connaissance suffisante des 1000 premiers mots les plus fréquents. Le déficit de certains étudiants devient encore plus flagrant au niveau 2000 et se généralise au niveau 3000. Ce manque de vocabulaire est un véritable problème, étant donné que notre analyse a confirmé la corrélation entre la taille du vocabulaire et les habiletés réceptives, tout particulièrement en compréhension orale. L'analyse a également révélé que la connaissance des premiers mille mots les plus fréquents est particulièrement importante, tant en compréhension orale qu'en compréhension écrite.

Notre étude s'est aussi penchée sur le rôle de certains facteurs démographiques sur la taille du vocabulaire et les habiletés réceptives. Nous avons pu démontrer que les apprenants ayant complété des études universitaires réussissent mieux en lecture. De plus, nous avons évoqué la possibilité que les individus de langue maternelle latine possèderaient davantage de vocabulaire et réussiraient mieux en lecture. La petite taille de l'échantillon a limité notre analyse, mais nous croyons avoir donné des pistes intéressantes pour de futures recherches.

Finalement, nous avons proposé divers moyens de pallier les lacunes en vocabulaire des étudiants. Nous croyons que tous les acteurs impliqués, en l'occurrence le MELS, les enseignants et les apprenants eux-mêmes, ont un rôle à jouer. Le Ministère doit d'abord définir des objectifs précis en matière de lexique, pour ensuite donner aux enseignants les moyens d'aider leurs étudiants à les atteindre.

En conclusion, il apparaît important de poursuivre les recherches en vocabulaire auprès des apprenants de francisation. Il reste beaucoup à faire afin de définir les meilleurs moyens d'améliorer l'enseignement du vocabulaire à nos immigrants.

ANNEXE A : LETTRE DE CONSENTEMENT

IDENTIFICATION

Chercheur responsable du projet : Isabelle Blais
Programme d'enseignement : Maîtrise en didactique des langues
Adresse courriel : blais.isabelle@courrier.uqam.ca

BUT GÉNÉRAL DU PROJET ET DIRECTION

Vous êtes invité(e) à prendre part à ce projet visant à mesurer la taille du vocabulaire des étudiants en francisation de niveau 6. Il vise également à comprendre comment le vocabulaire influence la compréhension écrite et la compréhension orale. Ce projet est réalisé dans le cadre d'un mémoire de maîtrise sous la direction de Tom Cobb, professeur du département de didactique des langues de la Faculté des sciences de l'éducation. Il peut être joint au (514) 987-3000 poste 2743 ou par courriel à l'adresse : cobb.tom@uqam.ca.

PROCÉDURE(S) OU TÂCHES DEMANDÉES AU PARTICIPANT

Votre participation consiste à faire un test de vocabulaire court et simple. Ce test, d'une durée approximative de 30 minutes, aura lieu dans votre salle de classe. La date et l'heure du test sont à confirmer par le responsable du projet. De plus, nous vous demandons d'autoriser le chercheur à avoir accès aux résultats des examens que vous passerez à la fin de la session.

AVANTAGES et RISQUES

Votre participation contribuera à l'avancement des connaissances par une meilleure compréhension de la relation entre la taille du vocabulaire et les habiletés langagières des étudiants de francisation. Il n'y a aucun risque lié à votre participation à cette étude.

ANONYMAT ET CONFIDENTIALITÉ

Il est entendu que les renseignements recueillis lors de la recherche sont confidentiels et que seuls, le responsable du projet et son directeur de recherche, Tom Cobb, auront accès à votre test de vocabulaire et aux résultats de vos examens finaux. Le matériel de recherche ainsi que votre formulaire de consentement seront conservés séparément sous clé par le responsable du projet pour la durée totale du projet. Les tests ainsi que les

formulaire de consentement seront détruits 5 ans après les dernières publications.

PARTICIPATION VOLONTAIRE

Votre participation à ce projet est volontaire. Cela signifie que vous acceptez de participer au projet sans aucune contrainte ou pression extérieure, et que par ailleurs vous êtes libre de mettre fin à votre participation en tout temps au cours de cette recherche. Dans ce cas les renseignements vous concernant seront détruits. 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 de vous identifier ne soit divulguée publiquement.

COMPENSATION FINANCIÈRE

Votre participation à ce projet est offerte gratuitement. 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 le responsable du projet pour des questions additionnelles sur le projet. Vous pouvez également discuter avec le directeur de recherche, Monsieur Tom Cobb, des conditions dans lesquelles se déroule votre participation et de vos 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.josée@uqam.ca).

REMERCIEMENTS

Votre collaboration est importante à la réalisation de ce projet et nous tenons à vous en remercier.

SIGNATURES :

Je reconnais avoir lu le présent formulaire de consentement et consens volontairement à participer 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 réfléchir à ma décision de participer. Je comprends que ma participation à cette recherche est totalement volontaire et que je peux y mettre fin en tout temps, sans pénalité d'aucune forme, ni justification à donner. Il me suffit d'en informer le responsable du projet.

Signature du participant :

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

QUESTIONNAIRE

Sexe : Masculin
 Féminin

Âge : _____

Langue maternelle : _____

Scolarité :

Diplôme d'études secondaires ou moins
Formation professionnelle ou études collégiales
Études universitaires non terminées
Études universitaires complétées

Contact avec la langue française :

À l'extérieur de l'école...

	Très souvent	Souvent	Parfois	Rarement/ Jamais
Vous parlez français	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vous écoutez du français	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vous lisez en français	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vous écrivez en français	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

ANNEXE C : TEST DE VOCABULAIRE (3 NIVEAUX)

Niveau 1k

Lisez attentivement les mots suivants et cochez ceux que vous connaissez.

Voici un exemple :

chien	<input checked="" type="checkbox"/>
-------	-------------------------------------

Notez que la liste contient des mots qui n'existent pas en français.

ville	<input type="checkbox"/>	manchir	<input type="checkbox"/>	part	<input type="checkbox"/>
bien	<input type="checkbox"/>	abjecter	<input type="checkbox"/>	participer	<input type="checkbox"/>
crétable	<input type="checkbox"/>	heure	<input type="checkbox"/>	divers	<input type="checkbox"/>
ostrat	<input type="checkbox"/>	entrée	<input type="checkbox"/>	prétemance	<input type="checkbox"/>
bois	<input type="checkbox"/>	style	<input type="checkbox"/>	pays	<input type="checkbox"/>
temps	<input type="checkbox"/>	commencer	<input type="checkbox"/>	naçon	<input type="checkbox"/>
métracte	<input type="checkbox"/>	abattir	<input type="checkbox"/>	admettre	<input type="checkbox"/>
fort	<input type="checkbox"/>	arriver	<input type="checkbox"/>	dont	<input type="checkbox"/>
charge	<input type="checkbox"/>	difficile	<input type="checkbox"/>	fil	<input type="checkbox"/>
reportation	<input type="checkbox"/>	source	<input type="checkbox"/>	nuit	<input type="checkbox"/>
dépriver	<input type="checkbox"/>	vernique	<input type="checkbox"/>	tentre	<input type="checkbox"/>
soi	<input type="checkbox"/>	jamais	<input type="checkbox"/>	début	<input type="checkbox"/>
aurois	<input type="checkbox"/>	face	<input type="checkbox"/>	ton	<input type="checkbox"/>
entrer	<input type="checkbox"/>	parfois	<input type="checkbox"/>	devenir	<input type="checkbox"/>
chaque	<input type="checkbox"/>	soudain	<input type="checkbox"/>	sureux	<input type="checkbox"/>
membre	<input type="checkbox"/>	fontre	<input type="checkbox"/>	lendemain	<input type="checkbox"/>
gillais	<input type="checkbox"/>	sécurité	<input type="checkbox"/>	anglais	<input type="checkbox"/>
prévoir	<input type="checkbox"/>	roile	<input type="checkbox"/>	pentrier	<input type="checkbox"/>
bientôt	<input type="checkbox"/>	moment	<input type="checkbox"/>	long	<input type="checkbox"/>
jousle	<input type="checkbox"/>	époque	<input type="checkbox"/>	vauche	<input type="checkbox"/>

Niveau 2k

Lisez attentivement les mots suivants et cochez ceux que vous connaissez.

Voici un exemple :

chien	<input checked="" type="checkbox"/>
-------	-------------------------------------

Notez que la liste contient des mots qui n'existent pas en français.

distance	<input type="checkbox"/>	dormir	<input type="checkbox"/>	roman	<input type="checkbox"/>
clair	<input type="checkbox"/>	aperne	<input type="checkbox"/>	rubir	<input type="checkbox"/>
soupaire	<input type="checkbox"/>	debout	<input type="checkbox"/>	convenir	<input type="checkbox"/>
ouest	<input type="checkbox"/>	valiant	<input type="checkbox"/>	intellectuel	<input type="checkbox"/>
plumbard	<input type="checkbox"/>	taire	<input type="checkbox"/>	octobre	<input type="checkbox"/>
transformer	<input type="checkbox"/>	satisfactoire	<input type="checkbox"/>	tirôt	<input type="checkbox"/>
odeur	<input type="checkbox"/>	peser	<input type="checkbox"/>	spécialement	<input type="checkbox"/>
mariage	<input type="checkbox"/>	réduction	<input type="checkbox"/>	lifrer	<input type="checkbox"/>
disabilité	<input type="checkbox"/>	luvois	<input type="checkbox"/>	angoisse	<input type="checkbox"/>
territoire	<input type="checkbox"/>	intéressé	<input type="checkbox"/>	retirer	<input type="checkbox"/>
bombe	<input type="checkbox"/>	complexe	<input type="checkbox"/>	oreille	<input type="checkbox"/>
plage	<input type="checkbox"/>	formule	<input type="checkbox"/>	expecter	<input type="checkbox"/>
feuille	<input type="checkbox"/>	loruste	<input type="checkbox"/>	siève	<input type="checkbox"/>
contrôle	<input type="checkbox"/>	changement	<input type="checkbox"/>	étoile	<input type="checkbox"/>
diroir	<input type="checkbox"/>	poids	<input type="checkbox"/>	bouche	<input type="checkbox"/>
immeuble	<input type="checkbox"/>	introids	<input type="checkbox"/>	quantité	<input type="checkbox"/>
bataille	<input type="checkbox"/>	courant	<input type="checkbox"/>	melindre	<input type="checkbox"/>
extrêmement	<input type="checkbox"/>	liste	<input type="checkbox"/>	asseoir	<input type="checkbox"/>
demure	<input type="checkbox"/>	nadoir	<input type="checkbox"/>	julir	<input type="checkbox"/>
acteur	<input type="checkbox"/>	dour	<input type="checkbox"/>	assurance	<input type="checkbox"/>

Niveau 3k

Lisez attentivement les mots suivants et cochez ceux que vous connaissez.

Voici un exemple :

chien	<input checked="" type="checkbox"/>
-------	-------------------------------------

Notez que la liste contient des mots qui n'existent pas en français.

abattre	<input type="checkbox"/>	dessus	<input type="checkbox"/>	inspirer	<input type="checkbox"/>
paquet	<input type="checkbox"/>	ultimation	<input type="checkbox"/>	grouper	<input type="checkbox"/>
juquette	<input type="checkbox"/>	situer	<input type="checkbox"/>	accrocher	<input type="checkbox"/>
fraction	<input type="checkbox"/>	détrame	<input type="checkbox"/>	rescuer	<input type="checkbox"/>
couloir	<input type="checkbox"/>	éliminer	<input type="checkbox"/>	fleuve	<input type="checkbox"/>
vicinité	<input type="checkbox"/>	partisan	<input type="checkbox"/>	debulant	<input type="checkbox"/>
incapable	<input type="checkbox"/>	délégué	<input type="checkbox"/>	avancé	<input type="checkbox"/>
brigeable	<input type="checkbox"/>	mérite	<input type="checkbox"/>	nutaindre	<input type="checkbox"/>
publication	<input type="checkbox"/>	argument	<input type="checkbox"/>	financement	<input type="checkbox"/>
formirique	<input type="checkbox"/>	piédeur	<input type="checkbox"/>	prévieux	<input type="checkbox"/>
outil	<input type="checkbox"/>	muscle	<input type="checkbox"/>	convenable	<input type="checkbox"/>
procédure	<input type="checkbox"/>	houroux	<input type="checkbox"/>	teinte	<input type="checkbox"/>
baser	<input type="checkbox"/>	surmonter	<input type="checkbox"/>	élu	<input type="checkbox"/>
boinase	<input type="checkbox"/>	trifler	<input type="checkbox"/>	eltrisse	<input type="checkbox"/>
coutume	<input type="checkbox"/>	creuser	<input type="checkbox"/>	mélange	<input type="checkbox"/>
vusier	<input type="checkbox"/>	outrir	<input type="checkbox"/>	détrame	<input type="checkbox"/>
infini	<input type="checkbox"/>	étonné	<input type="checkbox"/>	souffle	<input type="checkbox"/>
agiter	<input type="checkbox"/>	éphrasir	<input type="checkbox"/>	progresser	<input type="checkbox"/>
curiosité	<input type="checkbox"/>	taureau	<input type="checkbox"/>	psychologie	<input type="checkbox"/>
tube	<input type="checkbox"/>	caché	<input type="checkbox"/>	priet	<input type="checkbox"/>

BIBLIOGRAPHIE

Bauer, L. et Nation, P. 1993. «Word families». *International Journal of Lexicography*, vol. 6, no 4, p. 253-279.

Beeckmans, R. , Eyckmans, J. , Janssens, V. , Dufranne, M. et Van de Velde, H. 2001. «Examining the Yes/No vocabulary test: some methodological issues in theory and practice». *Language Testing*, vol. 18, no 3, p. 235-273.

Bonk, W. 2000. «Second language lexical knowledge and listening comprehension». *International Journal of Listening*, vol. 14, no 1, p. 14-31.

Cobb, T. 2000. «One size fits all? Francophone learners and English vocabulary size tests». *Revue canadienne des langues vivantes*, vol. 57, no 2, p. 295-324.

Cobb, T. et Horst, M. 2004. «Is there room for an AWL in French? » Dans Bogaards, P. et Laufer, B. (Eds.), *Vocabulary in a Second Language: Selection, acquisition, and testing*, Amsterdam: John Benjamins, p. 15-38.

David, A. 2008. «Vocabulary breadth in French L2 learners». *The Language Learning Journal*, vol. 36, no 2, p. 167-180.

Hu, M. et Nation, P. 2000. «Unknown vocabulary density and reading comprehension». *Reading in a foreign language*, vol. 13, no 1, p. 403-430.

Québec. Ministère de l'Immigration et des Communautés culturelles. Direction des affaires publiques et des communications. 2010. *Apprendre le Québec*. 167p. Récupéré le 23 mars 2012 de [http://www.immigration-quebec.gouv.qc.ca/publications/fr/divers/apprendre le quebec.pdf](http://www.immigration-quebec.gouv.qc.ca/publications/fr/divers/apprendre_le_quebec.pdf)

Québec. Ministère de l'Immigration et des Communautés culturelles. 2011. *Plan d'immigration du Québec pour l'année 2012*. 16 p. Récupéré le 23 mars 2012 de <http://www.micc.gouv.qc.ca/publications/fr/planification/Plan-immigration-2012.pdf>

- Le Bras, Y. et Lonsdale, D. 2009. *A Frequency Dictionary for French : Core Vocabulary for Learners*. London: Routledge Dictionary Series.
- Laufer, B. et Sim, D.D. 1985. «Reading and explaining the reading threshold needed for English for academic purposes texts». *Foreign Language Annals*, vol. 18, no 5, p. 405-4011.
- Laufer, B. 1992. «How much lexis is necessary for reading comprehension?» In *Vocabulary and Applied Linguistics*, eds. H. Bejoint and P. Arnaud. Macmillan. p. 126-132.
- 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. 2010. «Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension». *Reading in a foreign language*, vol. 22, no 1, p. 15-30.
- Milton, J. et Meara, P. 1995. «How periods abroad affect vocabulary growth in a foreign language». *Review of Applied Linguistics*, no 107-108, p. 17-34.
- Milton, J. 2009. *Measuring second language vocabulary acquisition*. Multilingual Matters. 278 p.
- Moon, R. 1997. «Vocabulary connections: multi-word items in English». Dans Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press. p. 40-63.
- Nation, P. et Waring, M. 1997. «Vocabulary size, text coverage and word lists». Dans Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge : Cambridge University Press, p. 6-19.
- OCDE. 2012. *Education indicators in focus*. 4 p. Récupéré le 20 novembre 2013 de <http://www.oecd.org/education/skills-beyond-school/49986459.pdf>
- O'Dell, F. 1997. «Incorporating vocabulary into the syllabus». In Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press. p. 258-278.
- Pellicer-Sanchez, A. et Schmitt, N. 2012. «Scoring Yes-No vocabulary tests: Reaction time vs. nonword approaches». *Language testing*, vol. 29, no 4, p. 489-509.

Read, J. 2000. *Assessing Vocabulary*. Cambridge : Cambridge University Press. 279p.

Schmitt, N. 1997. «Vocabulary learning strategies». . In Schmitt, N. et McCarthy, M., *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press. p. 199-227.

Schmitt, H. et Zimmerman, C. 2002. «Derivative word forms : What do learners know? » *TESOL Quarterly*, vol. 36, no 2, p. 145-171.

Schmitt, N., Jiang, X. et Grabe, W. 2011. «The percentage of words known in a text and reading comprehension». *The Modern Language Journal*, p. 26-43.

Sökmen, 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. 2008. «Vocabulary size and the skills of listening, reading and writing». *The Language Learning Journal*, 36, 2, p. 139-152.

Van Zeeland, H. et Schmitt, N. (2013) «Lexical coverage and L1 and L2 listening comprehension». *Applied Linguistics*, vol. 34, no 4, p. 457-479.