

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

QUAND LA LANGUE D'ENSEIGNEMENT N'EST PAS UNE LANGUE PREMIÈRE :
CONNAISSANCE DU VOCABULAIRE SCOLAIRE CHEZ DES ÉLÈVES
ALLOPHONES DU SECONDAIRE

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN LINGUISTIQUE

PAR
MICHEL BASTIEN

JANVIER 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse 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

Ce fut une grande entreprise. Si j'ai pu la mener à terme, dans un format réaliste et raisonnable, c'est grâce à l'indéfectible soutien et aux conseils avisés de plusieurs personnes. En ce qui concerne l'indéfectible soutien ET les conseils avisés, mes remerciements vont d'abord aux professeurs Lori Morris et Daniel Daigle. Leur confiance, leurs commentaires, constructifs et impitoyables, et leur incommensurable patience m'ont permis de réaliser un travail dont je suis aujourd'hui fier. Je remercie chaleureusement les professeurs Dominic Anctil, Tom Cobb, Pierre Cormier, Marlise Horst, Sophie Piron et Ophélie Tremblay, qui ont pris le temps d'évaluer cette thèse dans le plus grand respect et dans un esprit de collégialité, deux valeurs universitaires parmi les plus importantes à mes yeux.

Pendant plusieurs mois, j'ai pu améliorer des pans entiers de ma connaissance des statistiques grâce à M. Bertrand Fournier, maintenant à la retraite. Merci Bertrand, j'espère que tes loisirs ne te mettent pas trop dans le rouge.

Mon échantillon est grand. S'il est si grand, c'est grâce à N. Longpré, à ses collègues et à leurs élèves. Merci !

Des items ont été analysés, des questionnaires testés et commentés : merci à Jocelyne, Véronique, Caroline et Suzanne. Par ailleurs, cette dernière, fidèle à elle-même, est parvenue à traiter de la syntaxe dans une thèse sur le vocabulaire : encore merci !

Des collègues universitaires et professeurs ont bien voulu répondre aux différentes versions des tests de vocabulaire. Merci !

Ce projet de thèse a été rendu possible grâce au soutien financier du Fonds québécois de recherche Société et culture (FQRSC).

TABLE DES MATIÈRES

REMERCIEMENTS	ii
Table des matières	iii
Liste des figures	xi
Liste des tableaux	xiv
Résumé	xvii
Introduction	1
Chapitre I	
Problématique	5
1.1 Introduction	5
1.2 Caractéristiques de la langue dans le contexte scolaire	6
1.2.1 Langue maternelle et communication interpersonnelle	6
1.2.2 La langue à l'école : un langage décontextualisé	7
1.2.3 La littératie : un ensemble d'habiletés pour maîtriser le langage scolaire	8
1.2.4 Facteurs culturels et socioéconomiques	9
1.2.5 Importance de la lecture et du vocabulaire	10
1.2.6 Le développement du vocabulaire	10
1.2.7 Conclusion	12

1.3	Les élèves allophones	12
1.3.1	De qui s'agit-il ?	12
1.3.2	Description de l'immigration récente au Québec	13
1.3.3	Nombre et principaux lieux de résidence des immigrants au Québec	14
1.3.4	Pays d'origine des immigrants et des élèves allophones	15
1.3.5	Scolarité et connaissance du français des immigrants	16
1.3.6	Les langues parlées par les élèves	16
1.3.7	Répartition des élèves allophones dans le milieu scolaire	17
1.3.8	Conclusion	18
1.4	Le français, langue d'enseignement au Québec	18
1.5	Cheminement linguistique des élèves allophones dans le système scolaire québécois	19
1.6	Le contexte sociolinguistique d'apprentissage d'une langue seconde	21
1.6.1	Introduction	21
1.6.2	Les contextes de bilinguisme	21
1.6.3	Conclusion	24
1.7	L'acquisition de la langue d'enseignement par les allophones	25
1.8	La réussite scolaire chez les élèves allophones	26
1.9	Conclusion	29

Chapitre II

Cadre théorique	32
2.1 Introduction	32
2.2 Le vocabulaire	33
2.2.1 Le vocabulaire et la lecture	33
2.2.2 Le vocabulaire scolaire.....	36
2.3 La connaissance du vocabulaire	41
2.3.1 Connaitre un mot.....	41
2.3.2 L'étendue du vocabulaire	43
2.3.3 La profondeur de la connaissance du vocabulaire.....	45
2.3.4 Le vocabulaire vu comme un réseau	47
2.3.5 Les connaissances paradigmatiques	48
2.3.6 Les connaissances syntagmatiques.....	50
2.3.7 Un facteur à considérer : le degré d'abstraction d'un mot	52
2.3.8 Conclusion.....	56
2.4 Le développement du lexique en langue seconde (L2)	56
2.4.1 Un ou des lexiques ?.....	57
2.4.2 L'accès au mot en mémoire.....	58
2.4.3 Les représentations d'une langue à l'autre	58

2.4.4	Quatre modèles de lexique du locuteur bilingue	60
2.4.5	Conclusion.....	68
2.5	Mesurer la connaissance du vocabulaire	69
2.5.1	Mesurer l'étendue.....	70
2.5.2	Mesurer la profondeur de la connaissance	72
2.5.3	Administration papier ou assistée par ordinateur	76
2.5.4	Conclusion.....	77
2.6	Recension des écrits : la connaissance du vocabulaire chez les élèves allophones.....	78
2.6.1	Hacquebord (1994).....	79
2.6.2	Verhellen et Schoonen (Verhellen et Shoonen, 1998)	79
2.6.3	Ordonez, Carlo, Snow et McLaughlin (Ordóñez <i>et al.</i> , 2002)	81
2.6.4	Droop et Verhoeven (Droop et Verhoeven, 2003)	82
2.6.5	Lefrançois et Armand (Lefrançois et Armand, 2003)	83
2.6.6	Morris et Labelle (Morris et Labelle, 2008).....	84
2.6.7	Schoonen et Verhellen (Schoonen et Verhellen, 2008).....	84
2.6.8	Jean et Geva (Jean et Geva, 2008)	85
2.6.9	Bialystok, Luk, Peets et Yang (Bialystok, Luk, Peets et Yang, 2009)	87
2.6.10	Hellman (2011)	88

2.6.11 Conclusion.....	90
2.7 Objectifs de recherche.....	92
Chapitre III	
Méthodologie	97
3.1 Introduction	97
3.2 Éthique	98
3.3 Les participants	98
3.3.1 Description	98
3.3.2 Détermination des groupes d'élèves allophones et francophones	99
3.3.3 Description générale des élèves	100
3.3.4 Description des élèves allophones (groupe LM2).....	101
3.3.5 Déroulement	103
3.4 Le vocabulaire scolaire.....	104
3.5 La suite d'outils informatiques ÉVOCEL	109
3.6 Les outils de mesure	110
3.6.1 Mesure de l'étendue du vocabulaire.....	110
3.6.2 Mesure de la profondeur de la connaissance.....	111
3.6.3 Le questionnaire sociodémographique	116
3.7 Traitements des données, objectifs, variables et analyses.....	118

3.7.1 Objectif de recherche 1 : évaluer l'étendue du vocabulaire scolaire d'élèves allophones du secondaire	118
3.7.2 Objectif de recherche 2 : évaluer la profondeur de la connaissance du vocabulaire scolaire d'élèves allophones du secondaire, c'est-à-dire plus précisément évaluer les connaissances paradigmatiques et syntagmatiques de noms des domaines d'apprentissages du premier cycle du secondaire.....	119
3.7.3 Objectif de recherche 3 : mesurer la relation entre le degré d'abstraction des mots et la profondeur de la connaissance du vocabulaire des élèves.....	121
Chapitre IV	
Analyse des résultats	123
4.1 Introduction	123
4.2 Objectif de recherche 1 : évaluer l'étendue du vocabulaire scolaire des élèves allophones	123
4.3 Objectif de recherche 2 : évaluer la profondeur de la connaissance du vocabulaire scolaire des élèves allophones.....	128
4.3.1 Analyse des résultats, pour l'ensemble des domaines, selon la langue maternelle, pour la dimension SENS	129
4.3.2 Analyse de la dimension SENS, tous domaines, selon la langue maternelle, le niveau scolaire et le sexe.....	130
4.3.3 Analyse de la dimension SENS pour chacun des domaines d'apprentissages et les mots les plus fréquents selon la langue maternelle, le niveau de scolarité et le sexe	135
4.3.4 Analyse des résultats, pour l'ensemble des domaines, selon la langue maternelle, pour la dimension CONTEXTE	138

4.3.5 Analyse de la dimension CONTEXTE, tous domaines, selon la langue maternelle, le niveau et le sexe	139
4.3.6 Analyse de la dimension CONTEXTE pour chacun des domaines d'apprentissages et les mots les plus fréquents selon la langue maternelle, le niveau de scolarité et le sexe	145
4.3.7 Conclusion.....	157
4.4 Objectif de recherche 3 : mesurer la relation entre le degré d'abstraction des mots et la profondeur de la connaissance du vocabulaire des élèves.....	160
4.4.1 Description des résultats en fonction des facteurs lexicaux	160
4.4.2 Modélisation des effets des variables lexicales sur les résultats pour la variable SENS	168
4.4.3 Modélisation des effets des variables lexicales sur les résultats pour la variable CONTEXTE.....	171
4.4.4 Conclusion.....	173
4.5 Conclusion.....	176
Chapitre V	
Discussion	179
5.1 Introduction	179
5.2 État de la connaissance du vocabulaire scolaire des élèves allophones	180
5.2.1 L'étendue.....	180
5.2.2 La profondeur de la connaissance du vocabulaire.....	181
5.3 La connaissance du vocabulaire scolaire et le niveau de scolarité	184

5.4	La connaissance du vocabulaire et le sexe de l'élève.....	185
5.5	La connaissance du vocabulaire et les domaines d'apprentissages.....	186
5.6	Le vocabulaire scolaire et le lexique du locuteur bilingue.....	188
5.7	La connaissance du vocabulaire scolaire et le degré d'abstraction des mots	190
5.8	Le vocabulaire scolaire et le contexte sociolinguistique	192
5.9	L'impact éventuel de la connaissance du vocabulaire sur la lecture.....	193
5.10	L'impact éventuel de la connaissance du vocabulaire sur l'écriture	196
	Conclusion.....	197
	Références	234

LISTE DES FIGURES

Figure 3-1 - Moyennes des résultats en pourcentage pour les mots retenus en fonction du domaine (disciplines - mots fréquents), des connaissances paradigmatiques (P) et syntagmatiques (S) et du type de locuteur du français (LEL1 = lecteurs experts pour qui le français est une langue première ; LEL2 = lecteurs experts pour qui le français est une langue seconde)	115
Figure 4-1 Diagramme en boite des profils des résultats (scores en pourcentage) au test sur l'étendue du vocabulaire des élèves francophone (LM1) et allophones (LM2).....	125
Figure 4-2 Diagramme en boite des profils des résultats au test sur l'étendue du vocabulaire des élèves en fonction de la langue maternelle (LM1, LM2), du niveau scolaire (N1, N2, N3) et du sexe (F, M).	127
Figure 4-3 Diagramme en boite des profils des résultats pour la variable SENS, tous domaines, en fonction de la langue maternelle (LM1, LM2), du niveau scolaire (N1, N2, N3) et du sexe (F, M).	133
Figure 4-4 Illustration des moyennes des résultats pour la variable SENS en fonction des deux effets statistiques Niveau de scolarité (1 à 3) et Langue maternelle (1=francophones; 2=allophones). ...	135
Figure 4-5 Illustration des profils des résultats des élèves pour la variable SENS (connaissances paradigmatiques) en fonction de la langue maternelle, du niveau scolaire et des domaines d'apprentissages.	137
Figure 4-6 Diagramme en boite des profils des résultats pour la variable CONTEXTE, tous domaines, en fonction de la langue maternelle (LM1, LM2), du niveau scolaire (N1, N2, N3) et du sexe (F, M).	141
Figure 4-7 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Niveau de scolarité (1 à 3) et Langue maternelle (1=francophones; 2=allophones).	142
Figure 4-8 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 1	143
Figure 4-9 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 2.....	144

Figure 4-10 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 3.....	144
Figure 4-11 Illustration des moyennes des résultats pour la variable CONTEXTE en français en fonction des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2).....	148
Figure 4-12 Illustration des moyennes des résultats pour la variable CONTEXTE en Histoire en fonction des effets des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2) et de leur interaction.....	150
Figure 4-13 Illustration des moyennes des résultats pour la variable CONTEXTE en histoire en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 1.....	151
Figure 4-14 Illustration des moyennes des résultats pour la variable CONTEXTE en histoire en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 2.....	151
Figure 4-15 Illustration des moyennes des résultats pour la variable CONTEXTE en histoire en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 3.....	152
Figure 4-16 Illustration des moyennes des résultats pour la variable CONTEXTE en Science et technologie en fonction des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2).....	154
Figure 4-17 Illustration des moyennes des résultats pour la variable CONTEXTE pour les mots fréquents en fonction des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2).....	155
Figure 4-18 Illustration des profils des résultats des élèves pour la variable CONTEXTE (connaissances syntagmatiques) en fonction de la langue maternelle, du niveau scolaire et des domaines d'apprentissages.....	156
Figure 4-19 Illustration des profils des résultats des élèves pour les dimensions SENS (connaissances paradigmatiques) et CONTEXTE (connaissances syntagmatiques) en fonction de la langue maternelle et du niveau scolaire	159
Figure 4-20 Dispersion des résultats moyens en fonction du degré d'abstraction des mots pour la dimension SENS	161
Figure 4-21 Dispersion des résultats moyens en fonction du degré d'abstraction des mots pour la dimension CONTEXTE	162

Figure 4-22 Dispersion des résultats moyens en fonction de la fréquence des mots pour la dimension SENS	163
Figure 4-23 Dispersion des résultats moyens en fonction de la fréquence des mots pour la dimension CONTEXTE	164
Figure 4-24 Dispersion des résultats moyens en fonction de la longueur des mots pour la dimension SENS	165
Figure 4-25 Dispersion des résultats moyens en fonction de la longueur des mots pour la dimension CONTEXTE	165
Figure 4-26 Moyennes des résultats en fonction du degré d'abstraction et du niveau de scolarité pour la dimension SENS	175
Figure 4-27 Moyennes des résultats en fonction du degré d'abstraction et du niveau de scolarité pour la dimension CONTEXTE	176

LISTE DES TABLEAUX

Tableau 2-1 Connaitre un mot, en compréhension et en production, selon Nation (2001).....	42
Tableau 3-1 Distribution des élèves de l'échantillon en fonction du niveau de scolarité et du sexe.	99
Tableau 3-2 Distribution des élèves en fonction de la langue maternelle (LM), du niveau scolaire (1 à 3), de l'étendue des années de naissance et du sexe (filles, garçons).	100
Tableau 3-3 Pays de naissance des élèves allophones. Nombre de pays différents : 32.....	101
Tableau 3-4 Langues parlées à la maison chez les élèves allophones. Nombre de langues parlées différentes : 31.....	101
Tableau 3-5 Langues écrites par les élèves. Nombre de langues différentes: 24.....	102
Tableau 3-6 Nombre d'années du primaire en français chez les élèves allophones.	103
Tableau 3-7 - Nombre de mots total et nombre de noms pour chaque domaine.....	106
Tableau 3-8 Coefficients de corrélation de Spearman pour l'accord inter-juge (quatre juges : J1, J2, J3 et J4) sur le degré d'abstraction des noms.....	108
Tableau 3-9 Moyennes des indices lexicaux pour chaque domaine d'apprentissages et pour les mots les plus fréquents.....	109
Tableau 3-10 Procédures de cueillette d'informations relativement à la naissance.	116
Tableau 3-11 Compilation des résultats en fonction du domaine des mots et des variables dépendantes	120
Tableau 3-12 Les variables contrôle (VC), indépendantes (VI) et dépendantes prises en considération dans les analyse de variance pour l'objectif de recherche 2	121

Tableau 4-1 Statistiques descriptives des moyennes des résultats en pourcentage au test sur l'étendue du vocabulaire, en fonction de la langue maternelle (LM1= élèves francophones; LM2=élèves allophones).....	124
Tableau 4-2 Statistiques descriptives des moyennes des résultats en pourcentage au test sur l'étendue du vocabulaire, en fonction de la langue maternelle (LM1= élèves francophones; LM2=élèves allophones), du niveau scolaire (N1, N2, N3) et du sexe (F, M).	126
Tableau 4-3 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable SENS au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM1=élèves francophones; LM2=élèves allophones).....	130
Tableau 4-4 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable SENS au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).	132
Tableau 4-5 Moyennes des résultats en pourcentage pour la variable SENS au test sur la profondeur du vocabulaire pour chacun des domaines et pour les mots fréquents, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).	136
Tableau 4-6 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable CONTEXTE au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM1=élèves francophones; LM2=élèves allophones).....	138
Tableau 4-7 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable CONTEXTE au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).....	140

Tableau 4-8 Statistiques descriptives des moyennes des résultats en pourcentage pour la dimension CONTEXTE au test sur la profondeur du vocabulaire pour chacun des domaines et pour les mots fréquents du français, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).....	146
Tableau 4-9 Résumé des effets statistiques pour les variables Niveau de scolarité (1 à 3), Langue maternelle et Sexe sur les connaissances paradigmatisques.....	157
Tableau 4-10 Résumé des effets statistiques pour les variables Niveau de scolarité (1 à 3), Langue maternelle et Sexe sur les connaissances syntagmatiques.....	158
Tableau 4-11 Équations d'estimation généralisées pour les facteurs lexicaux (degré d'abstraction, fréquence, longueur) et les variables relatives aux élèves (niveau de scolarité, langue maternelle) sur la variable SENS.....	169
Tableau 4-12 Équations d'estimation généralisées pour les facteurs lexicaux (degré d'abstraction, fréquence, longueur) et les variables relatives aux élèves (niveau de scolarité, langue maternelle) sur la variable CONTEXTE.....	172
Tableau 4-13 Résumé des relations et interactions statistiques entre le degré d'abstraction et les autres variables lexicales, le niveau de scolarité et la langue maternelle pour les connaissances paradigmatisques (SENS) et les connaissances syntagmatiques (CONTEXTE).....	174

RÉSUMÉ

Le but de cette étude est de dresser un portrait de la connaissance du vocabulaire scolaire d'élèves allophones qui évoluent depuis plusieurs années dans le système scolaire francophone. Les élèves ont besoin de connaître les mots du vocabulaire scolaire pour comprendre les contenus scolaires. Parmi ces élèves, ceux qui parlent une ou des langues autres que le français à la maison et qui ne possèdent pas les mêmes bases lexicales que leurs camarades francophones au début de leur scolarité, sont particulièrement susceptibles de ne pas développer une connaissance suffisante de ce vocabulaire. Le nombre de mots connus autant que la profondeur de la connaissance de ces mots importent pour comprendre les textes scolaires. Au fur et à mesure que le nombre de mots connus augmente, des liens entre les mots se tissent, contribuant à approfondir la connaissance du vocabulaire. Dans notre travail, nous nous intéressons à la fois aux liens syntagmatiques et aux liens paradigmatiques, les premiers indiquant une connaissance lexicale de l'usage des mots en contexte, les seconds indiquant une connaissance lexicale du sens du mot au sein d'un système organisé, tel qu'il est présenté dans le monde scolaire. De plus, nous prenons en considération le degré d'abstraction des mots comme facteur d'apprentissage du vocabulaire, les mots plus abstraits étant plus difficiles à acquérir. La recension des écrits révèle que les élèves allophones ne semblent pas avoir développé une connaissance en profondeur des mots de la langue d'enseignement équivalente à celle développée chez les locuteurs natifs. Comme nous n'avons que très peu de données pour les élèves au secondaire, nous proposons les objectifs suivants : 1. Évaluer l'étendue du vocabulaire scolaire d'élèves allophones du premier cycle du secondaire. 2. Évaluer la profondeur de la connaissance du vocabulaire scolaire de ces élèves, c'est-à-dire plus précisément évaluer les connaissances paradigmatiques et syntagmatiques de noms des domaines d'apprentissages du premier cycle du secondaire. 3. Mesurer la relation entre le degré d'abstraction des mots et la profondeur de la connaissance du vocabulaire des élèves. Au total, 627 élèves de secondaire 1 à 3 âgés entre 12 et 17 ans fréquentant une école secondaire privée de Montréal ont répondu aux questionnaires et ont

procédé aux tests de vocabulaire. Les élèves proviennent d'un milieu multiethnique et socioéconomique moyen à élevé. Pour procéder aux diverses mesures, nous avons extrait un échantillon de mots du vocabulaire scolaire et développé une application informatisée disponible sur Internet mesurant la connaissance en étendue et en profondeur de ces mots. Pour mesurer l'étendue du vocabulaire, nous avons adapté un test de reconnaissance des mots. Les élèves doivent cocher oui ou non vis-à-vis d'items (des mots et des pseudomots) pour dire s'ils connaissent l'item ou non. Pour mesurer la profondeur de la connaissance, nous avons développé un test dans lequel les élèves doivent choisir deux mots qui partagent le sens avec un mot cible (connaissances paradigmatisques) et deux mots susceptibles d'apparaître dans le même contexte que le mot cible (connaissances syntagmatiques). Les résultats au test sur l'étendue montrent que les élèves allophones reconnaissent la plupart des mots. Les analyses de variance pour la mesure de la profondeur de la connaissance montrent des effets statistiques en fonction de la langue maternelle et du niveau de scolarité. Les élèves allophones obtiennent des résultats plus faibles que les élèves francophones à tous les niveaux de secondaire en ce qui concerne les connaissances paradigmatisques. Ils obtiennent également des résultats plus faibles pour les connaissances syntagmatiques, sauf en secondaire 3. Une modélisation linéaire d'équations d'estimation généralisées (GEE) montre que le degré d'abstraction est un facteur de la profondeur de la connaissance pour tous les élèves. Plus un mot est abstrait, moins il y a de chance de bien répondre, autant pour les connaissances paradigmatisques que pour les connaissances syntagmatiques. Nous soulignons l'impact négatif qu'une plus faible profondeur de la connaissance peut entraîner chez les élèves allophones en lecture comme en écriture. Par contre, les résultats des élèves allophones ne semblent pas plus affecté par le degré d'abstraction des mots que les élèves francophones, ce qui suggère qu'ils ont développé une représentation verbale des mots abstraits dans leur langue première sur laquelle ils appuient l'apprentissage des mots correspondant en français ou qu'ils parviennent à développer un lexique des mots abstraits directement en français, comme le font les locuteurs natifs.

Mots-clés : vocabulaire scolaire, allophone, profondeur de la connaissance lexicale, paradigmatique, syntagmatique

INTRODUCTION

Aller à l'école signifie poursuivre des activités pour développer des compétences dans plusieurs domaines du savoir humain. Dans chaque domaine d'apprentissages, de nombreux mots sont employés pour dénommer, décrire et expliquer les éléments et phénomènes propres au domaine. Ces mots constituent les assises sur lesquelles peuvent se développer les connaissances plus complexes relatives à un domaine. Ce sont des mots – *échantillon, procédé, époque* – spécialisés, plus rares et plus abstraits que les mots rencontrés quotidiennement à l'oral – *fille, garçon, matin*. Normalement, l'enfant qui commence l'école connaît déjà quelques-uns de ces mots, de même qu'il connaît suffisamment les structures syntaxiques de base de la langue d'enseignement pour consacrer son temps scolaire aux apprentissages relatifs aux domaines. Étayant son apprentissage sur les mots et structures qu'il connaît déjà, il développera de nouvelles compétences, dont l'apprentissage de nombreux nouveaux mots, au cours de son cheminement scolaire. Cependant, pour une catégorie d'élèves, la tâche peut être compliquée parce que ces élèves connaissent peu ou ne connaissent pas la langue d'enseignement quand ils commencent leur scolarité. Ce sont les élèves allophones, des élèves issus de l'immigration et dont le français est peu parlé ou n'est pas parlé du tout à la maison. C'est auprès de ces élèves que nous avons choisi de travailler dans le cadre de cette étude. Il ne s'agit pas pour nous d'évaluer les interventions qui ont eu lieu pendant la scolarité. Il s'agit de mieux décrire comment se produit l'acquisition du vocabulaire en langue seconde chez des élèves qui évoluent au sein d'un milieu dont la langue d'usage n'est pas la leur et quand cette langue est particulièrement apprise à partir de l'écrit. De plus, notre projet permet de mieux comprendre comment le lexique bilingue se développe dans la langue d'enseignement quand les mots sont plus rares, abstraits et spécialisés, éloignés du vocabulaire de tous les jours. Plus spécifiquement, notre étude contribue à l'avancement des connaissances sur l'importance relative des représentations syntagmatiques et paradigmatisques dans la profondeur de la connaissance du vocabulaire

scolaire de même que sur le rôle du degré d'abstraction comme facteur de la connaissance des mots.

Cette étude se divise en cinq parties. Dans la première partie, la problématique, nous expliquons plus en détail pourquoi nous procémons à une étude sur l'acquisition du vocabulaire scolaire chez les élèves allophones. Le vocabulaire scolaire constitue une dimension importante du langage scolaire, un langage que les élèves doivent apprendre et maîtriser pour réussir leur scolarité. Nous développons ce sujet dans cette partie puis décrivons la population visée par notre travail. Nous expliquons ensuite pourquoi les élèves allophones doivent apprendre le français, langue d'enseignement au Québec et soulignons les différents obstacles susceptibles de nuire à cet apprentissage. Cette partie nous amène à formuler notre question de recherche : Quel est le degré d'acquisition du vocabulaire scolaire chez des élèves allophones du système scolaire francophone québécois après plusieurs années au sein de ce système ?

Dans la deuxième partie, le cadre théorique, nous présentons les principaux concepts dont il est question dans notre travail et les relations entre ces concepts et la population étudiée. Nous définissons ainsi les concepts de vocabulaire scolaire et de connaissance du vocabulaire. Nous expliquons pourquoi nous devons mesurer autant l'étendue du vocabulaire que la profondeur de la connaissance de celui-ci. Nous précisons les dimensions les plus pertinentes pour le cadre de notre travail de la mesure de la profondeur de la connaissance du vocabulaire, c'est-à-dire les connaissances lexicales paradigmatiques et les connaissances syntagmatiques. Les premières correspondent à la connaissance des liens qui doivent être établis entre les mots pour mieux les organiser et les récupérer. Les secondes correspondent aux relations établies entre les mots se voisinant dans un contexte donné, permettant de constituer des unités de sens plus grandes. En plus de ces deux dimensions de la profondeur de la connaissance, nous soulignons que certaines caractéristiques des mots sont susceptibles d'influencer leur apprentissage, plus particulièrement leur degré d'abstraction. L'un de nos objectifs est de mesurer l'effet de ces caractéristiques sur les résultats au test sur la profondeur de la connaissance. Comme le vocabulaire scolaire est pour les élèves allophones un vocabulaire d'une langue seconde, nous brossons un portrait du développement du

vocabulaire en langue seconde à partir des modèles développés en recherche auprès des apprenants bilingues. Nous procédons ensuite à une recension des écrits pour rendre compte de l'état de la question chez les élèves allophones. Au terme de cette partie, nous formulons les trois objectifs de notre étude : mesurer l'étendue du vocabulaire, mesurer la profondeur de la connaissance du vocabulaire, et mesurer l'effet du degré d'abstraction des mots sur la profondeur de la connaissance chez des élèves allophones du secondaire.

Dans la partie suivante, la méthodologie, nous expliquons comment nous procédons pour atteindre nos objectifs. Nous décrivons d'abord les 627 élèves allophones et francophones d'une école secondaire privée de Montréal participant à notre étude. Les résultats des élèves francophones permettent d'établir une norme à laquelle nous comparons les résultats des élèves allophones. Nous décrivons ensuite les mots retenus et la suite informatique que nous avons développée pour mesurer l'étendue et la profondeur de la connaissance. Nous résumons également l'étude pilote que nous avons réalisée pour valider nos instruments. Nous complétons cette partie en expliquant comment nous traitons les données en fonction des différents objectifs, variables et analyses statistiques envisagés.

Dans la quatrième partie, l'analyse des résultats, nous décrivons les différents résultats obtenus pour chacun des objectifs. Les analyses de variance et de régression permettent d'établir des relations entre les résultats en fonction de variables comme la langue maternelle (allophone-francophone), le niveau de scolarité (secondaire 1 à 3), le sexe et le degré d'abstraction des mots. Les effets statistiques sont relevés et les plus importants soulignés à la fin de cette partie.

Au cours de la discussion, en cinquième partie, nous interprétons les différents résultats analysés dans la partie précédente. Nous répondons à notre question de recherche en résumant les résultats obtenus pour l'étendue du vocabulaire et pour la profondeur de cette connaissance. Nous expliquons ensuite les résultats en prenant en considération plusieurs facteurs, dont la langue maternelle, le niveau de scolarité, le sexe, les domaines d'apprentissages disciplinaires et le degré d'abstraction. Nous soulignons l'impact éventuel

de la connaissance du vocabulaire des élèves allophones sur la lecture et l'écriture. Nous complétons cette partie par un bilan critique de notre travail.

CHAPITRE 1

PROBLÉMATIQUE

1.1 Introduction

Dans cette première partie, nous démontrons l'importance de s'intéresser à la problématique de la connaissance du vocabulaire scolaire des élèves québécois allophones. Nos arguments reposent sur les préoccupations que pose l'acquisition de la langue d'enseignement quand elle est une langue seconde pour les élèves. Ces préoccupations sont soulevées depuis de nombreuses années. Messier (1997) rappelle que, dans les années 80, des études font état des inquiétudes des intervenants vis-à-vis des élèves allophones à l'école. Ces élèves vivent des difficultés d'intégration, accumulent du retard scolaire et abandonnent l'école plus tôt et plus fréquemment que les élèves scolarisés dans leur langue. Parmi les causes, des chercheurs ciblent les difficultés rencontrées par ces élèves dans l'apprentissage du langage scolaire et des référents culturels. Aujourd'hui, bien que nous en sachions un peu plus, des données manquent pour tracer un portrait linguistique des élèves allophones dans le réseau scolaire québécois, particulièrement à l'ordre secondaire. Plus précisément, nous ignorons le degré d'acquisition du vocabulaire scolaire chez ces élèves. Dans cette partie, nous montrons pourquoi la recherche doit s'intéresser à l'acquisition du vocabulaire scolaire chez les élèves allophones du premier cycle du secondaire. D'abord, nous expliquons que l'une des difficultés auxquelles sont confrontés tous les élèves est l'apprentissage du langage scolaire : le langage qui doit être appris à l'école est de plus en plus complexe et abstrait, notamment sur le plan du vocabulaire. Nous poursuivons en rappelant que la population immigrante est importante, particulièrement dans la grande région de Montréal ; nous procédons à une courte description démographique de cette population. Ensuite, nous rappelons que cette population évolue dans un contexte sociopolitique particulier et que ses enfants doivent, pour la plupart,

être scolarisés en français. Par la suite, nous décrivons le cheminement de ces enfants dans le système scolaire québécois. Nous rappelons que ces élèves se retrouvent potentiellement dans des situations de bilinguisme consécutif et de submersion dans le système scolaire, situations qui peuvent nuire à leur apprentissage du français et, éventuellement, à leur réussite scolaire. Un bilan complète cette partie en exposant le besoin de connaître l'état du vocabulaire en français des élèves allophones après plusieurs années passées dans le système scolaire francophone.

1.2 Caractéristiques de la langue dans le contexte scolaire

Au cours de sa scolarité, un élève doit développer une plus grande maîtrise de la langue, particulièrement à l'écrit. Contrairement au langage parlé à la maison, une partie du langage auquel est exposé l'élève à l'école est produit de façon décontextualisée (Armand et d'Anglejan, 1996a; Cummins, 2000; Giasson, 2011). Dans cette section, nous rappelons d'abord en quoi consiste l'acquisition de la langue pour un locuteur natif, en mettant l'accent sur ce qui distingue le langage scolaire de la langue parlée à la maison. Nous introduisons ensuite le concept de littératie, souvent utilisé pour nommer l'ensemble des habiletés attendues pour la maîtrise de l'écrit. Ensuite, nous énumérons quelques facteurs qui interviennent dans l'apprentissage de cette maîtrise, dont le besoin pour les élèves de développer un vocabulaire important. Nous terminons la section en rappelant que les élèves allophones commencent l'école désavantagés sur le plan linguistique comparativement aux locuteurs natifs.

1.2.1 Langue maternelle et communication interpersonnelle

Apprendre sa langue maternelle s'étend sur une douzaine d'années (Collier, 1989; Cummins, 2000). De la naissance à cinq ans, l'enfant apprend l'essentiel de la phonologie de sa langue maternelle, le vocabulaire, la grammaire, des connaissances sémantiques et pragmatiques (Collier, 1989), de même que plusieurs règles d'usage dans des contextes familiers (Cummins, 2000). Quand l'enfant commence l'école, ce développement n'est pas terminé. De 6 à 12 ans, l'enfant acquiert des règles plus complexes de syntaxe et de morphologie de

même que certains aspects de la phonologie ; il poursuit également le développement de son vocabulaire, ce qu'il fera tout au long de sa vie (Collier, 1989; Cummins, 2000).

Ainsi, pendant les premières années de sa vie, l'enfant développe ce que Cummins (2000) appelle les habiletés de communication interpersonnelle de base (BICS – *Basic interpersonal communicative skills*) : 'la compétence langagière acquise à travers les interactions interpersonnelles par la plupart des enfants avant de commencer l'école. Par contre, l'entrée à l'école va placer l'enfant devant de nouvelles compétences linguistiques à acquérir. Cummins appelle cet ensemble de compétences la compétence cognitive et langagière scolaire (CALP – *Cognitive academic language proficiency*). Ce découpage des habiletés linguistiques par Cummins ne fait pas l'unanimité chez les chercheurs, mais il semble y avoir un consensus sur le fait que l'école, du moins en occident, amène les élèves à développer une connaissance de la langue « décontextualisée » (Armand et d'Anglejan, 1996a; Cummins, 2000; Hacquebord, 1994; Painchaud, d'Anglejan, Armand et Jezak, 1994).

1.2.2 La langue à l'école : un langage décontextualisé

Armand et d'Anglejan (1996a) rappellent ce que plusieurs auteurs avant elles ont entendu par « langue décontextualisée ». Il s'agit d'une « compréhension du discours plus analytique, outil de construction et de transmission du savoir dans nos sociétés occidentales » (Armand et d'Anglejan, p. 74). La langue dite « contextualisée » se produit dans un contexte interpersonnel de communication où la « signification est négociée par les interlocuteurs et la transmission du message favorisée par de multiples indices situationnels et paralinguistiques (gestes, intonation, mimiques) » (Armand et d'Anglejan, p. 74). La langue décontextualisée, elle, « repose sur des indices situationnels ou contextuels réduits ou inexistant. La transmission du message s'appuie essentiellement sur le contenu linguistique et passe par un discours plus analytique » (Armand et d'Anglejan, p. 74).

Pour l'enfant qui évolue dans le contexte scolaire, cela signifie d'importants apprentissages sur le plan langagier. Il doit apprendre des subtilités de la phonologie de la langue et une grande quantité de vocabulaire ; il doit développer davantage ses connaissances sémantiques,

de la syntaxe de la langue, des structures du discours (au-delà de la phrase) et de la pragmatique (Cummins, 2000 ; Painchaud *et al.*, 1994). En plus, tout le système écrit est appris et développé, ce qui signifie développer des habiletés complexes d'écriture (Cummins, 2000 ; Painchaud *et al.*, 1994). La maîtrise du langage scolaire est utile à l'élève dans chacun des domaines d'apprentissages disciplinaires : chacun a ses usages et sa terminologie qui doivent être appris. Finalement, l'élève doit développer ses capacités métalinguistiques dans les cours de langue ou développer d'autres stratégies d'apprentissage dans d'autres domaines (Cummins, 2000).

1.2.3 La littératie : un ensemble d'habiletés pour maîtriser le langage scolaire

Pour maîtriser la langue décontextualisée, l'élève doit développer plusieurs habiletés associées à l'apprentissage du code écrit. L'ensemble de ces habiletés est appelé littératie (de *literacy* en anglais) (Armand, 2005; Armand et d'Anglejan, 1996b; August et Shanahan, 2006; Desrochers et Berger, 2011; Painchaud *et al.*, 1994). Selon l'Organisation de coopération et de développement économiques (OCDE), la littératie est « l'aptitude à comprendre et à utiliser l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités » (OCDE, s. d.). L'école en particulier privilégie l'usage de l'écrit comme moyen et objet d'apprentissages (Armand et d' Anglejan, 1996b; Desrochers et Berger, 2011).

Ce qui rend la situation plus difficile pour de nombreux élèves de certains milieux socioculturels, c'est qu'ils n'ont pas accès à ce vocabulaire en dehors de l'école (Corson, 1997). Le système de pensée sur lequel repose le monde scolaire est plus accessible à ceux que l'expérience à l'extérieur de l'école lie à ce système (Corson, 1997). Les systèmes de pensée occidentaux (auxquels réfèrent les différentes disciplines scolaires) sont enracinés dans l'histoire des institutions occidentales. Ceux qui grandissent dans la culture occidentale sont plus à même de l'apprendre, de la comprendre, puis de l'utiliser. Comme les mots du système de pensée scolaire se rencontrent essentiellement à l'écrit, pour avoir accès à la pensée scolaire, il faut participer à des activités de littératie. Les enfants dont la culture de la

littératie est importante à la maison sont mieux placés que ceux dont ce n'est pas le cas pour développer cette pensée (Corson, 1997).

Bien que les attentes vis-à-vis le développement de la littératie varient d'une culture à l'autre (Painchaud *et al.*, 1994), il semble exister un consensus chez les chercheurs quant aux habiletés à développer pour la maîtriser. À partir des études retenues dans le rapport « *Developing Literacy in Second-Language Learners* », August et Shanahan (August et Shanahan, 2006) ont regroupé ces habiletés ainsi : habiletés de prélecture, c'est-à-dire avoir une conception de l'écrit et de l'alphabet ; habiletés au niveau du mot, dont son décodage, sa lecture, la lecture de pseudomots, l'épellation ; habiletés au niveau du texte : vitesse de lecture, compréhension du texte ; habiletés d'écriture. Comme la compétence à l'oral est un facteur favorable au développement de la littératie, August et Shanahan ont identifié les habiletés relatives à cette compétence dans les diverses études : phonologie, vocabulaire, morphologie, grammaire et discours, en compréhension comme en production.

1.2.4 Facteurs culturels et socioéconomiques

En plus du développement des habiletés linguistiques nommées plus haut, d'autres facteurs interviennent dans le degré de maîtrise du langage scolaire chez les élèves. Des facteurs comme un statut socioéconomique familial peu élevé (Chall, Jacobs et Baldwin, 1990), une culture familiale et un milieu de vie qui emploient peu le langage scolaire (Armand et d'Anglejan, 1996 ; Chall *et al.*, 1990) semblent nuire à son développement. Des attitudes favorables à la littératie, une culture de l'écrit et des attentes manifestes quant à la réussite scolaire à la maison la favorisent (Chall *et al.*, 1990).

De plus, des caractéristiques mêmes de la langue dans le contexte scolaire semblent constituer un obstacle à son apprentissage, particulièrement après la quatrième année du primaire (Chall *et al.*, 1990). Avant la quatrième année, l'élève apprend à reconnaître des mots dont il connaît déjà le sens. Au-delà, le contenu des textes scolaires est plus complexe, technique et abstrait et dépasse l'expérience quotidienne vécue par la plupart des élèves.

L'élève doit alors développer ses connaissances, le langage et son vocabulaire, de même que sa capacité à réfléchir de façon critique.

1.2.5 Importance de la lecture et du vocabulaire

Selon Chall *et al.* (1990), après la quatrième année du primaire, de nombreux élèves expérimentent une chute particulière de la performance scolaire, souvent appelée « *Fourth grade slump* » (le déclin de la quatrième année, littéralement). Pour l'expliquer, Chall *et al.* (1990) rappelle que l'élève qui avance dans sa scolarité doit lire des textes dont les mots sont de moins en moins familiers et fréquents. Autour de la quatrième année, un changement important se produit : la tâche principale n'est plus d'apprendre à décoder (et à produire) le médium (le texte écrit) mais à comprendre le message véhiculé par le médium. L'élève est sensé avoir appris à lire pour apprendre à partir des textes ; le contenu notionnel des textes lus va au-delà de ce qui est déjà su, linguistiquement et cognitivement, par le lecteur, avec des nouveaux mots et de nouvelles idées. Les phrases sont plus longues et plus complexes. Les idées véhiculées par le texte sont plus difficiles à comprendre. Dans les années scolaires subséquentes, les textes et le matériel scolaire lus sont encore plus variés et complexes du point de vue de leur contenu, de la langue et de la demande cognitive nécessaire pour traiter le texte. Alors que la charge cognitive dans le traitement des textes et de la matière scolaire augmente, de nombreux élèves éprouveront des difficultés avec les mots abstraits, littéraires et peu fréquents.

1.2.6 Le développement du vocabulaire

Même s'ils sont peu fréquents, les mots rencontrés dans le cadre scolaire sont nombreux (Lété, Sprenger-Charolles et Colé, 2004; Nagy et Anderson, 1984). Pour lire et écrire dans ce cadre, les élèves doivent développer autant l'étendue de leur vocabulaire (le nombre de mots connus) que la profondeur de la connaissance de ce vocabulaire (Aitchison, 2005; Nation, 2001; Qian, 2002; Verhellen et Shoonen, 1998). Comment procéder ? Les recherches montrent que, chez les adultes, les mots sont organisés et marqués en mémoire pour être

récupérés rapidement (Aitchison, 2005). Développer le vocabulaire ne consiste pas seulement à mémoriser la forme des mots. Il faut connaître les mots sur les plans :

- de la forme : orale, écrite et dans la forme de ses parties
- du sens : la forme et son sens; le concept et les référents; les associations
- de l'usage : les fonctions grammaticales, les collocations, les contraintes d'utilisation (registre, fréquence) (Nation, 2001; Thornbury, 2002).

L'apprentissage des mots n'est pas instantané. Thornbury (2002) rappelle qu'il est progressif et qu'il prend du temps. Par ailleurs, Nation (2001) parle du fardeau de l'apprentissage d'un mot. Il s'agit de l'effort nécessaire pour l'apprendre. Différents mots sont plus ou moins difficiles à apprendre, selon les apprenants et leur expérience langagière.

L'apprentissage des mots et leur organisation en mémoire commence tôt dans la vie de l'élève. Singleton (Singleton, 1999) rappelle que l'enfant procède à une révision, à une réorganisation et à une consolidation de ses connaissances après l'explosion lexicale de la petite enfance, peu avant de commencer l'école. Cela se poursuivra tout au long de la scolarité. Alors que de nouveaux mots sont appris chaque jour, le lexique est constamment réorganisé, les mots étant reliés entre eux, et des relations plus complexes étant établies entre les mots et des domaines de mots et entre les concepts.

Le développement du lexique comprend donc une perception croissante des relations syntaxiques, sémantiques et conceptuelles entre les mots. Nous développons ce sujet dans le cadre théorique. Pour résumer, il suffit de dire ici qu'au cours de leur scolarité, les élèves doivent apprendre une grande quantité de mots (développer l'étendue de leur vocabulaire, donc) et établir des relations syntaxiques, sémantiques et souvent abstraites entre les mots (développer une connaissance en profondeur des mots, donc) pour pouvoir comprendre et produire le langage scolaire.

1.2.7 Conclusion

À la lumière de ce qui précède, nous observons que la réussite scolaire de l'élève passe par la maîtrise du langage scolaire. Cette maîtrise est caractérisée par une connaissance d'un large vocabulaire rencontré essentiellement à l'écrit, des habiletés morphosyntaxiques complexes et plusieurs habiletés métalinguistiques. Passée la quatrième année du primaire, de nombreux élèves dont la langue d'enseignement est la langue première éprouvent des difficultés scolaires parce qu'ils ne développent pas un vocabulaire suffisant pour comprendre les textes écrits rencontrés au cours de leur scolarité. Si c'est le cas pour ces élèves, il est permis de croire que la situation est pire pour des élèves qui commencent l'école sans bien connaître la langue d'enseignement. Développer un vocabulaire suffisant signifie à la fois de le développer en étendue (connaître un grand nombre de mots) et en profondeur (établir des liens sémantiques et syntaxiques entre les mots).

1.3 Les élèves allophones

1.3.1 De qui s'agit-il ?

La population étudiée dans ce travail évolue dans le milieu scolaire québécois. Il ne s'agit pas d'une population homogène. Il s'agit d'élèves qui se distinguent par des lieux de naissance, des langues et des cultures divers (MELS, 2006). Il existe plusieurs désignations pour identifier cette population. D'une part, le MELS regroupe les membres de cette population sous l'appellation « Issus de l'immigration ». Ces élèves sont nés à l'extérieur du Canada, et sont identifiés comme étant de première génération, ou sont nés au Canada, et dits de deuxième génération (MELS, 2006). Dans ce dernier cas, l'un des parents doit être né à l'extérieur du Canada, ou n'avoir comme langue maternelle ni le français ni l'anglais. Selon cette définition, des élèves issus de l'immigration peuvent avoir comme langue maternelle l'anglais ou le français, s'ils sont de première génération ou si l'un des parents, la mère en particulier, parle l'une de ces langues.

D'autre part, des chercheurs oeuvrant auprès de cette population adoptent une définition sur la base de critères linguistiques, sans tenir compte du lieu de naissance, pour identifier ces

membres (Armand, Beck et Murphy, 2009; Morris et Labelle, 2008). Il est question dans leurs travaux d'« élèves allophones ». Ce sont les élèves « qui parlent une langue autre que le français dans leur milieu familial » (Morris et Labelle, 2008, p. 9), qui peuvent être « nés dans des familles déjà installées au Québec [...] où le français n'est pas la langue dominante de communication et où la culture dans laquelle baigne l'enfant avant son arrivée à l'école peut se distinguer nettement de la culture franco-qubécoise typique » (Morris et Labelle, 2008, p. 9) ou les élèves dont la langue maternelle n'est ni le français, ni l'anglais (Armand *et al.*, 2009).

Par ailleurs, d'autres expressions ont cours : il sera à l'occasion question d'élèves issus des minorités linguistiques pour parler des élèves allophones ou d'élèves des minorités culturelles, mais dont le français peut être parlé à la maison dans le cas de ces derniers. Les élèves dont la langue d'enseignement est la langue maternelle sont identifiés comme les élèves de la majorité linguistique ou locuteurs natifs (sous-entendu : de la langue d'enseignement). Aux États-Unis, l'expression *Heritage Language* est utilisée pour désigner un élève dont la langue parlée à la maison n'est pas l'anglais, qui parle ou du moins comprend cette langue et qui est bilingue dans cette langue et en anglais (Valdés, 2001). En français, l'expression a été traduite par « langue d'origine » (par exemple, dans Zouali, 1997). Finalement, une autre façon de distinguer les élèves allophones des élèves issus de la majorité linguistique est de les surnommer respectivement les élèves L2 (langue seconde) et les élèves L1 (langue première).

Dans la description qui suit de la population immigrante au Québec, autant le terme « allophones » que l'expression « issus de l'immigration » sont employés, selon qu'il s'agisse de personnes dont la langue maternelle n'est pas le français dans le premier cas ou des personnes visées par les documents du MELS sur l'immigration dans le second.

1.3.2 Description de l'immigration récente au Québec

Le Québec a de tout temps été une terre d'immigration. Peuplé d'abord par les peuples d'origine asiatique, depuis la colonisation européenne il est occupé principalement par des

immigrants de langues française et anglaise (MELS, 1998). Jusque dans les années 60, des immigrants en provenance d'Europe et d'Amérique du Nord s'ajoutent à la population. Depuis les années 60, le Québec accueille des immigrants en provenance d'Asie, des Antilles, d'Afrique et d'Amérique du Sud (ICC, 2008; MELS, 1998, 2006). Chaque année depuis 1996, plus de 30 000 immigrants se sont établis au Québec (MELS, 2006). En 2008, le Québec en a reçu 45 264 (Armand, Beck et Murphy, 2009). À leur arrivée, un peu plus de la moitié a déclaré connaître le français (Armand *et al.*, 2009). Depuis cinquante ans, ce sont tout de même des dizaines de milliers d'immigrants qui ne connaissent pas le français qui se sont établis au Québec. Ces immigrants ont eu des enfants, à l'extérieur du Québec ou au Québec. Ces enfants ont évolué dans un milieu familial non-francophone dans une province d'accueil qui est à majorité francophone.

1.3.3 Nombre et principaux lieux de résidence des immigrants au Québec

De 1997 à 2006, le Québec a accueilli 292 278 immigrants selon Immigration et communautés culturelles (2008). Près de 80 % des immigrants s'établissent dans la région métropolitaine de Montréal (RMM). La RMM comprend les agglomérations de Montréal, de Longueuil et de Laval. À leur arrivée, les immigrants s'établissent surtout à Montréal (68,3 %). Ensuite, certains (environ 7 %) quittent l'île de Montréal pour s'établir à Laval ou à Longueuil ou s'éloignent de la RMM (ICC, 2008). Le tableau 1-1 présente la distribution des immigrants de 1997 à 2006, selon le groupe d'âge, pour les trois agglomérations et pour la RMM entière.¹

¹ Dans la présentation des statistiques du ministère Immigration et communautés culturelles, nous nous en tenons aux données relatives à la région métropolitaine pour ne pas alourdir le texte de trop nombreuses données. D'autant plus que notre étude porte sur des élèves de la région métropolitaine. Nous procédons de même avec les données du MELS quand c'est possible.

Tableau 1-1 Lieu de résidence des immigrants par agglomération et pour la RMM de 1997 à 2006 selon le groupe d'âge (ICC, 2008)

Groupe d'âge	Montréal	Laval	Longueuil	Total RMM
0-14 ans	35 688	4 637	4 568	44 893
15-24 ans	27 952	2 626	2 440	33 018
25-34 ans	74 039	6 908	5 762	86 709
35-44 ans	36 159	3 445	3 501	43 105
45-64 ans	14 913	1 407	1 534	17 854
65 ans et +	2 636	444	373	3 453
Information non disponible	42	3	11	56
Total	191 429	19 470	18 189	229 088

Ces données permettent de constater que la population immigrante est relativement jeune et comprend de nombreux enfants et adolescents qui ont joint ou joindront le système scolaire de la région.

1.3.4 Pays d'origine des immigrants et des élèves allophones

Les pays d'origine des immigrants couvrent tous les continents. Ils proviennent en majorité d'Asie, ensuite d'Afrique, suivie de l'Europe et de l'Amérique puis, dans une faible proportion, de l'Océanie. Cependant, les enfants de ces immigrants ne sont pas tous nés à l'étranger. Selon le MELS (MELS, 2006), la majorité des élèves issus de l'immigration sont nés au Québec. Le tableau 1-2 montre les plus importantes régions d'origine des élèves issus de l'immigration en 2003- 2004.

Tableau 1-2 Principales régions de naissance des élèves issus de l'immigration en 2004 (MELS, 2006)

Nés au Québec	Asie orientale	Europe occidentale	Afrique du Nord	Caraïbes et Bermudes	Europe orientale	Autres
63,1 %	3,8 %	3,5 %	3,4 %	3,2 %	3,1 %	Moins de 2 %
	République populaire de Chine, incluant Hong Kong, Taïwan, etc.	Principalement Allemagne de l'Ouest et Belgique	Principalement Maroc et Algérie	Haïti, République Dominicaine, Jamaïque, etc.	Roumanie, Russie, Bulgarie, etc.	

Le nombre d'élèves issus de l'immigration en fonction des pays d'origine fluctue. De 1994 à 2004, le MELS (2006) observe une diminution en provenance des Bermudes et des Caraïbes

et une augmentation des élèves en provenance d'Asie, d'Europe occidentale et d'Afrique du Nord.

1.3.5 Scolarité et connaissance du français des immigrants

Le MELS (2006) mentionne que la connaissance préalable de la langue française est un élément favorable à l'immigration au Québec. Pour le gouvernement, la connaissance du français est un facteur essentiel d'intégration des immigrants au Québec (ICC, s. d.). De plus, les immigrants doivent manifester une volonté d'assurer la continuité et la stabilité du fait français (ICC, s. d.). La vaste majorité des immigrants connaissent le français ou déclarent le français comme langue maternelle (ICC, 2008). Par contre, il faudrait savoir dans quelle mesure le français est ou n'est pas maîtrisé. Nous n'avons pas vu de données publiées à ce sujet.

En plus de la connaissance du français, la scolarité est un critère important de sélection pour l'immigration. Selon le ministère de l'Immigration et Communautés culturelles (MICC, site web), là « scolarité est considérée seulement si le candidat a réussi un programme d'études sanctionné par un diplôme. Le diplôme étranger est évalué en fonction de sa correspondance dans le système d'éducation québécois. [...] Pour avoir des chances d'être sélectionnés, les candidats doivent minimalement détenir un diplôme correspondant à un diplôme d'études secondaires générales ou professionnelles du Québec ». Selon les données du MICC, les immigrants ont en majorité plus de 12 ans de scolarité, ce qui correspond au moins aux nombres d'années d'un diplôme d'études secondaires professionnelles. Par contre, près de 22 % ont 11 ans et moins de scolarité.

1.3.6 Les langues parlées par les élèves

Les élèves issus de l'immigration ont déclaré 237 langues depuis 1994. De ce nombre, 80 % déclarent l'une des langues suivantes : français, anglais, arabe, espagnol, italien, créole, chinois, vietnamien, grec et portugais. Le français (29 %) est la langue maternelle la plus importante chez les élèves issus de l'immigration selon les données du MELS (2006). Une autre langue maternelle est déclarée chez 71 % des élèves issus de l'immigration. Par contre,

le MELS (2006) rappelle que le système de déclaration de l'effectif scolaire des jeunes en formation générale, qui recueille des données sur les élèves de la formation générale des jeunes, ne permet la déclaration que d'une seule langue maternelle. Si plus d'une langue est parlée à la maison, le système de déclaration ne permet pas de le savoir.

En interrogeant la base de données de Statistiques Canada (site web), il est possible d'avoir une idée de la langue maternelle et des langues parlées à la maison. Quand la langue maternelle n'est pas le français, le français est peu susceptible d'être parlé à la maison : chez 34 % des gens de la région métropolitaine, le français n'est pas la langue maternelle et il n'est pas parlé à la maison chez 29,52 % des gens.

1.3.7 Répartition des élèves allophones dans le milieu scolaire

Selon le MELS (2006), en 1994-1995, les élèves issus de l'immigration comptent pour 13,7 % de l'ensemble des élèves du préscolaire, du primaire et du secondaire (158 910 élèves contre 997 544). En 2003-2004, la proportion a grimpé à 18,1 % et les enfants des immigrants sont au nombre de 201 314 à fréquenter l'école. De ce nombre, 85 % fréquentent l'école publique et 15 % l'école privée (comparativement à 89,3 % dans le réseau public pour l'ensemble des élèves). Toujours en 2003-2004, les élèves issus de l'immigration de première génération fréquentent une école de langue française dans une proportion d'environ 90 %. Ceux de deuxième génération fréquentent une telle école dans une proportion d'environ 68 %, comparativement à environ 88 % pour l'ensemble des élèves. Nous rappelons que les élèves de première ou de deuxième génération, tel que défini par le MELS, peuvent avoir le français ou l'anglais comme langue maternelle.

Sur l'île de Montréal, les élèves issus de l'immigration constituent de 45 % à 53 % des élèves montréalais entre 1994-1995 et 2003-2004. Les régions de la Montérégie et de Laval accueillent généralement environ 20 % des élèves issus de l'immigration (MELS, 2006). Les données amenées par Armand *et al.* (2009) montrent que la proportion des élèves allophones dans le système scolaire à Montréal semble avoir légèrement diminué : en 2008, ces élèves représentent 46,37 % de l'effectif scolaire à la Commission scolaire de Montréal (CSDM),

43,88 % de celui de la Commission scolaire English Montréal (CSEM) et 43,82 % de celui de la Commission scolaire Marguerite-Bourgeoys (CSMB). Peut-être la population s'est-elle déplacée en banlieue, les statistiques sur le déplacement des familles d'immigrants vers les couronnes nord et sud de Montréal le suggérant (ICC, 2008).

1.3.8 Conclusion

Ce tableau de l'immigration récente au Québec montre que, malgré des critères favorisant l'immigration francophone, le Québec accueille des milliers d'immigrants allophones chaque année. Une bonne part de ces immigrants est scolarisée, mais un nombre important d'immigrants ont peu fréquenté l'école. Les enfants des immigrants joignent les rangs scolaires chaque année. Alors que le français n'est pas d'usage à la maison chez beaucoup d'entre eux, le français est la langue principale parlée et valorisée au Québec, particulièrement à l'école, comme nous l'exposons dans la prochaine section.

1.4 Le français, langue d'enseignement au Québec

Au Québec, la population francophone, majorité sociologique et linguistique, valorise sa langue et veut en assurer le maintien par, entre autres, l'obligation de fréquenter l'école française pour toute la population scolaire québécoise (Armand, 2005; MELS, 1998). La Charte de la langue française favorise le statut du français en tant que langue de scolarisation, langue de travail et langue commune de la vie publique (Armand, 2005).

En ce qui concerne la scolarisation, le ministère de l'éducation, des loisirs et des sports (MELS) établit comme principes éducatifs l'égalité des chances, la maîtrise du français et l'éducation à la citoyenneté démocratique dans un contexte pluraliste. L'une des missions de l'école est l'instruction, c'est-à-dire l'acquisition des connaissances, des habiletés et des attitudes nécessaires pour comprendre et transformer le monde ainsi que pour continuer à apprendre tout au long de sa vie (MELS, 1998). De plus, l'école doit contribuer à la maîtrise du français et à son utilisation, puisqu'elle est « la langue commune, grâce à laquelle les Québécois et les Québécoises de toutes origines peuvent communiquer entre eux et participer au développement de la société québécoise » (MELS, 1998, p. 7). L'école a l'obligation de

remplir cette mission auprès de tous les élèves, quelques soient leurs caractéristiques ou leur provenance (MELS, 1998).

Dans le cas de l'élève issu de l'immigration, il s'agit de l'amener à s'intégrer, c'est-à-dire à lui permettre de s'adapter à sa société d'accueil à long terme, et « distinct de l'assimilation » (MELS, 1998, p. 4). Selon le MELS, l'intégration est réussie quand « la personne immigrante ou ses descendants participent pleinement à l'ensemble de la vie collective de la société d'accueil et ont développé un sentiment d'appartenance à son égard » (MELS, 1998, p. 4). Pour ce faire, l'élève immigrant a besoin d'apprendre et de maîtriser le français pour réussir ses apprentissages scolaires de même que les codes sociaux pour « établir [...] des relations significatives qui transcendent les barrières linguistiques et culturelles et pour participer à la vie collective [...] » (MELS, 1998, p. 4)².

Pour aider les élèves immigrants à commencer l'école dans une langue qui n'est pas la leur, un service d'accueil et de soutien à l'apprentissage du français a été instauré. Pour bénéficier de ce service, l'élève doit être non francophone et inscrit pour la première fois à l'enseignement en français, ne pas connaître suffisamment le français pour suivre les cours dans une classe ordinaire, être inscrit dans une école où toutes les activités se déroulent en français et ne pas participer à un échange d'élèves (MELS, 2006). Nous décrivons ce service et son application dans la section qui suit.

1.5 Cheminement linguistique des élèves allophones dans le système scolaire québécois

Pour intégrer l'élève allophone au système scolaire francophone, deux écueils sont à éviter selon Armand (2005) : celui de l'insertion en classe régulière « sans soutien suffisant et approprié » et celui « de la marginalisation dans un réseau scolaire parallèle qui limite les contacts linguistiques et sociaux avec des locuteurs dont la maîtrise du français est

² Nous rappelons qu'au Québec, selon la Loi 101, l'enseignement se donne en français, mais il y a des exceptions et la législation à ce sujet est contestée par des parents désirant voir évoluer leur enfant dans le système scolaire anglophone.

supérieure » (p. 145). Ainsi, depuis 1997, les commissions scolaires évaluent les élèves immigrants qui arrivent au Québec pour vérifier leur connaissance du français. Ceux qui n'en ont pas une connaissance suffisante ont droit à des services de soutien à l'apprentissage du français (Armand, 2005). Ces services sont d'une durée de dix mois pour le préscolaire, vingt pour le primaire et trente pour le secondaire (Armand, 2005). Les services de soutien consistent au passage en classe d'accueil pour les nouveaux arrivants ou au soutien à l'apprentissage du français pour les élèves intégrés qui manifestent des difficultés en français. En plus des services de soutien, il existe une mesure pour améliorer la connaissance de la langue et de la culture d'origine appelée Programme d'enseignement des langues d'origine (PELO), quand le bassin d'élèves d'un même groupe ethnique et le financement de la commission scolaire le permettent (MELS, 2006).

Selon le MELS (2006), entre 1994 et 2004, de 15,8 % à 18,3 % des élèves immigrants ont reçu des services de soutien en français pendant l'année scolaire. C'est peu, considérant que près de 71 % des élèves nouvellement arrivés ont une autre langue maternelle. De plus, nous n'avons pas relevé de document indiquant qu'il y avait des mesures pour les élèves allophones nés au Québec. Selon Armand (2005), au préscolaire, dans deux commissions scolaires francophones de Montréal, c'est l'intégration dans des classes régulières qui est la pratique la plus répandue, alors qu'une troisième offre des classes d'accueil pour la moitié des élèves. Au primaire et au secondaire, ces commissions scolaires offrent des classes d'accueil pour une durée d'un an. Ensuite, il y a intégration dans la classe régulière si le niveau atteint en français le permet, sinon, il y a poursuite dans la classe d'accueil.

Cette situation peut faire en sorte que des élèves allophones ne développent pas parfaitement leurs habiletés ni en langue première, ni en langue seconde. Nous développons cette affirmation dans la prochaine section.

1.6 Le contexte sociolinguistique d'apprentissage d'une langue seconde

1.6.1 Introduction

Le contexte social, politique et scolaire dans lequel évolue un élève apprenant une langue seconde influe sur le succès de l'apprentissage. Hamers (Hamers, 2004), Cummins (Cummins, 2000), Fazio et Lyster (Fazio et Lister, 1998) de même que Leung et Franson (Leung et Franson, 1989), parmi d'autres chercheurs, ont résumé comment différents contextes peuvent faire en sorte qu'un enfant développe sa compétence en L2 ou, au contraire, qu'il accumule un retard langagier ou cognitif dans une langue ou dans les deux. Nous définissons quelques contextes de bilinguisme pour mieux comprendre comment peuvent évoluer la maîtrise de la langue maternelle et celle de la langue de la majorité chez un enfant allophone.

1.6.2 Les contextes de bilinguisme

Les enfants allophones qui grandissent au Québec sont exposés à au moins deux langues : leur langue maternelle (mais il peut y avoir plusieurs langues parlées à la maison) et la langue d'enseignement, qui est le français, (sans tenir compte de l'anglais, langue majoritaire à l'échelle continentale et sur le plan culturel, pour plusieurs). Ces enfants sont donc appelés à devenir bilingues (ou trilingues ou plurilingues). Être bilingue (ou trilingue, etc.) ne signifie pas maîtriser deux langues (ou trois, etc.) : il y a différents contextes de bilinguisme (Cummins, 2000; Hamers, 2004). Deux contextes de bilinguisme en particulier permettent de mieux comprendre les conditions qui feront que les élèves allophones seront plus susceptibles de maîtriser l'une ou l'autre des langues.

Le premier contexte est celui de bilinguisme simultané (Hamers, 2004). L'enfant est exposé à deux langues pendant les années de développement du langage et ces deux langues sont utilisées dans le processus de socialisation à partir de la naissance. Dans un cas, les deux langues peuvent être apprises simultanément et se développent en remplissant les mêmes fonctions, dont les fonctions liées à la littératie. L'enfant développe alors un bilinguisme équilibré et additif. Cette situation se produit dans des familles mixtes (le père et la mère

n'ont pas la même L1) dont les parents sont instruits, utilisent les deux langues quotidiennement et possèdent des ressources écrites dans les deux langues à la maison. Dans un autre cas, les deux langues sont apprises simultanément, mais sans les fonctions de la langue liées à la littératie. Cette situation se produit dans des familles peu instruites dont les parents n'ont pas la même L1 et dont les fonctions de la langue relative à la littératie ne sont pas valorisées. Lorsqu'un enfant exposé à ce type de bilinguisme commence l'école, il peut être exposé aux mêmes difficultés que celles d'un enfant monolingue de même statut socioéconomique : celles de devoir tout apprendre des fonctions langagières liées à la littératie lors de son entrée dans le monde scolaire. Dans un dernier cas, les deux langues sont apprises simultanément, mais l'une n'est utilisée que pour des fonctions de communication, alors que l'autre l'est pour toutes les fonctions. C'est ce qui se produit quand les parents parlent la même langue, mais que la personne qui garde les enfants en parle une autre. L'enfant pourra apprendre la langue de cette personne, mais pour ses fonctions de communication seulement. Dans les trois cas de figure, il n'y a pas d'inconvénients à devenir bilingue. Il peut même y avoir certains avantages cognitifs : il est alors question de bilinguisme additif.

Un deuxième contexte susceptible de décrire la situation dans laquelle baignent les enfants allophones est un contexte de bilinguisme consécutif. Hamers (2004) décrit deux types de bilinguisme consécutif. Dans un premier cas, il est précoce. Une langue maternelle est d'abord apprise, suivi de l'apprentissage d'une langue seconde, après les années de développement du langage mais pendant l'enfance (avant 8-10 ans). Cette langue seconde est apprise soit dans le voisinage, soit à l'école. Un premier cas d'espèce se produit lorsque la langue première et la langue seconde sont apprises conséutivement, que toutes les fonctions de la langue sont développées en L1 puis, plus tard, en L2 et que les deux langues sont valorisées dans le réseau social de l'enfant. Le produit est un bilinguisme équilibré comme ce qui peut se produire avec le bilinguisme additif.

Dans un deuxième cas, les deux langues sont acquises conséutivement, mais l'enfant n'a pas appris les fonctions de la langue première relatives à la littératie, qui peut être dévalorisée socialement dans la famille, avant de commencer son apprentissage de la langue seconde,

langue privilégiée dans la société. Commencer l'apprentissage d'une langue en même temps que l'apprentissage des fonctions liées à la littératie peut entraîner certains problèmes d'apprentissage de la langue ou des fonctions (Hamers, 2004). C'est ce qui peut se produire dans le cas d'une scolarisation de submersion dans la langue de la majorité :

If the conditions do not permit the child to valorise his L1 sufficiently as compared to his L2, he will not be able to use his L1 for new literacy-related functions when he starts acquiring them, because he can only rely on his limited knowledge of L2. He will try to use an underdeveloped L2 to learn new functions. In this case, acquiring new language functions as well as a new language without the support that comes from the valorization process for the first language, might be too difficult a task for the child³ (Hamers, 2004, p. 91).

Nous nous retrouvons alors devant une situation de bilinguisme soustractif. La tâche étant trop difficile, la langue comme outil cognitif ne sera pas développée. De plus, dans certains cas, quand deux langues sont apprises consécutivement ou simultanément, la langue la moins valorisée peut disparaître ou s'atrophier avant que l'enfant ait développé les fonctions liées à la littératie en cette langue. Alors, non seulement cela pourra nuire au développement cognitif de l'enfant, mais il deviendra monolingue dans la langue dominante.

Cette interaction entre l'apprentissage de la langue (ou de langues) et le développement cognitif a été abondamment explorée dans la littérature (Hamers, 2004 ; Cummins, 2000 et (Collier, 1989) en font la synthèse). Pour Cummins (2000), un développement continu dans les deux langues (bilinguisme additif), en lecture et en écriture, est une condition pour assurer le développement cognitif, linguistique et scolaire. Lorsque des élèves développent faiblement leur littératie en L1 et L2 (par exemple, dans un contexte de submersion), leur habileté à comprendre ce qui se passe en classe et à intervenir sera hypothéquée. De plus,

³ « Si les conditions ne permettent pas à l'enfant de valoriser sa L1 suffisamment par rapport à sa L2, il ne sera pas en mesure d'utiliser de nouvelles fonctions liées à la littératie dans sa L1 quand il commence à les acquérir, car il ne peut compter que sur une connaissance limitée de la L2. Il va essayer d'utiliser une L2 sous-développée pour apprendre les nouvelles fonctions. Dans ce cas, l'acquisition de nouvelles fonctions ainsi que d'une nouvelle langue sans le soutien qui vient du processus de valorisation de la langue première pourraient être des tâches trop difficiles pour l'enfant. »

pour Cummins (2000), un seuil critique de compétence doit être atteint dans au moins une des deux langues pour ne pas nuire à l'apprentissage des langues ou au développement cognitif. Il nomme cette proposition hypothèse du seuil critique. Par ailleurs, Cummins (2000) affirme que, dans la mesure où le seuil critique est dépassé, le développement d'une langue contribue grandement au développement d'une seconde, proposant ainsi une compétence sous-jacente aux deux langues d'une personne bilingue, ce qu'il appelle l'hypothèse de l'interdépendance.

À notre connaissance, il n'y a pas de recherche qui se soit spécifiquement penchée sur le développement langagier des élèves allophones québécois dans leur langue première avant et après la scolarisation, particulièrement en ce qui concerne le développement de la littératie. Il est dès lors difficile de tracer un portrait juste du type de bilinguisme développé par ces élèves. Par contre, comme la plupart de ces élèves sont scolarisés en français, dès leur entrée à l'école, il est probable que plusieurs d'entre eux se trouvent placés dans une situation de bilinguisme consécutif soustractif, sans que la littératie n'ait été développée en langue maternelle.

1.6.3 Conclusion

Nous avons vu que le MELS intégrait les élèves issus de l'immigration dans les classes régulières, après que certains aient bénéficié de mesures de soutien. Selon le MELS (1998), une fois les bases de la communication acquises, il est souhaitable que l'élève soit intégré dans la classe ordinaire, avec son groupe d'âge et dans son établissement d'appartenance. L'élève pourra non seulement continuer à y développer ses compétences dans la langue d'enseignement et de communication, mais il trouvera aussi plus de possibilités de s'intégrer socialement dans un milieu de proximité. Rappelons que ce ne sont pas tous les élèves allophones qui bénéficient de soutien en français, particulièrement les élèves allophones nés au Québec. En fait, il semble que les besoins de soutien ne soient pas les mêmes, selon la langue maternelle ou le pays d'origine de l'élève. Dans ces conditions, comment se comparent le développement langagier et la réussite scolaire des élèves allophones à ceux des élèves dont la langue d'enseignement est la langue première ?

1.7 L'acquisition de la langue d'enseignement par les allophones

Comme nous l'avons souligné précédemment, il n'y a pas, à notre connaissance, d'études qui ont porté spécifiquement sur l'acquisition du vocabulaire scolaire du niveau secondaire auprès des élèves allophones. Par contre, depuis plusieurs années, des chercheurs se sont penchés sur la connaissance de la langue d'enseignement auprès d'élèves allophones, particulièrement auprès d'élèves du primaire. Nous présentons ici de façon très générale les résultats d'études susceptibles d'éclairer notre problématique.⁴

Les résultats de plusieurs études montrent que les élèves allophones n'éprouvent pas de difficulté à développer leurs habiletés de décodage du texte écrit dans une langue qui n'est pas leur langue maternelle (August et Shanahan, 2006; Droop et Verhoeven, 2003; Lefrançois et Armand, 2003; Morris et Labelle, 2008). Ils obtiennent des résultats semblables à ceux des locuteurs natifs. Par contre, des différences entre les deux populations apparaissent en ce qui concerne la connaissance du vocabulaire, les habiletés morphosyntaxiques, la compréhension de texte et la production écrite. Les élèves locuteurs natifs obtiennent de meilleurs résultats dans la plupart des études. En fait, seule l'étude de Morris et Labelle (2008) apporte des résultats qui semblent en partie contredire les autres résultats, particulièrement en ce qui concerne le vocabulaire. Il faut dire que les mots évalués par l'outil de mesure étaient de haute fréquence. Les élèves allophones, comme tous les élèves, sont plus susceptibles d'avoir appris ces mots. Dans les études résumées par August et Shanahan (2006), les habiletés associées à la compréhension au niveau du texte, c'est-à-dire le vocabulaire, la sémantique, la syntaxe, les connaissances contextuelles et textuelles, sont rarement développées chez les élèves allophones au point de rejoindre le développement atteint par leurs pairs monolingues.

Les études montrent que des facteurs comme le développement de la littératie dans la L1, la compétence à l'oral dans la L2 et le statut socioéconomique sont positivement corrélés à

⁴ Les études mesurant la connaissance du vocabulaire sont décrites plus en détail dans le cadre théorique.

l'acquisition de la L2 chez les allophones. Cependant, les études de Aarts et Verhoeven (Aarts, Rian; Verhoeven, Ludo, 1999) et de Morris et Labelle (2008) rappellent que la scolarité des parents et la valorisation de la littératie peuvent favoriser l'acquisition de celle-ci malgré un statut socioéconomique défavorable.

Il faut souligner la rareté des études portant sur les élèves allophones quand ils ont évolué depuis plus de cinq ans dans le système scolaire. Seule l'étude de Hacquebord (Hacquebord, 1994) indique que ces élèves n'ont pas rattrapé les locuteurs natifs dans la connaissance du vocabulaire et l'impact de ce retard sur la lecture n'est pas clair. En général, il a été estimé que les élèves allophones scolarisés dans leur langue maternelle et en langue seconde pouvaient espérer rattraper les élèves locuteurs natifs (Collier, 1989). Par contre, cela devient incertain pour les élèves allophones scolarisés uniquement en langue seconde. Ils peuvent mettre de 7 à 10 ans pour obtenir des résultats dans les domaines d'apprentissages comparables à ceux de leurs pairs issus de la majorité, quand ils y parviennent.

1.8 La réussite scolaire chez les élèves allophones

Si nous n'avons pas une idée précise de la connaissance du vocabulaire scolaire des élèves allophones, nous pouvons nous faire une idée de l'importance de conduire une étude à ce sujet en jetant un regard sur le taux de réussite scolaire dans cette population. August et Shanahan (2006) rappellent que, aux États-Unis, la performance en littératie de nombreux élèves allophones tombe sous la norme de façon croissante au fur et à mesure que les années scolaires passent. En ce qui concerne le taux de décrochage, August et Shanahan rappellent que, en 1999, 31 % de jeunes adultes allophones n'avaient pas complété leurs études secondaires, comparativement à 10 % chez les jeunes adultes qui parlaient anglais à la maison. La maîtrise de la langue semble un facteur important : chez les élèves allophones, 51 % de ceux qui éprouvent des difficultés en anglais n'ont pas complété leur scolarité, comparativement à 18 % de ceux qui maîtrisent l'anglais. Ainsi, « the lower the level of English proficiency, the lower the probability of completing high school » (p. 47).

En ce qui concerne les élèves allophones du Québec, un rapport du MELS (2006) apporte quelques renseignements à ce sujet. Dans ce rapport, il est fait état du retard scolaire chez les élèves. Le retard scolaire est calculé en fonction de l'âge normal pour suivre le parcours scolaire sans avoir redoublé. En général, les élèves issus de l'immigration sont proportionnellement plus nombreux à présenter un retard scolaire. En 1994-95, 27,4 % de ces élèves accusaient un retard, contre 21,6 % de tous les élèves. En 2003-2004, il s'agissait de 19,9 % des élèves issus de l'immigration contre 16,7 % de l'ensemble des élèves. Le retard apparaît particulièrement chez les élèves de première génération ; il est même moins élevé chez les élèves de deuxième génération que pour l'ensemble des élèves. Cependant, ces chiffres ne présentent pas un portrait des élèves allophones. Nous rappelons que, parmi les élèves issus de l'immigration, une part connaît le français. De plus, la façon de calculer les proportions du MELS réduit quelque peu l'importance du retard chez les élèves issus de l'immigration : il faudrait calculer le pourcentage des autres élèves qui accusent un retard sans inclure la portion des élèves issus de l'immigration. À la lumière de ces arguments, il est possible que les données du MELS embellissent la situation véritable des élèves allophones.

Le MELS a analysé ses données en fonction de certains paramètres comme le sexe, la langue maternelle et la région de naissance. Par contre, il est important de souligner que le MELS ne tient pas compte du facteur socioéconomique. En ce qui concerne le sexe, le retard scolaire est plus important pour les garçons que pour les filles, qu'ils soient ou non issus de l'immigration. En 2003-2004, l'écart était le même et ne semblait pas dépendre de la variable immigration. Quand la langue maternelle est prise comme variable, l'analyse des données montre que les élèves dont la langue maternelle appartient aux langues indo-européennes latines, germaniques ou isolées (comme le grec), sémitiques ou tonales d'Asie, le taux de retard est sous ou dans la moyenne. Les élèves dont la langue maternelle appartient aux langues pidgins et créoles, africaines noires, iraniennes et agglutinantes accusent davantage un retard scolaire.⁵ C'est le cas également pour les élèves qui sont nés en Amérique centrale,

⁵ Il s'agit d'une taxonomie des langues, discutable, adoptée par le MELS selon le regroupement de Malherbe (1983 ; dans MELS, 2006).

aux Caraïbes et aux Bermudes, en Afrique orientale et centrale, en Asie occidentale et centrale, en Asie du Sud-Est et méridionale. Au contraire, les élèves nés en Amérique du Nord, en Europe occidentale et septentrionale et en Océanie montrent un retard sous ou dans la moyenne. Par ailleurs, il semble exister un biais, dans la mesure où de nombreux élèves nouvellement arrivés commencent l'école à un âge plus avancé que l'ensemble des élèves.

En plus des élèves accusant un retard scolaire, le MELS rend compte des élèves souffrant d'un trouble d'apprentissage, déclarés EHDAA.⁶ En général, la proportion d'élèves issus de l'immigration déclarés EHDAA est semblable à celle de l'ensemble des élèves. Par contre, les élèves dont la langue maternelle est une langue pidgin, créole ou de l'Afrique noire sont plus nombreux à être déclarés comme tels. C'est l'inverse pour les élèves dont la langue maternelle est une langue tonale d'Asie et indo-européenne slave.⁷

À notre connaissance, les seules données disponibles sur le taux de diplomation des élèves issus de l'immigration proviennent du Bulletin statistique de l'éducation (MEQ, 1998). Selon ce document, les élèves allophones⁸ montrent un rendement scolaire comparable à celui des élèves dont la langue maternelle est le français ou l'anglais. Les résultats aux épreuves uniques du secondaire et les taux de réussite sont semblables. Dans certains domaines, les résultats des élèves allophones sont plus élevés que ceux des autres élèves. Par contre, certains groupes linguistiques éprouvent plus de difficultés « et affichent des résultats nettement inférieurs à la moyenne des élèves » (p. 10). D'après le document, les facteurs qui peuvent expliquer cette situation sont des déficiences pré-migratoires, une faible maîtrise de la langue d'enseignement, un milieu de vie défavorisé, une sous-évaluation des besoins des

⁶ Élève « ayant des difficultés d'apprentissage, des problèmes de comportement, une déficience intellectuelle, une déficience motrice légère ou organique, une déficience langagiére, une déficience physique grave, des troubles marqués du développement ou toute autre déficience, tous degrés de gravité confondus » (MELS, 2006, p. 37). Cependant, les difficultés d'apprentissage et les troubles de la conduite et du comportement ne font plus partie de la définition.

⁷ Nous ne croyons pas que la langue maternelle soit en soi un facteur ; ce sont probablement un ensemble de facteurs socioéconomiques qui interviennent.

⁸ Le Bulletin emploie le terme « allophone », pas « issus de l'immigration », qui a été adopté plus tard.

élèves, l'absence de soutien de la part des parents. Par ailleurs, en ce qui concerne la maîtrise du français, le document affirme que les élèves issus de l'immigration « affichent un bon taux de réussite », indiquant ainsi une maîtrise suffisante de la langue d'enseignement, « du moins sur le plan fonctionnel » (p. 11).

Pour terminer, il faut souligner les difficultés importantes que rencontrent les élèves allophones arrivant à l'école à l'adolescence (MELS, 2006 ; Armand, 2005 ; MEQ, 1998). Ces élèves cumulent rapidement plusieurs années de retard et seul un de ces élèves sur trois obtient son diplôme (MEQ, 1998). Il s'agit d'une problématique liée à une population particulière et ces élèves ne seront pas pris en considération dans le cadre de notre travail.

1.9 Conclusion

Dans cette première partie, nous avons décrit une situation qui, de prime abord, paraît simple : des élèves qui ne connaissent pas le français doivent évoluer dans un système scolaire francophone. Ils doivent à la fois apprendre une langue, le code écrit et les domaines d'apprentissages disciplinaires, alors que les élèves locuteurs natifs du français ont l'avantage de déjà connaître la langue. Les éléments résumés dans les différentes sections de la problématique révèlent au contraire une situation complexe. D'abord, même pour des élèves dont le français est la langue première, le langage scolaire est difficile à acquérir, particulièrement son vocabulaire et sa syntaxe. Les mots à apprendre sont nombreux, souvent abstraits et principalement rencontrés à l'écrit. Cela implique que l'habileté de lecture doit être acquise de même que des habitudes de lecture établies pour apprendre le vocabulaire scolaire. Pour les élèves dont le français est une langue seconde, à ces facteurs s'ajoute leur peu de maîtrise du français. Ils sont sensés éventuellement maîtriser le français pour poursuivre et réussir leur scolarité au même titre que les élèves locuteurs natifs du français.

Cependant, les élèves allophones, rappellent Morris et Labelle (2008, p. 10), « commencent leur vie scolaire avec un retard linguistique important [...] par rapport aux élèves dont la langue maternelle correspond à la langue d'enseignement et qui arrivent à l'école avec d'excellentes représentations phonologiques dans leur L1 [...], des milliers de mots de

vocabulaire [...], une bonne maîtrise orale de morphologie flexionnelle et dérivationnelle [...] et la capacité de produire des phrases d'une certaine complexité syntaxique ». Ce retard linguistique peut être compensé par le développement de la langue maternelle de l'élève allophone favorisant le développement langagier et cognitif propice à l'adaptation à la culture scolaire ou par certains contextes de scolarisation.

Plusieurs éléments soulevés dans les différentes sections suggèrent que des élèves allophones ne parviennent pas à développer suffisamment le langage scolaire pour réussir leurs études. Il a été estimé que l'acquisition de cette langue peut prendre de 7 à 10 ans, quand elle est acquise. La plupart des études réalisées auprès de cette population l'ont été au primaire. Or, pour mesurer l'état de la connaissance du français des élèves allophones, nous avons besoin d'études réalisées auprès d'élèves allophones qui sont dans le système scolaire depuis au moins cinq ans et jusqu'à dix ans. De plus, August et Shanahan (2006) soulignent le manque d'études faites auprès d'élèves du secondaire : « The majority of research on language-minority students has been conducted with children in elementary schools and predominantly in the primary grades. Studies of language-minority students in the middle-school and secondary school years are needed »⁹ (p. 70). À l'extérieur du Québec, seule une étude (Hacquebord, 1994) a porté sur des élèves allophones du secondaire. Au Québec, il n'y a pas, à notre connaissance, de telles études.

Ainsi, il y a un besoin de mesurer les habiletés langagières, notamment le vocabulaire, auprès de ces élèves. En ce qui concerne le vocabulaire, les mots du vocabulaire scolaire sont nombreux, leurs caractéristiques complexes. Les élèves doivent développer leur connaissance de ce vocabulaire pour comprendre les textes et en produire dans les différents domaines d'apprentissages. Est-ce que les élèves allophones y sont parvenus ? Pour le vérifier, nous voulons répondre à cette question de recherche dans le cadre de notre étude :

⁹ « La majorité des recherches sur les élèves des minorités linguistiques ont été menées avec des enfants dans les écoles élémentaires et principalement dans les classes primaires. Des études auprès d'étudiants des minorités linguistiques des niveaux secondaire et collégial sont nécessaires. »

Quel est le degré d'acquisition du vocabulaire scolaire chez les élèves allophones du système scolaire francophone québécois après plusieurs années dans ce système ?

Pour commencer à répondre à cette question, nous proposons dans la prochaine partie un cadre théorique qui nous permettra de préciser les objectifs de recherche et guidera l'interprétation des résultats de notre étude.

CHAPITRE II

CADRE THÉORIQUE

2.1 Introduction

Dans la première partie, nous avons montré l'importance de mesurer le degré d'acquisition du vocabulaire scolaire chez les élèves allophones du système scolaire francophone québécois au secondaire. Comme nous l'avons rappelé, même des élèves dont la langue d'enseignement est la langue première éprouvent des difficultés à élargir leurs connaissances du vocabulaire scolaire. Comme les élèves allophones doivent apprendre la langue d'enseignement en même temps que le contenu des domaines d'apprentissages, nous croyons qu'ils sont plus susceptibles d'accuser du retard comparativement aux élèves francophones. Dans cette partie, nous définissons les principaux concepts abordés dans le cadre de notre étude et précisons le cadre dans lequel nous la situons. Comme le vocabulaire scolaire se retrouve particulièrement à l'écrit, nous montrons d'abord l'importance du vocabulaire en lecture puis nous montrons comment la mesure de la connaissance du vocabulaire peut être un facteur de la performance de lecture. Ensuite, nous résumons les diverses définitions du vocabulaire scolaire et précisons celle que nous adoptons dans le cadre de notre travail. Par la suite, comme le vocabulaire est constitué de mots, nous élaborons ce que c'est que de connaître un mot. Nous montrons que ce ne sont pas toutes les dimensions de la connaissance du vocabulaire que nous voulons mesurer et précisons notre sujet d'étude en définissant deux de ces dimensions : l'étendue de la connaissance et la profondeur de la connaissance. Nous précisons éventuellement quels éléments de la profondeur de la connaissance nous retenons pour notre travail, à la lumière des études sur la connaissance du vocabulaire chez les élèves allophones. Par ailleurs, nous montrons que, parmi plusieurs caractéristiques des mots, le degré d'abstraction est un facteur d'apprentissage qu'il est important de prendre en compte. Ensuite, pour mieux comprendre la population des élèves allophones, nous décrivons

comment le vocabulaire d'une langue est développé chez des personnes pour qui cette langue n'est pas la langue première. Puis nous procérons à une recension des écrits sur l'acquisition du vocabulaire chez les élèves allophones. Nous terminons cette partie en énonçant les objectifs de notre étude.

2.2 Le vocabulaire

2.2.1 Le vocabulaire et la lecture

Avant de définir notre concept principal, le vocabulaire scolaire, nous voulons expliquer son origine et les liens entre le vocabulaire, ce vocabulaire particulier et la lecture. Notre travail s'inscrit dans la lignée des recherches basées sur un corpus et les techniques de profilage par fréquence lexicale (Cobb et Horst, 2001, 2004; Horst et Cobb, 2006; Nation, 2001). Le principe guidant cette démarche de recherche est que, considérant le grand nombre de mots à apprendre (voir plus loin) pour maîtriser une langue, il vaut mieux identifier des mots en fonction de leur fréquence et de leur utilité (connaissance disciplinaire, par exemple) et les enseigner en priorité. Dans notre travail, nous ne voulons pas rechercher des mots à enseigner, mais utiliser la démarche de profilage pour identifier les mots dont des élèves ont besoin dans le cadre de leur scolarité. Comme nous l'expliquons plus loin, les élèves ont avantage à connaître les mots les plus fréquents pour lire, mais également des mots plus spécialisés, particulièrement dans le cadre scolaire.

Dans ce cadre, un facteur important de la réussite est l'habileté de lecture (Chall *et al.*, 1990; Corson, 1997; Schmitt, Jiang et Grabe, 2011). Au fur et à mesure que les élèves avancent dans leur scolarité, ils doivent lire des textes dont les mots sont de moins en moins familiers et fréquents (Chall *et al.*, 1990). Les phrases sont plus longues et plus complexes. Les idées véhiculées par le texte plus difficiles à comprendre. À partir de la quatrième année, un changement important doit se produire: la tâche principale n'est plus d'apprendre le médium (l'écrit) mais le message. Le contenu des textes lus va au-delà de ce qui est déjà su, linguistiquement et cognitivement, par le lecteur, avec des nouveaux mots et de nouvelles idées. Pour comprendre les textes, l'élève doit alors apprendre des milliers de mots (Chall *et*

al., 1990). Cependant, de nombreux apprenants de certains milieux socioculturels n'ont pas accès à ce vocabulaire en dehors de l'école (Corson, 1997). La lecture des textes scolaires est alors une importante - et souvent la plus importante - source d'apprentissage pour plusieurs élèves au-delà de la salle de classe (Schmitt *et al.*, 2011).

L'un des principaux facteurs de la compréhension est la connaissance préalable des mots du texte à lire (Schmitt *et al.*, 2011). En lisant, un lecteur peut inévitablement rencontrer des mots qu'il ne connaît pas. Plus il y a de mots inconnus dans un texte lu, moins la compréhension a des chances de se produire, soit parce que le contenu du texte est oublié pendant que le lecteur cherche la signification des mots inconnus, soit parce que le nombre de mots inconnus ne permet simplement pas le traitement du texte (Schmitt *et al.*, 2011). Plusieurs chercheurs ont donc tenté de mesurer le pourcentage de mots qui doivent être préalablement connus pour lire un texte sans nuire à la compréhension.¹⁰ Le chiffre le plus souvent avancé est de 95 % : un lecteur qui connaît 95 % des mots d'un texte devrait pouvoir comprendre le texte. Ce chiffre a été remis en question et des pourcentages de 98 à 99 % ont même été proposés (Schmitt *et al.*, 2011). Il n'y a pas consensus en ce qui concerne le pourcentage exact. En fait, il semble que plus le vocabulaire d'un texte est connu préalablement, meilleure sera la compréhension (Schmitt *et al.*, 2011). Par contre, le pourcentage minimal peut varier selon le degré de compréhension attendu. Même en connaissant très peu de mots, des lecteurs peuvent retirer de l'information d'un texte lu, comme le thème abordé, par exemple (Schmitt *et al.*, 2011). Cependant, pour espérer obtenir une note de plus de 60 % à un test de compréhension, ce sont près de 98 % des mots du texte qui doivent être connus préalablement (Schmitt *et al.*, 2011).

Combien de mots les élèves doivent-ils connaître pour comprendre les textes lus dans le cadre scolaire ? Certains mots sont-ils plus importants que d'autres ? Les recherches basées sur un corpus ont tenté de répondre à ces questions en identifiant les mots des textes écrits et en les regroupant en familles et par fréquence. Un corpus comme l'*American Heritage Word*

¹⁰ Lire Nation (2001), Cobb et Horst (2001) ou Schmitt *et al.* (2011) pour une recension à ce sujet.

Frequency Book (Carroll, Davies et Richman, 1971, dans Nation, 2001) contient un peu plus de cinq millions de mots tirés de milliers de textes publiés dans l'univers scolaire des États-Unis. Un corpus souvent utilisé actuellement est le *British National Corpus*, comprenant 100 millions de mots tirés de textes écrits (90 %) et oraux de diverses sources pour représenter le mieux possible les mots de l'anglais britannique du 20^{ème} siècle (Site web du BNC). En français, Lété, Sprenger-Charolles et Colé (2004) ont constitué un corpus et établi les fréquences d'occurrences des mots de manuels scolaires des élèves français âgés de 6 à 11 ans. Ils ont constitué deux lexiques : celui des formes orthographiques, avec 48 886 entrées, et celui des lemmes¹¹ avec 23 812 entrées.

Considérant le très grand nombre de mots à apprendre, particulièrement pour des apprenants qui doivent développer leur littératie en langue seconde, des listes plus restreintes ont été établies (Cobb et Horst, 2004). Il s'est agi d'identifier les mots dont les apprenants d'une langue seconde ont besoin à différentes étapes de leur apprentissage de cette langue en fonctions des objectifs scolaires (Horst et Cobb, 2006). Des listes de mots ont été établies en fonction de la fréquence des mots ou de leur spécialité (Nation, 2001). Le but était d'identifier un nombre limité de mots à enseigner à des apprenants pour qu'ils puissent lire les textes dans le cadre de leur formation. Pour identifier ces mots, les chercheurs ont eu recours à du profilage par fréquence lexicale (Cobb et Horst, 2004).

Le profilage par fréquence lexicale consiste à identifier les mots puis les familles de mots au sein de corpus de textes de plusieurs millions de mots en fonction de leur fréquence. Ensuite, les familles de mots les plus fréquentes par tranche de 1000 sont identifiées et leur couverture (le pourcentage de mots qu'ils recouvrent dans un texte donné), calculée. Par exemple, en anglais, les 1000 familles de mots les plus fréquentes couvrent près de 75 % des mots des textes écrits (Cobb et Horst, 2004). Les 2000 familles de mots les plus fréquentes en couvrent près de 80 % (Cobb et Horst, 2004). Ainsi, une personne qui connaît ces 2000 familles de mots connaît préalablement 80 % des mots des textes qu'elle est susceptible de lire. En

¹¹ C'est-à-dire la forme canonique d'un mot, la forme telle qu'elle est entrée dans un dictionnaire.

français, Cobb et Horst (2004) ont estimé que les 2000 familles de mots les plus fréquentes d'un corpus de textes tirés des journaux français *Le Monde* et belge *Le Soir* couvraient environ 83 % des mots de textes informatifs traduits de la revue *Reader's Digest*. Récemment, Le Bras et Lonsdale (Le Bras et Lonsdale, 2009) ont publié un dictionnaire de fréquence basé sur un corpus de 23 millions de mots dans lequel on trouve les 5000 lemmes français les plus fréquents. Nous ne connaissons pas la couverture de ces mots les plus fréquents dans les textes courants.

Nous avons rappelé plus tôt que ce sont 95 à 98 % des mots d'un texte qui doivent être connus préalablement pour que la compréhension du texte ne soit pas hypothéquée par une méconnaissance du vocabulaire. Nous observons que connaître les 2000 mots les plus fréquents ne suffit pas. Apprendre systématiquement de nouvelles tranches de 1000 n'est pas une solution puisque, pour un petit pourcentage de couverture obtenu, il faut apprendre des milliers de mots de plus (Cobb et Horst, 2004). Pour limiter le nombre de mots à apprendre, des chercheurs comme Coxhead (Coxhead, 2000) ont tenté d'identifier les mots spécialisés les plus importants à connaître, c'est-à-dire qui permettaient le mieux de combler le déficit entre les 80 % couverts par les mots les plus fréquents et les 95 à 98 % à connaître. Coxhead a ainsi créé l'*Academic Word List* (AWL), liste constituée de 570 familles de mots les plus fréquentes à travers des textes scolaires, mais moins fréquentes que les 2000 items de la *General Service List* de West (West, 1953). Des listes de vocabulaire comme la liste de Coxhead peuvent couvrir jusqu'à 10 % des mots d'un texte scolaire, en plus des 80 % couverts par les 2000 mots les plus fréquents.

Cependant, la validité de la liste de Coxhead a été remise en question dernièrement (Hyland et Tse, 2007). Cette remise en question a été l'occasion de préciser ce qui était entendu par le concept de vocabulaire scolaire. Nous développons ce sujet dans la prochaine section.

2.2.2 Le vocabulaire scolaire

Quand des apprenants d'une langue ont maîtrisé les premiers milliers de mots les plus utiles à la compréhension, il importe qu'ils apprennent des mots plus spécialisés (Nation, 2001). Tout

dépendant du but des apprenants, ils peuvent apprendre le vocabulaire partagé par différentes disciplines (*Academic vocabulary*) ; plus tard, le vocabulaire d'une discipline en particulier peut être étudié (*Technical Vocabulary*) (Nation, 2001). D'autres auteurs utilisent l'expression *Academic Vocabulary* pour désigner les deux groupes de vocabulaire (Baumann et Graves, 2010), que nous traduisons par vocabulaire scolaire. D'autres expressions ont cours : vocabulaire scolaire général, littératie scolaire, connaissances préalables scolaires, domaines de connaissance, compétence scolaire, connaissances linguistiques, vocabulaire spécifique à un domaine, vocabulaire de contenu, langage scolaire et habileté langagièrre scolaire (Baumann et Graves, 2010).

Nous employons principalement trois expressions dans notre travail : le langage scolaire, le domaine de connaissances ou d'apprentissages disciplinaires et le vocabulaire scolaire. Le langage scolaire est utilisé dans les discussions menées autour des registres linguistiques (Baumann et Graves, 2010). Il s'agit du registre de la langue que les élèves sont supposés adopter dans le traitement des domaines d'apprentissages disciplinaires (Baumann et Graves, 2010). Ceux-ci sont les champs de la connaissance portant sur un sujet général donné, appelés également disciplines. Les domaines d'apprentissages n'ont pas seulement un contenu qui leur est propre, mais également une manière particulière d'organiser ce contenu (Baumann et Graves, 2010). Le corps de la connaissance autour d'un domaine donné est organisé autour de concepts ou principes propres à ce domaine et distingue les domaines les uns des autres. Par exemple, la biologie est organisée en fonction de systèmes alors que l'histoire est organisée le long d'une période de temps ou selon une région géographique. Chaque domaine a son lexique qu'il faut connaître pour développer une compétence dans ce domaine (Baumann et Graves, 2010).

À la lumière de ces considérations, nous retrouvons deux définitions principales du vocabulaire scolaire : le vocabulaire scolaire général et le vocabulaire spécifique à un domaine de connaissances. Le vocabulaire scolaire général est composé des mots qui reviennent dans les textes à travers les différentes disciplines scolaires. Ce sont des mots abstraits dont le sens peut changer d'un contenu disciplinaire à l'autre (par exemple : *forme*, en français et en géométrie ; *processus* en histoire ou en science et technologie) (Baumann et

Graves, 2010). Coxhead (2000), dont nous avons parlé plus tôt, les appelle « mots scolaires » : ce sont les items lexicaux apparaissant fréquemment et uniformément à travers un large éventail de matériel scolaire. Pour constituer sa liste de mots scolaires (*Academic Word List*, présentée dans la section précédente), Coxhead a assemblé un corpus de 3,5 millions de mots à partir de textes de niveau collégial (américain) dans des domaines comme l'histoire, la linguistique, l'économie, le marketing, la loi, la biologie, la chimie et la physique. Elle a ensuite exclu les 2000 mots les plus fréquents de l'anglais de la liste. Ces 2000 mots fréquents proviennent de la *General Service List* de West (1953). Puis elle a conservé les mots qui apparaissaient 100 fois et plus à travers le corpus et dans 15 des domaines ou plus sur les 28 domaines retenus. Elle a ensuite regroupé les mots en une liste de 570 familles de mots. C'est cette liste qu'elle a appelée « liste des mots scolaires » (*Academic Word List*).

L'autre définition du vocabulaire scolaire désigne les mots spécifiques à un domaine de connaissances. Il s'agit de la définition la plus employée (Baumann et Graves, 2010). Le vocabulaire spécifique à un domaine doit être enseigné pour s'assurer que les élèves possèdent les connaissances préalables pour comprendre les contenus rencontrés à l'école (Baumann et Graves, 2010). Il existe quelques synonymes : vocabulaire technique, vocabulaire spécifique au contenu (*content-specific*), termes techniques, langage des domaines scolaires. Les termes et expressions spécifiques à un contenu trouvés dans les manuels scolaires ou d'autres écrits techniques constituent ce vocabulaire (Baumann et Graves, 2010).

En plus des mots couverts par ces deux définitions, d'autres groupements de mots, et même des symboles, qui sont rencontrés et utilisés dans l'un ou l'autre milieu scolaire peuvent également être définis par l'expression « vocabulaire scolaire ». La classification suivante permet de recouvrir la plupart de ces mots et signes.

1. **Le vocabulaire scolaire général.** Ce sont les mots qui apparaissent le plus fréquemment à travers les domaines scolaires sans faire partie des 2000 à 3000 mots les plus fréquents de la langue. Ces mots sont souvent polysémiques, les définitions

variant selon les domaines. Par exemple : *analyze, assume, code, conduct, context, role, period* (Baumann et Graves, 2010). Nous retrouvons sous cette définition la liste *AWL* établie par Coxhead (2000). Hyland et Tse (2007) ont remis en question l'utilité d'une telle liste de mots décontextualisés. D'abord, dans le cas de l'*AWL*, les mots ont été profilés à travers des textes d'un corpus vieux de plus de 30 ans (Hyland et Tse, 2007). Certains mots utiles actuellement ont peut-être été exclus alors que d'autres moins pertinents ont été conservés. Plus préoccupant, un nouveau profilage des mots de la liste avec un corpus de textes récents provenant de plusieurs disciplines a montré que les mots n'avaient pas la même fréquence d'une discipline à l'autre et que certains mots d'un domaine étaient surreprésentés (Hyland et Tse, 2007). De plus, la classe de certains mots et leur sens variaient d'une discipline à l'autre (Hyland et Tse, 2007). Selon Hyland et Tse (2007), l'effort requis pour apprendre ces mots et leurs variations ne permettait de retirer que peu d'avantages.

2. **Le vocabulaire spécifique à un domaine.** Il s'agit des mots et expressions de relativement basse fréquence spécifiques à un contenu apparaissant dans les manuels et autres ouvrages techniques. Par exemple, en mathématiques : *apex, bisect, geometry*; en science : *anticyclone, barometric pressure*; en sciences humaines : *atoll, butte, escarpment, tectonic plate* (Baumann et Graves, 2010). Considérant les problèmes posés par l'enseignement du vocabulaire scolaire général, Hyland et Tse (2007) suggèrent plutôt de mettre l'accent sur l'enseignement du vocabulaire spécifique à un domaine, ce vocabulaire répondant mieux aux besoins des apprenants dans le cadre de leurs études dans ce domaine.
3. **Le vocabulaire de la littérature.** Les mots que les auteurs d'œuvres littéraires utilisent pour décrire les personnages, les situations, les actions, etc. Par exemple : *awkward, chortled, mellow, suavely*. (Baumann et Graves, 2010)
4. **Le vocabulaire du métalangage.** Les termes utilisés pour décrire le langage de la littératie et les consignes, de même que les mots pour décrire les processus, les structures ou les concepts utilisés dans les textes scolaires. Par exemple : *genre, glossary, infer, interrogative, main idea, calculate, compare, model, observe*. (Baumann et Graves, 2010)

5. Les symboles. Icônes, binettes, dessins, notations mathématiques, électroniques, etc. (Baumann et Graves, 2010).

Bien que l'attention de la recherche basée sur les corpus et sur le profilage par fréquence lexicale présentée jusqu'à présent ait principalement insisté sur l'enseignement du vocabulaire, l'approche a également permis le développement de la recherche en langue seconde, notamment nombreux outils de mesure du vocabulaire sont issus de cette démarche (Horst et Cobb, 2006).¹² C'est dans ce cadre que nous situons notre propre étude. Nous voulons vérifier si le vocabulaire scolaire a été appris par les élèves allophones du système scolaire francophone québécois après plusieurs années dans ce système. Considérant les besoins des élèves pour réussir leur scolarité, tous les éléments de la classification qui précède mériteraient d'être évalués. Cependant, nous sommes limité par l'ampleur du travail qu'une telle évaluation demanderait. D'autant plus qu'il n'existe pas, à notre connaissance, de liste d'items couvrant tous les éléments du vocabulaire scolaire au Québec définis dans la classification. Constituer une liste du vocabulaire scolaire général est hors de notre portée dans le cadre de cette étude. Par contre, nous pouvons établir une liste du vocabulaire spécifique aux domaines d'apprentissages, c'est-à-dire les mots du français, de la géographie, de l'histoire, des mathématiques et de science et technologie (voir la partie Méthodologie).

Comme nous l'avons présenté plus tôt, les élèves ont besoin de connaître les mots les plus fréquents de la langue s'ils veulent comprendre les textes qu'ils lisent. En plus des mots du vocabulaire scolaire, nous tenons compte des quelques mots les plus fréquents du français, dans le cadre de notre étude. Normalement, les élèves devraient maîtriser davantage ces mots que les mots du vocabulaire scolaire (Nation, 2001). Tenir compte de la connaissance des mots les plus fréquents nous permettra de mieux comprendre le degré d'acquisition du

¹² Nous présentons plusieurs de ces outils dans la section 2.5.

vocabulaire scolaire en le comparant à celui du vocabulaire des mots fréquents. Cela permettra également d'estimer éventuellement la compétence en lecture des élèves.¹³

2.3 La connaissance du vocabulaire

Dans la section précédente, nous avons montré que les élèves devaient connaître préalablement le plus de mots possibles pour pouvoir comprendre les textes lus au cours de leur scolarité. Dans cette section, nous définissons ce qu'est connaître un mot. Nous précisons ensuite les dimensions de la connaissance du vocabulaire et un facteur important de cette connaissance que nous jugeons pertinents de mesurer considérant la problématique. Nous terminons en présentant divers outils de mesure des dimensions de la connaissance du vocabulaire retenues dans le cadre de notre étude.

2.3.1 Connaitre un mot

Il existe de nombreuses définitions, et depuis longtemps, de ce qu'est connaître un mot.¹⁴ Les définitions retenues par Qian (Qian, 2002), Nation (2001) et Thornbury (2002) résumées ici recouvrent les dimensions importantes de la connaissance d'un mot puis, plus généralement, de la connaissance du vocabulaire.

D'abord, selon Nation (2001) et Thornbury (2002), connaître un mot, c'est avoir en mémoire à long terme les informations suivantes à son sujet :

- sa forme : orale, écrite et dans la forme de ses parties ;
- son sens : la forme et son sens; le concept et les référents; les associations ;

¹³ Nous développons ce sujet dans la section 2.5

¹⁴ Lire Qian (2002) ou Nation (2001) pour un résumé à ce sujet.

- son usage : les fonctions grammaticales, les collocations, les contraintes d'utilisation (registre, fréquence).

Les informations mémorisées sont récupérées différemment selon que le mot est utilisé en compréhension ou en production. Le tableau 2-1 présente la distinction établie par Nation (2001) entre la connaissance d'un mot en compréhension et en production.

Tableau 2-1 Connaitre un mot, en compréhension et en production, selon Nation (2001).

Connaitre un mot comme <i>underdeveloped</i> sur le mode réceptif implique :	Connaitre un mot comme <i>underdeveloped</i> sur le mode productif implique :
<p>de le reconnaître quand il est entendu ; d'être familier avec sa forme écrite pour le reconnaître quand il est lu ; de reconnaître qu'il peut être composé de différentes parties et que ces parties peuvent être liées à son sens ; de savoir qu'il a un sens ; de savoir quel sens a le mot dans un contexte donné ; de connaître le concept derrière le mot, ce qui permet de le comprendre dans plusieurs contextes ; de savoir que des mots sont interreliés (<i>underdeveloped</i> avec <i>overdeveloped</i>, <i>backward</i> et <i>challenged</i>, par exemple) ; de reconnaître si le mot a été bien utilisé dans le contexte ; de savoir que d'autres mots apparaissent régulièrement dans son contexte (<i>territories</i> et <i>areas</i>, par exemple) ; de savoir que ce mot n'est pas un mot fréquent, ni péjoratif en anglais.</p>	<p>de pouvoir le prononcer correctement, y compris les intonations ; de pouvoir l'écrire correctement ; de pouvoir le construire à partir de ses différentes parties dans leur forme appropriée de pouvoir produire le mot qui veut dire <i>underdeveloped</i> ; de pouvoir produire des synonymes et des antonymes ; de pouvoir utiliser le mot correctement dans une phrase originale ; de pouvoir également utiliser les mots qui accompagnent habituellement le mot ; de savoir si le mot est approprié, tout dépendant de la situation, du registre.</p>

Cette définition ne concerne que la connaissance d'un seul mot. Comme le suggèrent différents éléments de la définition de Nation, nous utilisons les mots en contexte, en relation avec d'autres mots. De plus, le traitement de tous ces mots doit pouvoir se faire à grande vitesse (Aitchison, 2005). Compte tenu de ces conditions, Qian (2002) propose quatre dimensions à prendre en considération dans la définition de la connaissance des mots d'une langue :

1. l'étendue du vocabulaire, c'est-à-dire la quantité de mots connus;

2. la profondeur de la connaissance, en particulier les caractéristiques des mots comme la prononciation, l'épellation, les propriétés morphosyntaxiques, sémantiques, de collocation, de phrase, ce qui rejoint la connaissance du mot selon Nation;
3. l'organisation du lexique, c'est-à-dire l'entreposage en mémoire, la connexion et la représentation des mots dans le lexique mental (c'est-à-dire le système organisé des mots dans la mémoire à long terme);
4. le degré d'automaticité de la connaissance en réception et production, c'est-à-dire la vitesse de récupération et de traitement des mots. Récupération et traitement qui comprennent l'encodage et le décodage phonologiques et orthographiques, l'accès aux propriétés structurelles et sémantiques dans le lexique, la représentation et l'intégration lexico-sémantiques, le traitement et la composition morphosyntaxiques.

Dans le cadre de notre travail, ce sont plus précisément les deux premières dimensions qui nous intéressent. Qian (2002) rappelle à quel point l'importance de l'étendue et de la profondeur de la connaissance du vocabulaire en compréhension a été soulignée dans la recherche en langue première. En langue seconde, il a souvent été question de l'étendue (Vermeer, 2001), mais la profondeur de la connaissance est davantage prise en considération dans la recherche des vingt dernières années (voir notre recension des écrits en 2.6 pour quelques résumés à ce sujet). Nous décrivons plus en détail les deux concepts et leurs relations avec la compréhension dans les prochaines sections.

2.3.2 L'étendue du vocabulaire

Nous avons montré que ce sont de 95 à 98 % des mots d'un texte qu'un lecteur devait connaître préalablement pour ne pas compromettre la compréhension. Cela implique de connaître de nombreux mots avant d'entreprendre la lecture de textes courants, sans parler de la lecture des textes plus spécialisés, composés de mots moins fréquents. Normalement, un adulte connaît des dizaines de milliers de mots auxquels il a accès en une fraction de seconde (Aitchison, 2005). Il est impossible de connaître exactement combien de mots une personne connaît, mais il est possible d'estimer le lexique d'une personne dans une certaine mesure. Le

nombre de mots estimé varie selon les études, tout dépendant de la méthodologie utilisée (Aitchison, 2005 ; Nation, 2001 ; Vermeer, 2001). Cela dépend, entre autres s'il s'agit d'une mesure des mots compris ou produits et si l'étude porte sur chaque mot individuel ou sur des mots de la même famille.

En anglais, le nombre de mots comptés dans un très grand dictionnaire s'élève à près de 114 000 familles de mot, sans compter les noms propres (Nation, 2001). Nagy et Anderson ont estimé que le nombre de mots retrouvés dans l'anglais scolaire s'élevait à environ 88 500 familles de mots. Selon différentes sources, le vocabulaire passif d'un adulte est estimé à 20 000 familles de mot (Nation, 2001; Thornbury, 2002) ou jusqu'à 50 000 mots (Aitchison, 2005). Dès l'âge de 6 ans, le vocabulaire passif est estimé à 14 000 mots (Aitchison, 2005).

En français, Lexique, un important outil de référence (lexique.org), propose plus de 128 000 mots distincts dont plus de 54 000 lemmes, sans compter les noms propres, symboles, abréviations et mots d'origines étrangères (Bernard Lété, 2004). L'équipe de Lexique a identifié 48 886 mots (23 812 lemmes) différents à partir d'un corpus de 1,9 million de mots tirés de manuels scolaires français de la première à la cinquième année (Lété, 2004). Ce sont donc des milliers, sinon des dizaines de milliers, de mots qui peuvent être appris.

Il n'est dès lors pas surprenant que la taille du vocabulaire soit un paramètre important mesuré dans la recherche sur la connaissance du vocabulaire (Qian, 2002; Read, 2004; Vermeer, 2001). Nous présentons dans la section 2.5 différents outils de mesure de la taille du vocabulaire. Nous soulignons ici que, pour mesurer la taille du vocabulaire, il a fallu que les chercheurs développent des tests qui permettent de mesurer un grand nombre de mots dans un temps raisonnable. La tâche devait être simple (Read, 2004). Le principal inconvénient de ces tests est qu'ils ne permettent pas de mesurer jusqu'à quel point un mot est connu (Read, 2004). Voilà pourquoi des tests de mesure de la profondeur de la connaissance ont été développés (Read, 2004 ; Qian, 2002). Nous expliquons dans la prochaine section ce que nous entendons par profondeur de la connaissance.

2.3.3 La profondeur de la connaissance du vocabulaire

Comme les réponses obtenues dans les tests sur l'étendue donnaient des indices superficiels de la connaissance d'un mot, il a fallu développer des tests qui mesurent la connaissance du mot plus en profondeur (Read, 2004). Le concept de profondeur de la connaissance donc permet de distinguer les tests de vocabulaire plus conventionnels et traditionnels (choix multiple, dictée à trous, dictée, reconnaissance d'erreur, par exemple, selon (Wesche et Paribakht, 1996) des tests conçus plus récemment pour mesurer précisément la profondeur de la connaissance (Read, 2004). Différentes définitions de la profondeur peuvent être regroupées en trois grandes conceptions (Read, 2004). Nous les résumons successivement puis précisons les éléments que nous retenons pour notre étude.

1. À la suite du constat qu'il peut être difficile de donner une définition exacte des mots, la profondeur de la connaissance désigne la variation dans la précision de la connaissance du sens d'un mot, du vague au très précis (Read, 2004). Le *Vocabulary Knowledge Scale (VKS)* de Wesche et Paribakht (1996) est un exemple de test qui mesure la profondeur de la connaissance en ce sens (voir section 2.5, plus loin). Par ailleurs, la précision de la connaissance peut être affectée par des caractéristiques des mots (Read, 2004). Par exemple, la définition des mots de haute fréquence est souvent vague et imprécise, particulièrement hors contexte : *chose, faire, beau, ici, quelqu'un*. D'autres mots sont polysémiques, le sens prenant différentes teintes, avec une portée d'usage variable; le sens peut même varier largement : *forme, puce, bizarre, briser, propre, piger*. Les mots se distinguent également selon un usage de tous les jours ou un usage plus technique : *fruit, langue, réaction, argument*. Les mots qui se prêtent le mieux à une définition précise sont les mots techniques, comme *phonème, immunodéficience, fabrication*. Cette variabilité dans les caractéristiques sémantiques des mots est perçue par les locuteurs adultes natifs d'une langue (Read, 2004). Par exemple, pour des mots comme *carburateur, mollusque, logarithme* et *molécule*, les locuteurs natifs peuvent identifier le contexte ou le champ d'utilisation sans nécessairement connaître le sens particulier de ces mots. La connaissance de mots spécialisés et de basse fréquence peut refléter d'ailleurs

les intérêts, la profession et le niveau d'éducation de même que l'origine socioculturelle d'une personne (Read, 2004).

2. Une autre conception de la profondeur de la connaissance rejoint la définition de la connaissance d'un mot de Nation (2001) que nous avons présentée plus tôt. Elle désigne la connaissance d'un mot qui inclut non seulement ses propriétés sémantiques, mais orthographiques, phonologiques, morphologiques, grammaticales, de collocation et pragmatiques (Nation, 2001; Ordóñez, Carlo, Snow et McLaughlin, 2002; Qian, 2002). Cette conception complique le développement d'outils de mesure, s'il faut prendre en considération toutes les dimensions (Read, 2004). Read rappelle l'étude de cas longitudinale de Schmitt (1998 ; dans Read, 2004) dans laquelle une entrevue prenait deux heures pour faire dire à un élève ce qu'il connaissait au sujet de cinq aspects pour onze mots. Même à l'écrit, une telle procédure est longue et difficile. En plus des difficultés pratiques, il faut s'interroger sur la généralisabilité de la procédure, c'est-à-dire sur la valeur théorique de nombreuses données pour un si petit nombre de mots.
3. La profondeur de la connaissance peut être conçue comme la densité d'un réseau, c'est-à-dire qu'un mot appris est intégré dans un réseau lexical (Read, 2004; Schoonen et Verhallen, 2008; Vermeer, 2001). Contrairement aux deux autres approches, qui s'intéressent aux mots pris individuellement, cette approche s'intéresse au développement de liens entre des groupes de mots dans le lexique mental (Read, 2004). Le mot est lié à d'autres mots auxquels il peut être associé et desquels il peut être distingué. Comme l'étendue du vocabulaire d'un apprenant augmente, les nouveaux mots doivent s'intégrer dans un nouveau réseau ou un réseau existant (Read, 2004). La profondeur de la connaissance est vue comme l'habileté à distinguer des mots liés sémantiquement et, plus généralement, la connaissance des différentes façons de lier des mots individuels. Cette vision est fondée sur la distinction établie entre les relations paradigmatisques et syntagmatiques retrouvée en linguistique structuraliste, dans les théories sémantiques de Lyons (1995 ; dans Read, 2004) et de Cruse (1986 ; dans Read, 2004).

Des trois conceptions de la profondeur du vocabulaire, c'est la troisième qui nous semble la plus adéquate pour permettre la réalisation de notre étude. D'abord, comme nous procérons à une cueillette de données auprès d'élèves d'une école secondaire, les contraintes de temps sont importantes. Les tests développés dans le cadre de la troisième conception permettent de procéder rapidement à leur passation auprès de nombreux élèves (voir la section 2.5). Ensuite, la problématique et notre recension des écrits (voir la section 2.6) suggèrent que les élèves allophones ne sont pas susceptibles d'éprouver des difficultés avec toutes les dimensions de la connaissance du mot. Ces élèves ne semblent pas particulièrement avoir de difficultés avec la forme des mots, alors que des retards à la fin du primaire sont observés sur les plans sémantique et syntaxique. Les connaissances sémantiques et syntaxiques peuvent en partie être mesurées dans le cadre de la troisième conception, à travers les connaissances paradigmatisques et syntagmatiques de même qu'en prenant en compte le degré d'abstraction des mots. Nous développons ce sujet dans les prochaines sections.

2.3.4 Le vocabulaire vu comme un réseau

Nous avons vu que des dizaines de milliers de mots sont stockés dans la mémoire adulte. Les recherches montrent qu'ils sont organisés et marqués pour être récupérés rapidement (Aitchison, 2005). Dans cette perspective, développer son vocabulaire ne consiste pas seulement à mémoriser de nombreux mots. Inévitablement, des liens se créent entre les mots. Plus il y a de liens, plus le mot est connu en profondeur (Aitchison, 2005; Nation, 2001; Qian, 2002; Verhellen et Shoonen, 1998). Par ailleurs, de nouveaux sens peuvent être attribués à un mot connu ou un sens peut s'enrichir (Aitchison, 2005; Verhellen et Shoonen, 1998). De plus, différentes structures de relations peuvent être construites entre les mots (Singleton, 1999).

Ce sont d'importants réseaux qui sont donc développés entre les mots avec le temps et l'expérience. Les connexions lexicales mentales sont loin de ressembler à l'organisation lexicale proposée par un dictionnaire ou un lexique. Il s'agit d'un réseau complexe et dense, au sein duquel des liens sont solides et durables, alors que d'autres peuvent être créés rapidement au besoin (Aitchison, 2005). Il s'agit d'une structure fluide, quoique certains liens

durent longtemps. De nouveaux liens doivent pouvoir être créés rapidement et les plus vieux doivent pouvoir être modifiés (Aitchison, 2005). Dans cette structure, les mots sont souvent présentés comme les nœuds d'un réseau (Vermeer, 2001). Ces nœuds peuvent être connectés sous différents angles : selon le thème (*table-chaise-fauteuil*), la prononciation (*câble-fable-sable*), la morphologie (*similitude-gratitude-altitude*), la conception (*tasse-verre-chope*) ou le registre (*crisser-déposer*, comme dans « *crisse-le là* » ; « *dépose-le là* »). Plus un réseau est dense autour d'un mot, plus grande est la connaissance de ce mot (Vermeer, 2001).

Comment la densification se produit-elle ? S'appuyant sur les travaux de Harris (1992 ; dans Singleton, 1999) et Anglin (1970 ; dans Singleton, 1999), Singleton (1999) rappelle que l'enfant procède à une révision, à une réorganisation et à une consolidation de ses connaissances après l'explosion lexicale de la petite enfance, peu avant de commencer l'école. Cela se poursuivra tout au long de la scolarité. Alors que de nouveaux mots sont appris chaque jour, le lexique est réorganisé de deux façons. D'abord, les items du lexique sont organisés dans des domaines de mots reliés entre eux. Ensuite, l'enfant développe des relations plus complexes entre les mots et les domaines de mots et entre les concepts. Cette réorganisation comprend trois dimensions : 1. Un déplacement des relations syntagmatiques vers les relations paradigmatisques entre les mots. 2. L'établissement de relations de plus en plus abstraites entre les mots, dont nous discutons dans la section 2.3.7). 3. La classification lexicale et sémantique du monde en termes d'hyperonymie et d'hyponymie (Singleton, 1999).

2.3.5 Les connaissances paradigmatisques

Le déplacement des relations syntagmatiques vers les relations paradigmatisques se produit entre l'âge de cinq et dix ans (Singleton, 1999). Ce déplacement est démontré par les résultats obtenus dans des tâches d'association de mots (Meara, 2009). Les enfants les plus jeunes proposent des mots dont la catégorie grammaticale est différente de celle du mot cible (relation syntagmatique : *table* et *manger*, par exemple), alors que les plus vieux proposent plus systématiquement un mot de la même catégorie syntaxique (relation paradigmatique : *table* et *chaise*, par exemple). Les résultats obtenus ainsi ont suggéré que le développement

du lexique comprend une perception croissante des relations syntaxiques, sémantiques et conceptuelles entre les mots et l'association des mots à leur classe syntaxique (Singleton, 1999).

C'est particulièrement pendant les années scolaires que se développent les relations paradigmatisques entre les mots, l'école favorisant un tel traitement (Schoonen et Verhallen, 2008). Les relations paradigmatisques sont des relations hiérarchiques verticales entre les mots, comme l'hyperonymie et l'hyponymie (Vermeer, 2001). À l'école, les élèves développent leur lexique en étant confrontés à de nouveaux sens et à de nouvelles relations entre les mots. Ces nouveaux sens et relations sont développés par généralisation, catégorisation et abstraction (Schoonen et Verhallen, 2008). Par exemple, un enfant connaît d'abord le mot *oiseau* à travers des expériences et perceptions concrètes à la maison, à la télévision ou dans un livre. Par la suite, le sens du mot est étendu et des liens avec d'autres mots sont établis. Le sens du mot est éventuellement décontextualisé, c'est-à-dire détaché de l'expérience de l'enfant avec le référent, au fur et à mesure qu'il apprend des caractéristiques de ce qu'est un *oiseau* : il a des plumes, mais il n'est pas nécessairement petit. Il en fait une catégorie plus générale : les oiseaux construisent des nids, pondent des œufs. Toutes sortes d'aspects sont ajoutés au concept OISEAU. Éventuellement, en plus d'avoir établi des relations syntagmatiques avec le concept OISEAU (OISEAU-VOLE, OISEAU-NID, OISEAU-ŒUF), il y a établissement de relations paradigmatisques (OISEAU-ANIMAL, OISEAU-CANARI).

Ainsi, et particulièrement dans le monde scolaire, la connaissance des mots s'organise logiquement, dans des relations d'antonymie, de synonymie, de gradation, de coordination, d'hyperonymie et d'hyponymie (Schoonen, Verhallen, 2008 ; Qian, 2002 ; Vermeer, 2001). Les relations paradigmatisques sont hiérarchiques et caractérisées par l'inclusion dans une classe taxonomique. Par ailleurs, le travail de Miller et Fellbaum (1991 ; dans Read, 2004) montre que les relations paradigmatisques varient en fonction de la catégorie grammaticale d'un mot. Les relations taxonomiques (hiérarchiques) entre les hyperonymes et les hyponymes sont caractéristiques des noms communs, alors que les relations d'antonymie et de gradation sont des principes de classification des adjectifs. En ce qui concerne les verbes,

s'il est possible d'établir des relations hiérarchiques avec la plupart comme avec les noms, les premiers forment tout de même une classe sémantique plus complexe. Peu importe le type, établir des relations paradigmatisques entre les mots aide l'élève à comprendre leur sens dans les textes scolaires, le système scolaire favorisant ce type de relation entre les mots (Schoonen et Verhellen, 2008).

2.3.6 Les connaissances syntagmatiques

Nous avons déjà dit que les mots n'apparaissaient habituellement pas seuls et que la connaissance d'un mot impliquait la connaissance du ou des contextes dans lequel apparaissait le mot. Pour se représenter le sens d'un mot avec les autres mots rencontrés dans un contexte donné, le locuteur doit faire appel à ses connaissances syntaxiques et sémantiques (Gleitman, Cassidy, Nappa, Papafragou et Trueswell, 2005). Il peut également faire appel à sa connaissance des blocs de mots apparaissant fréquemment dans une même séquence, c'est-à-dire les collocations (Metin et Karaoğlan, 2011; Nation, 2001; Ward, 2007; Wray, 2002). Cette connaissance des blocs de mots nous l'appelons connaissances syntagmatiques, qu'elles soient le traitement syntaxique de chacun des mots du syntagme ou tout un syntagme mémorisé comme unité lexicale (Nation, 2001; Wray, 2002). Wray (2002) distingue ces deux façons de traiter l'information linguistique comme étant analytique et holistique. Le traitement analytique implique l'interaction des mots et des morphèmes avec les règles syntaxiques pour créer ou décoder une information linguistique originale. Le traitement holistique signifie la récupération d'un bloc de mots mémorisés.

Le traitement analytique demande des connaissances syntaxiques préalables. Ces connaissances syntaxiques contribuent ensuite à la compréhension d'un mot dans son contexte (Gleitman *et al.*, 2005). Ainsi, les apprenants d'une langue développent des représentations des formes lexicales par entrelaçage syntaxique (« *syntactic bootstrapping* »). Le processus d'apprentissage du vocabulaire commence par un paireage mot-monde, dans lequel les apprenants développent d'abord un stock d'items lexicaux (des noms surtout) porteurs de concepts de base et concrets. À partir de ce stock de mots faciles à apprendre, les apprenants développent le lexique en se fondant sur des structures linguistiques de plus en

plus complexes, mais suivant une séquence d'acquisition dans le temps (Gleitman *et al.*, 2005). Ainsi, pour qu'un apprenant apprenne un nouveau verbe, il faut qu'il connaisse au préalable un certains nombres de noms et il faut qu'il ait construit les représentations linguistiques qui révèleront la structure d'arguments du verbe exprimé par son interlocuteur dans un contexte donné (Gleitman *et al.*, 2005).

En ce sens, l'information syntaxique, même si elle semble limitée, favorise l'apprentissage des mots : l'acquisition d'items plus abstraits demande à l'apprenant de porter attention à la distribution des items déjà connus à l'intérieur de structures syntaxiques (Gleitman *et al.*, 2005). Le contexte syntaxique devient alors une source d'information pour construire le sens des items. Éventuellement, une correspondance structure-réalité est établie, en plus de la correspondance mot-réalité. Par exemple, la structure particulière dans laquelle apparaît un verbe révèle la structure argumentative qu'il encode. Si on observe une image dans laquelle un chien offre un ballon à un éléphant et que l'on entend « Le chien donne le ballon à l'éléphant » ou « L'éléphant reçoit le ballon du chien », cela permet d'établir la structure argumentative des verbes donner ou recevoir. Ainsi, la position des éléments dans une phrase énoncée dans un contexte perceptible établit leur rôle comme sujet ou objet grammatical et permet de donner un sens au verbe. Les enfants d'au moins trois ans font ces inférences syntaxiques (Gleitman *et al.*, 2005).

Avec l'expérience, le locuteur d'une langue mémorise les blocs de mots les plus fréquents plutôt que de les traiter d'une manière procédurale chaque fois qu'il les rencontre (Nation, 2001). Les locuteurs d'une langue les utilisent et les traitent à une plus grande vitesse comme unité que s'ils avaient à les reconstruire (Nation, 2001). De plus, certaines erreurs faites par des locuteurs montrent que des collocations sont utilisées comme des unités lexicalisées (Nation, 2001). Ces blocs, ou collocations, sont récurrents dans une langue (Metin et Karaoglan, 2011 ; Nation, 2001 ; Wray, 2002). Leur emploi par un locuteur est souvent un critère pour l'identifier comme locuteur natif (Nation, 2001). Les collocations varient en taille (nombre de mots d'une séquence), en type (mots outils en collocation avec des mots de contenu, mots de contenu avec d'autres mots de contenu), dans la distance entre les mots en collocation et dans la portée des collocations possibles pour un mot (par exemple, établir un

pont, son camp, une liaison) (Nation, 2001). Les corpus de textes énormes que nous avons désormais à notre disposition permettent de mieux les répertorier (Metin et Karaoglan, 2011 ; Nation, 2001).

Les collocations sont une des relations parmi d'autres associées à une compréhension et à une production appropriée du vocabulaire (Nation, 2001). Un aspect de la connaissance du mot est d'avoir une représentation des contextes dans lesquels une forme donnée peut être utilisée pour exprimer du sens. Cette connaissance contextuelle peut impliquer la situation générale, le thème ou un contexte particulier. Les autres mots d'une collocation sont en grande partie de l'information fournie dans un contexte particulier pour un mot donné, qui précisent le sens de ce mot ou qui forment une unité de sens plus grande (Nation, 2001). La connaissance du thème est la connaissance d'informations relatives au thème abordé par un texte. Par exemple, la compréhension précise du mot *balle* dans un texte dépend du thème dont il est question dans le texte : la chasse, le golf, un animal de compagnie ou un jour à la ferme (Nation, 2001). Éventuellement, l'usage de certaines collocations les figent en des expressions lexicalisées au sens plus ou moins transparent (eau douce) ou opaque (bateau mouche). Nous ne nous intéressons pas à ce phénomène dans le cadre du présent travail.

Tenir compte des collocations comme connaissances syntagmatiques dans la mesure de la profondeur du vocabulaire est particulièrement pertinent dans le cas du vocabulaire scolaire tel que défini dans notre travail. Les mots des disciplines apparaissent fréquemment dans des syntagmes complexes avec d'autres noms ou adjectifs spécialisés également (Ward, 2007). Ces collocations peuvent être plus spécialisées et spécifiques à un domaine d'apprentissages que le mot seul (par exemple, gaz -> compression d'un gaz ; gaz naturel) (Ward, 2007). Les élèves doivent connaître autant les mots pris individuellement que le sens global des collocations s'ils veulent comprendre un texte.

2.3.7 Un facteur à considérer : le degré d'abstraction d'un mot

Plusieurs caractéristiques des mots sont des facteurs d'apprentissage : leur fréquence, leur longueur, les collocations, les registre et contrainte fonctionnelle, le comportement

syntaxique, les formes de base et dérivation, les propriétés idiosyncrasiques, la polysémie, l'homonymie, le degré d'abstraction, etc. (Ferrand, 2001; Nation, 2001; Qian, 2002). Dans le cadre qui nous préoccupe, le degré d'abstraction est un facteur qu'il vaut la peine de prendre en considération. D'abord, un grand nombre de mots du vocabulaire des domaines d'apprentissages sont abstraits (Ward, 2007). Les disciplines emploient la nominalisation pour désigner en un mot des phénomènes ou entités complexes (les noms *lexicalisation* ou *variable*, par exemple) (Ward, 2007). Ensuite, il semble que les mots abstraits ne soient pas appris ni traités aussi facilement et rapidement que les mots plus concrets (Altarriba, Bauer et Benvenuto, 1999). Des connaissances préalables sont requises, dont des connaissances syntaxiques (Gleitman *et al.*, 2005). Les élèves allophones qui sont moins susceptibles d'avoir développé ces connaissances, particulièrement les connaissances syntaxiques de la langue d'enseignement (August et Shanahan, 2006; Morris et Labelle, 2008), risquent d'éprouver des problèmes dans l'apprentissage des mots plus abstraits. En langue seconde, plusieurs auteurs soulignent la distinction entre l'apprentissage des mots abstraits et celle des mots concrets (ou *imageables*) dans le développement du lexique du locuteur bilingue (De Groot et Keijzer, 2000; Pavlenko, 2009) (voir la section 2.4). Voilà pourquoi nous croyons que le degré d'abstraction est un facteur à considérer dans la mesure de la profondeur de la connaissance du vocabulaire. Jusqu'à présent, nous avons vu que la représentation des mots pouvaient se construire, entre autres, en établissant des relations syntagmatiques et paradigmatisques entre eux. Dans la littérature, les exemples utilisés pour illustrer ces relations sont souvent des mots perceptibles. Il s'agit de mots concrets et « *imageables* » (Ferrand, 2001). Par contre, considérant que les élèves doivent apprendre des mots abstraits, nous expliquons ici comment l'humain parvient à se représenter des mots avec lesquels il n'a jamais eu d'expérience ou qui ne sont pas perceptibles.

Les mots selon qu'ils sont plus concrets ou plus abstraits ne sont pas appris dans le même ordre (Altarriba *et al.*, 1999; Singleton, 1999). L'humain procède à un traitement des mots de plus en plus abstrait, ce qui est révélé dans des études où les résultats montrent que les enfants plus âgés et les adultes traitent les mots de façon plus abstraite que les plus jeunes (Altarriba *et al.*, 1999 ; Singleton, 1999). Ils savent généraliser, contrairement aux jeunes

enfants qui surgénéralisent. Ils savent mieux distinguer les objets et percevoir ce qui peut les rassembler ou les différencier. Ils ont développé une capacité de discrimination ou de différenciation : ils savent comment faire abstraction de certaines différences et identifier une caractéristique commune partagée.

Les plus jeunes enfants vont également généraliser, mais sans capacité à discriminer (Singleton, 1999). Par exemple, les enfants pourront étiqueter CHIEN des chiens et des chevaux et ne sauront étiqueter QUADRUPÈDE ces différents animaux. Les adultes sauront regrouper sous QUADRUPÈDE les chiens, chats, chevaux, etc. Par ailleurs, le développement vers des conceptions plus abstraites est lié au développement des relations hiérarchiques entre les mots. Un jeune enfant saura regrouper sous FLEUR des mots comme *rose*, *tulipe* ou sous ARBRE des mots comme *chêne* ou *érable*. Ce n'est que plus tard qu'il saura regrouper ces objets et ces étiquettes sous PLANTE. De plus, la métaphore et l'emploi de sens plus abstrait sont davantage utilisés chez les plus vieux. Un enfant de trois ans ne saura pas accoler un adjectif comme *brillant* ou *dur* à une personne, alors qu'à partir de l'âge de 12 ans, plusieurs savent désormais assembler les mots avec plus de flexibilité (Singleton, 1999). Il n'y a pas que l'âge qui soit un facteur d'apprentissage des mots, selon qu'ils sont concrets ou abstraits. Il semble que les mots abstraits soient plus difficiles à apprendre que les mots concrets. Par exemple, un mot comme *savoir* est plus difficile à apprendre que *chien* ou *sauter*. Le mot *savoir* dépend de connaissances générales et syntaxiques préalables pour être appris (Gleitman *et al.*, 2005).

Comme le suggère ce qui a été vu précédemment, les mots abstraits sont appris sur des bases différentes de celles des mots concrets. Les mots concrets sont appris davantage avec l'expérience du monde, alors que les mots abstraits le sont avec l'expérience de la langue. Des chercheurs ont proposé que les représentations des mots abstraits en mémoire soient qualitativement différentes de celles des mots concrets (Crutch, Connell et Warrington, 2009). Des travaux récents en neuropsychologie auprès de personnes aphasiques et dyslexiques profondes ont montré que c'était le cas. Les mots abstraits sont principalement représentés par des associations avec d'autres mots alors que les mots concrets le sont par leur similarité taxonomique entre eux. Crutch *et al.* (2009) rappellent que les écrits, fictifs ou

pas, sont construits à partir d'une grande part de mots abstraits. En apparence, les mots abstraits et concrets partagent certaines similitudes, par exemple un effet de fréquence dans une tâche de rappel. Cependant, à fréquence égale, les mots concrets sont systématiquement rappelés plus rapidement que les mots abstraits. Leurs propriétés sémantiques, leur longueur, leur nombre de syllabes, leur structure morphologique et phonologique sont différents.

Pour expliquer la différence dans les temps de récupération des mots selon leur degré d'abstraction, il a été proposé que les mots concrets pouvaient être représentés dans différents systèmes de la mémoire, contrairement aux mots plus abstraits (Altarriba *et al.*, 1999; De Groot et Keijzer, 2000; Holcomb, Kounios, Anderson et Caroline, 1999). Il s'agit de la théorie du double encodage (Paivio, 1991, dans Holcomb *et al.*, 1999). Selon cette théorie, il existe deux systèmes de représentations sémantiques interreliés : le système de représentations verbales et le système de représentations imagées (« *imaginal* »). Le système de représentations verbales traite l'information linguistique alors que le système de représentations imagées traite l'information non linguistique. Les représentations ne sont pas disponibles de la même manière en mémoire selon le degré d'abstraction des mots. Les mots concrets et abstraits sont représentés dans le système de représentations verbales, mais seuls les mots concrets sont connectés au système de représentations imagées. Il est plus facile de se rappeler des mots concrets que des mots qui ne sont pas dans le système de représentations imagées, car le double accès semble favoriser une récupération plus efficace (De Groot et Keijzer, 2000; Ferrand, 2001).

Le degré d'abstraction est donc un facteur qu'il est important de prendre en compte dans le cadre de notre travail. Cependant, d'autres facteurs peuvent intervenir dans le traitement des mots comme leur longueur et leur fréquence (De Groot et Keijzer, 2000; Ferrand, 2001; Zahar, Cobb et Spada, 2001). Nous devons alors également prendre ces deux autres facteurs en considération. Un facteur qui peut intervenir avec le degré d'abstraction et que nous tentons de contrôler dans notre étude est la polysémie (Hyland et Tse, 2007) (voir la conception de la mesure de la profondeur de la connaissance dans la partie Méthodologie). Des mots très fréquents sont à la fois abstraits et polysémiques (par exemple, *chose*) (Hyland et Tse, 2007). Certains mots des domaines d'apprentissages sont plutôt spécialisés et

conservent un sens précis dans un domaine donné (par exemple, *adjectif* en français, *cercle* en mathématiques) mais d'autres mots sont souvent polysémiques à travers les domaines (par exemple, *processus*, en science et technologie, en géographie et en histoire) (Hyland et Tse, 2007). Selon nous, le phénomène de la polysémie est suffisamment complexe et particulier pour devoir être étudié à part. Nous expliquons dans la partie méthodologie comment nous avons exclu le facteur polysémie dans le choix et l'organisation des items de notre mesure de la profondeur de la connaissance.

2.3.8 Conclusion

Nous avons vu dans cette section qu'un être humain peut apprendre des milliers mots. Ces mots ne sont pas mémorisés comme une simple liste. Au contraire, ils sont organisés, ce qui facilite leur récupération en mémoire. Des liens syntagmatiques et paradigmatisques sont établis entre eux. De plus, bien que de nombreux mots soient appris et traités par association entre une forme et un objet perceptible, les mots peuvent être appris grâce à la maîtrise de structures syntaxiques ou à l'aide d'autres connaissances stockées en mémoire. Le sens des mots du langage scolaire dont il est question dans ce travail repose particulièrement sur une maîtrise de la syntaxe et de connaissances préalables : ces mots sont plus abstraits et plus complexes morphologiquement. De plus, ils ne sont pas susceptibles d'avoir été traités hors du contexte scolaire. Dans un tel contexte, comment des élèves pour qui le langage scolaire n'est pas la langue première développent-ils le vocabulaire de ce langage ? Le développement suit-il le même cheminement que celui des élèves pour qui la langue d'enseignement est la langue première ? Pour nous aider à répondre à ces questions, nous décrivons dans la prochaine section comment se construit le vocabulaire d'un apprenant d'une langue seconde.

2.4 Le développement du lexique en langue seconde (L2)

Pour mieux comprendre les différences et ressemblances éventuelles dans la connaissance du vocabulaire scolaire entre des élèves locuteurs natifs et des élèves allophones, nous devons comprendre comment le vocabulaire se développe quand la langue cible n'est pas une langue première. Dans cette section, nous exposons les différents modèles de développement du

vocabulaire chez l'apprenant d'une langue seconde. L'expérience montre qu'une personne peut connaître des mots dans deux langues, ou plus, sans que cela ne semble poser de problème particulier. Les questions qui se posent dans le cadre de notre travail concernent le développement d'un lexique dans une langue seconde par des locuteurs qui n'ont peut-être pas toutes les connaissances linguistiques pour le faire avec succès. Pour répondre à ces questions, nous devons au préalable décrire comment les mots de langues différentes sont traités et représentés en mémoire et comment leurs représentations, et lesquelles, sont éventuellement reliées entre elles.

2.4.1 Un ou des lexiques ?

Les mots d'une langue seconde sont-ils traités comme le sont les mots d'une langue première ? Des chercheurs ont proposé que le traitement des mots dans le lexique mental de la L2 était différent de leur traitement dans le lexique de la langue première (L1) (Singleton, 1999). Le premier reposerait surtout sur des opérations phonologiques alors que le second reposeraient plutôt sur des opérations sémantiques. Cependant, les recherches récentes suggèrent plutôt une gradation dans les différents traitements des mots du lexique en L2. Plus de ressources cognitives sur le plan de la forme seraient dédiées au traitement des mots nouveaux d'une L2 (comme au début de l'apprentissage de la L1, d'ailleurs). Ainsi, les mots difficiles à prononcer sont appris plus lentement que ceux qui ne posent pas de difficulté (Read, 2000). Le traitement du sens des mots prend plus d'importance au fur et à mesure que la forme des mots est maîtrisée : les apprenants débutants stockent le vocabulaire en fonction de la forme sonore des mots alors que les plus avancés les mémorisent en fonction du sens (Read, 2000 ; Singleton, 1999).

De nombreuses recherches suggèrent que les lexiques en langue seconde et en langue première d'un même locuteur sont connectés entre eux, mais ils ne forment pas un seul et même lexique (Singleton, 1999)(Kroll et Sunderman, 2003; Singleton, 1999). De plus, il semble que les relations entre un mot donné en L2 et un autre en L1 vont varier d'un individu à l'autre, tout dépendant du contexte dans lequel le mot est traité, comment il a été appris, jusqu'à quel point il est connu et dans quelles mesures les liens sémantiques sont communs

d'une langue à l'autre (Singleton, 1999). Certains modèles ont proposé que les représentations des mots dans le lexique étaient distinctes d'une langue à l'autre (Singleton, 1999). Ces modèles tendent à affirmer que l'accès aux mots est sélectif, c'est-à-dire qu'il est possible d'activer en mémoire dans une seule langue la forme des mots rencontrés (Singleton, 1999). Des modèles plus récents ont proposé que les représentations des mots d'une langue à l'autre étaient intégrées. Selon ces modèles, l'accès lexical est non sélectif, c'est-à-dire que la forme des mots dans les différentes langues est activée lorsqu'un locuteur les rencontre (Singleton, 1999). Nous décrivons ces modèles dans la section 2.4.4.

2.4.2 L'accès au mot en mémoire

Actuellement, il semble y avoir consensus en ce qui concerne l'accès aux mots chez la personne bilingue : cet accès peut être non sélectif, autant pour les formes graphémiques que pour les formes phonémiques (Meuter, 2009). En d'autres mots, lorsqu'un locuteur voit ou entend un mot dans une langue donnée, les représentations dans les diverses langues du locuteur relatives à ce mot peuvent être activées. Le modèle le plus cité est le modèle bilingue d'activation interactive (*Bilingual Interactive Activation* ou BIA, Dijkstra et Van Heuven, 1998, dans Meuter, 2009; voir description, plus loin). Il s'agit d'un modèle de la reconnaissance des mots qui propose d'importantes connexions entre les représentations des mots à l'intérieur d'une langue et entre les langues, dont des fonctions d'inhibition et d'excitation des liens bidirectionnelles entre les traits, les lettres, les mots et les niveaux de langue. Les mots candidats sont activés peu importe la langue, caractéristique d'un lexique non sélectif. Par contre, des contraintes dues au contexte peuvent limiter l'activation de candidats dans une des langues ou inhiber entièrement l'une des langues. D'autres facteurs comme le degré de compétence dans la L2, des chevauchements entre les langues (à travers les mots apparentés, par exemple) peuvent également contraindre la sélection lexicale.

2.4.3 Les représentations d'une langue à l'autre

Si les formes phonologiques et morphosyntaxiques diffèrent d'une langue à l'autre dans la plupart des modèles de représentation mentale, il semble que les représentations de sens et de

concepts sont partagées par les langues, en partie ou totalement (Pavlenko, 2009). Cette conclusion est justifiée par le fait que les bilingues peuvent traduire la plupart des mots d'une langue à l'autre et par les résultats dans des tâches de décision lexicale avec amorce et celles d'identification d'image (Altarriba et Basnight-Brown, 2009; Pavlenko, 2009). En fait, plus récemment, plutôt que de s'intéresser aux liens entre la forme des mots et les concepts, les chercheurs se sont intéressés à la structure des catégories linguistiques, c'est-à-dire les représentations mentales liées à la forme des mots et aux notions grammaticales. La question clé n'est plus « Est-ce que l'entreposage des concepts est séparé ou partagé? » mais « Qu'est-ce qui est partagé et qu'est-ce qui est séparé pour des concepts lexicaux particuliers ? » (Pavlenko, 2009).

Dans cette perspective, les concepts lexicaux sont définis comme des représentations multimodales incluant des informations visuelles, auditives, perceptuelles (texture) et kinesthésiques (sensori-motrice) stockées en mémoire (Pavlenko, 2009). Ces représentations sont dynamiques, sujettes aux changements, au développement et aux différences générationnelles et individuelles. Dans ce dernier cas, les différences sont attribuées à des expériences personnelles variées vis-à-vis du concept cible. Comme les concepts sont activés dans un contexte donné, cela permet d'activer différentes représentations pour un même mot, selon ce contexte.

Il arrive que des concepts d'une L1 n'aient pas d'équivalent dans la L2 et vice versa. En fait, des différences conceptuelles peuvent exister d'une langue à l'autre dans la façon de catégoriser les concepts, dans les frontières qui sont établies entre eux et dans la prototypicalité qui leur est attribuée (Pavlenko, 2009). Cela fait en sorte que des traductions ne sont pas toujours équivalentes sur le plan conceptuel : certains mots de différentes langues sont partiellement équivalents alors que d'autres ne trouvent pas d'équivalent dans l'autre langue. À l'opposé, il y a équivalence (ou quasi-équivalence) quand les concepts de deux langues partagent une même structure et une même frontière dans leurs catégories linguistiques. Par exemple, dans le domaine des contenants, les noms *tas* (en néerlandais) et *tasse* (en français) sont catégorisés de la même façon dans les deux langues (Pavlenko, 2009).

En plus des concepts qui peuvent ne pas trouver d'équivalence, la façon de catégoriser des concepts semblables peut varier d'une langue à l'autre. Parfois, deux catégories distinctes dans une langue sont soumises à une seule catégorie plus large dans une autre langue. Par exemple, toujours dans le domaine des contenants, la catégorie *jar* en anglais pourrait faire partie de la catégorie *frasco*, en espagnol; mais cette catégorie pourrait également englober d'autres objets comme *bottle* et *container* (Pavlenko, 2009). Certaines catégories d'une langue peuvent ne pas exister dans une autre langue. Par exemple, les apprenants russes de l'anglais auront à construire l'idée des *rowhouses* ou des *rowhomes* de l'anglais nord-américain (Pavlenko, 2009), ces constructions urbaines n'existant pas en Russie (Pavlenko, 2009).

Apprendre des mots comme *rowhouse* ou des catégories comme *contenant* dans une L2 demeure relativement facile : leur référent est perceptible, concret (Pavlenko, 2009). Par contre, des mots abstraits comme *intimité* ou des mots désignant une émotion comme *frustration* risquent d'être plus difficiles à apprendre parce que les apprenants doivent identifier correctement la portée des situations et des contextes dans lesquels ces mots peuvent être employés (Pavlenko, 2009). Ce qui pourra d'abord être appris au sujet d'un mot abstrait sera une définition explicite, puisque le locuteur n'a pas accès à une représentation conceptuelle multimodale du mot. Ensuite, il pourra tenter de catégoriser le mot et le lier à des objets ou à des situations. Il parviendra éventuellement à utiliser spontanément le mot dans des contextes valables, comme le font les locuteurs natifs de la langue cible (Pavlenko, 2009).

2.4.4 Quatre modèles de lexique du locuteur bilingue

À la lumière de ce qui précède, plusieurs chercheurs ont tenté de modéliser le lexique du locuteur bilingue. Nous présentons ici les quatre principaux modèles souvent cités dans les recensions des écrits.

2.4.4.1 Le modèle d'interaction active bilingue (*Bilingual Interactive Activation* ou BIA)

Le modèle d'activation interactive bilingue plus (BIA) (Dijkstra, Van Heuven et Grainger, 1998, dans (Dijkstra et van Heuven, 2002) est un modèle de la reconnaissance visuelle des mots qui prend en compte le rôle de la phonologie pendant la lecture bilingue. Ce modèle a pour origine le modèle d'activation interactive (*Interactive Activation model*) de reconnaissance visuelle des mots chez les monolingues proposé par McClelland et Rumelhart (1981, dans Dijkstra et van Heuven, 2002). Le modèle BIA comprend quatre niveaux de représentation : un niveau des traits, de la lettre, du mot et de la langue (appelé *supralexical language node*). L'activation dans ce modèle a lieu de façon bidirectionnelle, à l'intérieur d'un niveau comme entre les niveaux. Le modèle est interactif dans la mesure où les nœuds de haut niveau peuvent envoyer des informations dans les nœuds de bas niveau et vice versa, comme l'illustre l'Image 2-1.

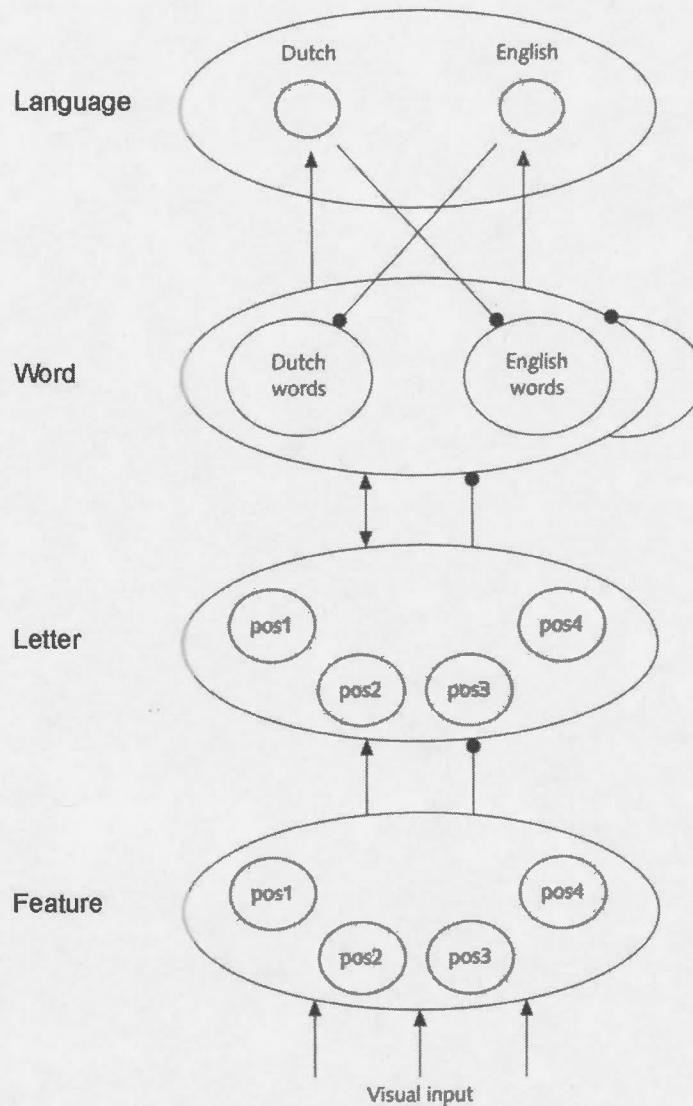


Image 2-1 - Le modèle d'activation interactive bilingue plus (BIA) de Dijkstra, Van Heuven et Grainger, 1998 (tiré de Dijkstra et van Heuven, 2002, p. 177)

Dans le modèle, des informations sensorielles à partir de stimulus orthographiques activent les traits graphiques représentés en mémoire. Ces traits éveillent les représentations de lettres qui à leur tour éveillent les représentations lexicales dans les deux langues. Ces représentations lexicales activent les nœuds langagiers (supralexicaux). Un nœud langagier

activé envoie des rétroactions à tous les nœuds lexicaux dans cette langue et ces nœuds excitent à leur tour les représentations des lettres de la langue. L'activation d'un nœud langagier particulier est influencée par la reconnaissance lexicale précédente. Un mot reconnu précédemment en L1 augmente la sensibilité à l'activation dans ce nœud langagier (les liens avec une flèche) et décroît la sensibilité à l'activation dans le nœud de la L2 (les liens avec un rond plein). C'est à travers ces processus de rétroactions que les personnes bilingues réduisent les interférences d'une langue pendant le traitement d'une autre. Le modèle suggère que, à la suite de la phase d'activation sans sélection, les informations langagières contextuelles permettent la suppression de représentations lexicales incompatibles avec le contexte. Le modèle rend ainsi compte des occasions d'interférences interlangagières et du fait que les locuteurs bilingues sont capables de tenir ces interférences au minimum.

2.4.4.2 Le modèle hiérarchique modifié

Dans le modèle hiérarchique modifié (Kroll et Stewart, 1994), le contenu conceptuel est totalement ou partiellement partagé entre les langues ou spécifiques à l'une ou l'autre langue. Dans ce modèle, les mots sont stockés séparément selon leur forme dans chacune des langues. Les formes sont liées à différents systèmes de catégorisation et de conceptualisation. Ces systèmes identifient les catégories qui sont partagées d'une langue à l'autre mais chaque langue entraîne également le développement d'un système de catégories qui lui est propre. L'Image 2-2 illustre le modèle hiérarchique modifié.

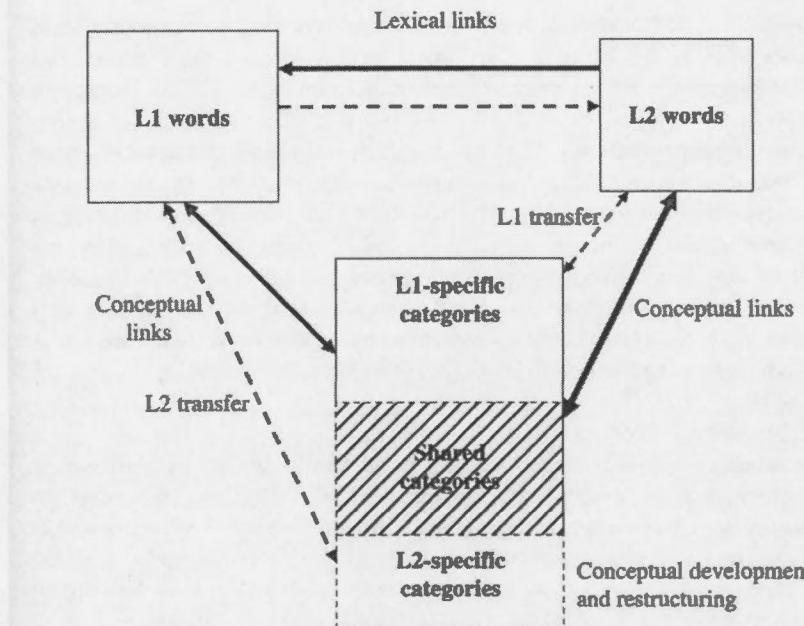


Image 2-2 Le modèle hiérarchique modifié de Kroll et Stewart, 1994) (tiré de Pavlenko, 2009, p. 147)

La différence de liens entre les mots de L2 et les mots de L1 dans le modèle illustre l’existence d’une asymétrie dans le développement du lexique du locuteur bilingue. Celui-ci, particulièrement au début de son apprentissage, va transférer ses connaissances relatives à un mot de la L1 à son correspondant dans la L2 qu’il a nouvellement rencontré (Kroll et Sunderman, 2003). Le locuteur débutant traduit ainsi plus facilement et systématiquement les mots qu’il a rencontrés en L2 vers la L1. Il s’agit ici des mots concrets, peu d’études ayant porté sur les mots abstraits. Au fur et à mesure que la compétence en L2 s’acroît, des liens directs sont établis entre les mots de la L2 et le système de représentations sémantiques, sans passer par une traduction des mots en L1 (Kroll et Sunderman, 2003). Cependant, à l’exception des bilingues les plus compétents, les liens entre les mots et les concepts seront toujours plus forts dans la L1 que dans la L2 (Kroll et Sunderman, 2003).

2.4.4.3 Le modèle à traits distribués (*Distributed feature model*)

Le modèle à traits distribués (Van Hell et De Groot, 1998, dans Pavlenko, 2009) s'inspire d'un modèle monolingue selon lequel les concepts se ressemblent ou se distinguent selon le nombre de traits sémantiques qu'ils partagent (Kroll et Sunderman, 2003). L'Image 2-3 présente ce modèle.

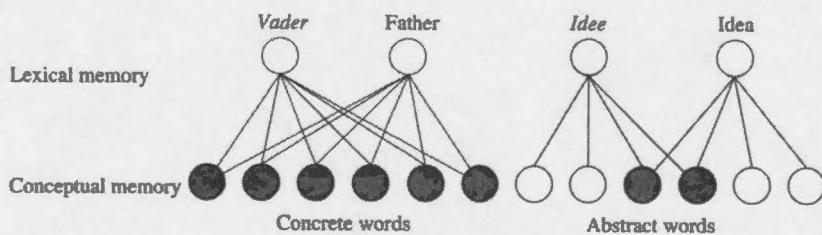


Image 2-3 Le modèle à traits distribués de Van Hell et De Groot, 1998 (tiré de Pavlenko, 2009, p. 144)

Van Hell et De Groot ont établi une distinction entre les noms concrets et abstraits de même qu'entre les mots qui sont apparentés et ceux qui ne le sont pas (*cognates, noncognates*) (Meuter, 2009). Dans le modèle, les noms concrets (et les mots apparentés, comme *Vader* et *Father*, illustrés par le modèle) possèdent des sens plus similaires entre les langues. À l'inverse, les représentations des mots abstraits (comme les mots *Idee* et *Idea*) peuvent être très différentes puisque le sens de ces mots est largement déterminé par le contexte sociolinguistique dans lequel ils se sont développés (Meuter, 2009). Un modèle alternatif à celui de Van Hell et De Groot propose que les mots apparentés, comme ils partagent une même morphologie, sont agglutinés dans le lexique, peu importe la langue. Les mots qui ne sont pas apparentés sont distribués dans leur lexique respectif.

2.4.4.4 Le modèle de Jiang

Jiang (Jiang, 2000) propose un modèle du développement du vocabulaire d'une langue seconde en situation formelle. L'auteur distingue trois aspects interreliés dans ce développement : les représentations, l'acquisition et le traitement. Le système de représentations des mots est d'abord développé dans la langue première. Une entrée lexicale

chez un locuteur natif l'amène à construire une représentation sémantique, syntaxique, morphologique et une représentation de sa forme (phonologique et orthographique) (Image 2-4). Les représentations sémantiques et syntaxiques sont portées par le lemme de l'entrée lexicale alors que l'information morphologique (les différentes formes du mot) et les informations relatives à ses formes sont catégorisées sous le lexème.

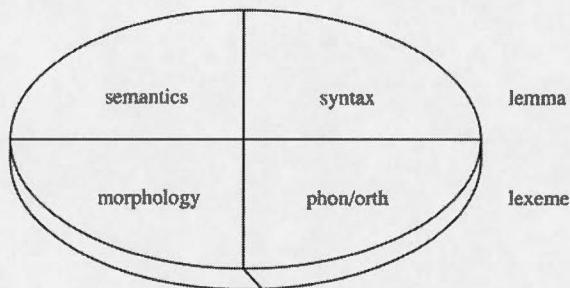


Image 2-4 La structure des représentations d'une entrée lexicale, en L1 (Jiang, 2000, p. 48)

De plus, les diverses représentations d'un mot de la L1 sont associées à sa représentation dans le système conceptuel. Le défi pour un apprenant d'une L2 est de développer un système de représentations du mot propre à la L2 et de l'associer à un système conceptuel qui a d'abord été développé avec la L1.

Selon Jiang, ce développement de la connaissance d'un mot dans une langue seconde se déroule en trois étapes. Au cours de la première étape (Image 2-5), l'apprenant consacre surtout ses efforts à une représentation phonologique et orthographique du mot dans la L2. Le lemme du mot est peu développé dans la L2 et le traitement se produit sous la conduite de la conception du mot en L1 (par le biais de la traduction, par exemple), comme l'illustre l'Image 2-5.

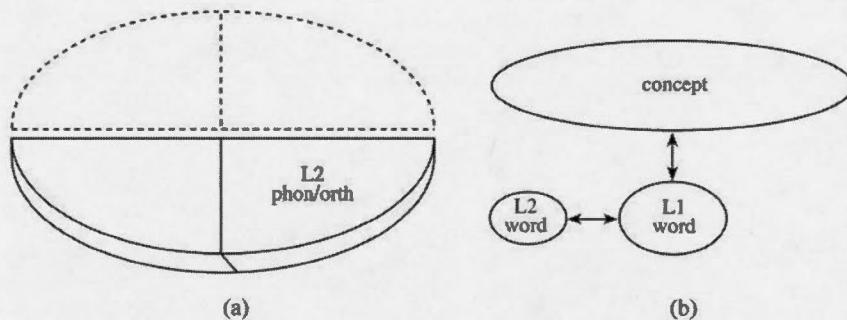


Image 2-5 La représentation du mot et son traitement au début de l'apprentissage d'une L2 (Jiang, 2000, p. 51)

Au cours de la deuxième étape (Image 2-6), le nombre d'associations entre le mot de la L2 et sa traduction dans la L1 augmente, au point d'être représenté dans les dimensions sémantiques et syntaxiques du mot de la L1. Par contre, les représentations morphologiques du mot en L1 ne sont pas associées au mot en L2, celles-ci étant spécifiques à la langue et peu susceptibles d'être transférées.

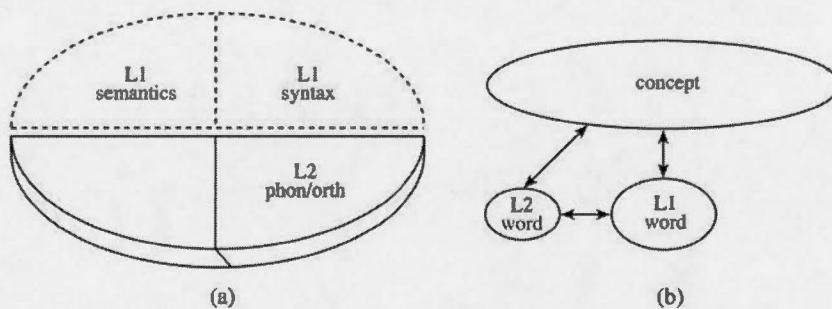


Image 2-6 La représentation du mot et son traitement à la deuxième étape de son apprentissage en L2 (Jiang, 2000, p. 53)

L'auteur donne en exemple le mot anglais *table* avec sa traduction *zhuozi* en mandarin. Les deux mots réfèrent à l'objet table et sont des noms, mais *table* prendra un -s au pluriel alors que la forme *zhuozi* ne change pas du singulier au pluriel. L'information morphologique en

L1 est donc peu susceptible d'aider à l'utilisation de la forme dans la L2 (à l'exception de quelques cas ; voir références dans Jiang, 2000, p. 52).

Une autre caractéristique importante de cette étape de développement est la faiblesse de la relation entre le mot de la L2 et le système de représentations conceptuelles. Cela s'explique par le fait que les représentations du lemme du mot sont copiées en grande partie à partir de ses représentations en L1. Il se peut également que des informations aient été perdues dans le processus de traduction.

Éventuellement, à la troisième étape, le mot est entièrement représenté par ses caractéristiques dans la L2 (Image 2-7). Il s'agit de l'étape d'intégration, où les représentations et le traitement du mot en L2 ressemblent à ceux d'un mot en L1.

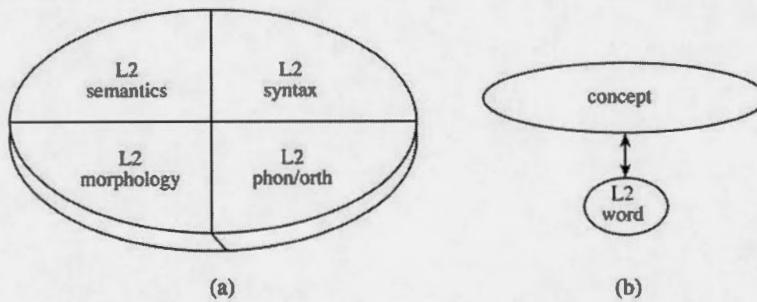


Image 2-7 L'étape d'intégration d'un mot de la L2, dont les représentations ressemblent à celles d'un mot dans la L1 (Jiang, 2000, p. 53)

2.4.5 Conclusion

Les notions théoriques et les modèles du lexique mental bilingue décrits dans cette section sont fondés sur des données recueillies auprès de locuteurs maîtrisant déjà leur langue première. Cependant, les caractéristiques linguistiques des élèves allophones peuvent différer de celles des populations bilingues étudiées par les chercheurs cités plus haut. Nous avons vu que ce ne sont pas tous les élèves allophones qui ont développé des habiletés de littératie dans leur langue première (voir la section 1.6). Plusieurs ne connaissent de leur L1 que ce qui est utilisé à la maison, à l'oral. D'autres suivent des cours dans leur langue d'origine (PELO),

mais il s'agit d'une minorité. Comment, alors, se développe le lexique d'une langue seconde si celui de la langue première n'est pas ou est peu développé?

Parce qu'ils font dépendre les premiers apprentissages des mots d'une L2 par la traduction dans les mots de la L1, les modèles de Kroll et Stewart et de Jiang ne permettent pas de répondre à cette question. Il semble que la mémoire conceptuelle soit impliquée dans l'apprentissage des mots d'une L2 dès les premiers moments de l'acquisition de celle-ci (De Groot et Keijzer, 2000). Le modèle à traits distribués de Van Helle et De Groot rend mieux compte du phénomène. De plus, nous avons déjà expliqué que les mots abstraits ne sont pas stockés et traités comme le sont les mots concrets (voir section 2.3.7). Le modèle à traits distribués de Van Helle et De Groot rend également compte de cette distinction entre mots concrets et mots abstraits. Il pourra ainsi nous aider à interpréter nos résultats en ce qui concerne le degré d'abstraction des mots.

2.5 Mesurer la connaissance du vocabulaire

Il existe de nombreux outils de mesure pour évaluer les différents aspects de la connaissance du vocabulaire (Nation, 2001; Read, 2000). Les études présentées dans la section 2.5 et d'autres encore (Read, 2000; Wesche et Paribakht, 1996) montrent également une grande diversité de modalités de passation, de l'entrevue semi-dirigée au test de lecture à voix haute en passant par la production de réponses écrites. Cependant, à notre connaissance, il n'existe pas de test qui permette de mesurer le vocabulaire scolaire en français, que ce soit en étendue ou dans la profondeur de la connaissance. Il nous faudra développer un tel outil de mesure en nous inspirant des outils développés ailleurs. Dans le cadre de notre travail, nous voulons développer un outil qui s'adresse à une population d'élèves du secondaire, qui puisse être administré dans les limites imposées par le cadre scolaire (contrainte de temps, passation en groupe sans possibilité de tricher, etc.). Nous cherchons des outils qui permettent d'évaluer la quantité de mots connus par les élèves et de vérifier la qualité des liens établis entre les mots dans le lexique mental de ces élèves (c'est-à-dire la profondeur de la connaissance du vocabulaire). Dans cette section, nous présentons différents outils de mesure, d'abord de

l'étendue du vocabulaire, puis de sa profondeur. Nous décrivons les outils retenus, en présentant les différents arguments qui nous ont aidé à arrêter notre choix.

2.5.1 Mesurer l'étendue

Pour mesurer l'étendue du vocabulaire, il est possible d'utiliser des tâches à choix multiple, d'association avec un synonyme ou une définition, de traduction ou d'items à cocher pour indiquer si on connaît le mot ou pas (Read, 2000). Nous présentons successivement quelques-unes de ces tâches, telles qu'elles sont administrées dans des études ou des tests d'admission ou de classement à l'université. Pour la plupart de ces tâches, on estime l'étendue du vocabulaire en projetant la proportion du nombre de mots connus dans la tâche sur le nombre total de mots de l'échantillon (Read, 2000 ; Vermeer, 2001). Par exemple, on calcule que quelqu'un qui connaît 10 mots sur 12 des mots qui font partie des 1000 mots les plus fréquents en connaît $10/12 * 1000 = 833$.

Le premier outil présenté ici est le *Voluntary Levels Test* de Nation (dans Read, 2000). Il permet de mesurer l'étendue du vocabulaire sur cinq niveaux de fréquence des mots en anglais. Les premiers 2000, la tranche des 3000, celle des 5000, les mots universitaires (plus de 5000) et les 10000 et plus. Selon Nation, les 3000 mots les plus fréquents sont les mots de haute fréquence que tous les apprenants devraient connaître. Ce ne sont pas tous les mots qui sont présentés au répondant. Pour chaque niveau, il y a 36 mots et 18 définitions. Les mots sont présentés en groupes de 6, comme suit :

1. apply	choose by voting: _____ become like water: _____ make: _____
2. elect	
3. jump	
4. manufacture	
5. melt	
6. threaten	

Tous les mots d'un groupe appartiennent à une même classe de mot. Les mots apparaissent sans contexte, mais le test est rapide à passer. Il présente une forme qui peut être adaptée au vocabulaire scolaire. Il existe d'autres variantes avec des phrases à compléter par le mot cible (dont les deux ou trois premières lettres sont fournies). Dans ce dernier cas, il s'agit alors

davantage d'une mesure du vocabulaire en production que du vocabulaire en réception. Pour évaluer l'étendue du vocabulaire, il faut ensuite estimer le nombre de mots connus pour chaque niveau de fréquence en calculant la proportion de bonnes réponses sur le nombre total de mots pour chaque niveau. Ces dernières années, Nation a mis à jour la liste des fréquences en appuyant les analyses sur le *British National Corpus* dont nous avons parlé au début de cette partie (Nation et Beglar, 2007).

Un autre outil pour mesurer l'étendue du vocabulaire est l'*Eurocentres Vocabulary Size Test* développé par Meara (Meara et Buxton, 1987; Meara, 1992). Le but est de mesurer la taille du vocabulaire à partir d'échantillons de mots de différentes fourchettes de fréquence, à raison de 20 mots par niveau de fréquence de 1000. Le répondant doit dire s'il connaît ou non chaque mot d'une liste. Il s'agit d'un test administré par ordinateur. Pour s'assurer que les répondants ne répondent pas à tort et à travers, le test présente des pseudomots dans la liste. Il y a un pseudomot pour deux mots. Ce test donne la possibilité de répondre rapidement à un très grand nombre de mots. Le résultat est pondéré par les erreurs, c'est-à-dire lorsqu'un pseudomot est choisi comme un mot connu. Les résultats obtenus avec ce genre de test montrent qu'ils permettent d'estimer l'étendue du vocabulaire auprès de la plupart des apprenants de langue seconde, à part quelques exceptions (Meara, 1992). De plus, c'est un test qui est facile et rapide à développer.

Actuellement, la dernière version du test de Meara s'appelle X_Lex et est disponible en ligne.¹⁵ Le test mesure la connaissance des 5000 mots les plus fréquents de la langue. Les mots sont présentés un à un et le participant décide s'il connaît un mot ou pas en cliquant sur le bouton approprié. Les items sont des mots et des pseudomots. Deux scores sont obtenus : un résultat brut et un résultat qui tient compte des pseudomots déclarés connus. Ce dernier score donne une idée du nombre de mots connus sur 5000. Il existe également une version pour mesurer la fréquence des 5000 à 10 000 mots les plus fréquents de la langue.

¹⁵ www.logistics.co.uk

Une variante de ce test est utilisée dans la batterie de tests développés par Alderson (Alderson, 2005) appelée DIALANG. DIALANG est un outil sur Internet pour diagnostiquer la compétence langagière dans l'une ou l'autre de 14 langues européennes.¹⁶ Il existe trois versions de l'outil pour chaque langue, correspondant à trois niveaux de difficultés. Pour déterminer à quelle version sera confronté un participant, celui-ci passe le *Vocabulary Size Placement Test* (VSPT). Il s'agit d'un test à réponses Oui-Non, comme celui de Meara. Le test comprend 75 items, dont 25 pseudomots. Le participant coche « Oui » ou « Non » devant chaque mot en répondant à la question : « S'agit-il d'un mot de cette langue? ». La taille du vocabulaire est habituellement mesurée en extrapolant le résultat d'un test (Read, 2000 ; Vermeer, 2001).

2.5.2 Mesurer la profondeur de la connaissance

Plusieurs outils ont été développés pour mesurer cette connaissance (Read, 2000). Nous présentons les principales mesures développées au cours des dernières années, dont les outils pour mesurer la profondeur de la connaissance en termes de réseau.

Un premier outil dont il est souvent question dans la mesure de la profondeur de la connaissance du vocabulaire est le *Vocabulary Knowledge Scale* de Wesche et Paribakht (1996). Son but est de mesurer le vocabulaire appris de façon fortuite dans le cadre d'une étude ou d'une formation. Des mots sont présentés au répondant qui doit choisir pour chaque mot la proposition qui décrit le mieux sa connaissance du mot. Le répondant doit compléter la proposition, au besoin. Voici comment se présente le choix de propositions :

¹⁶ www.lancs.ac.uk/researchenterprise/dialang/about.htm

I - Je ne me souviens pas avoir vu ce mot avant.
 II - J'ai déjà vu ce mot, mais je ne sais pas ce qu'il veut dire.
 III - J'ai déjà vu ce mot et je crois qu'il veut dire : _____ (synonyme ou traduction).
 IV - Je connais ce mot. Il veut dire _____ (synonyme ou traduction).
 V - Je peux utiliser ce mot dans une phrase : _____
 (Écrire une phrase avec le mot. Si tu réponds à cette section, réponds aussi à la section IV.)

Le test est long à administrer et ne permet pas de mesurer une grande quantité de mots. De plus, le répondant doit produire des réponses. Dans le cadre de notre travail, ce test n'est pas approprié puisqu'il faut produire des mots et non seulement les comprendre. Les deux habiletés sont en partie distinctes (Nation, 2001) et les tests en compréhension mieux réussis que ceux en production (De Groot et Keijzer, 2000). Comme nous situons notre travail dans un cadre plus large de compréhension des textes scolaires, nous préférons nous en tenir à un test qui ne mesure que la compréhension du vocabulaire de ces textes.

S'inspirant des tâches d'association de mots (un mot est présenté et le participant doit écrire tous les mots auxquels ce mot lui fait penser), Read (2000) a développé le *Word-associates Test* (Wat). Comme les apprenants d'une langue seconde ont souvent un vocabulaire limité, Read a développé une tâche d'association de mots en reconnaissance seulement. Un mot cible est présenté et il faut sélectionner parmi un choix de mots les mots auxquels il peut être associé. Par exemple :

Edit

arithmetic	<u>film</u>	pole	publishing	revise	risk	surface	text
------------	-------------	------	------------	--------	------	---------	------

(Les réponses sont soulignées.)

Dans cette première version du test, trois types d'association pouvaient être établies :

1. Paradigmatique. Les deux mots sont des synonymes ou partagent un sens, l'un étant plus général que l'autre. Par exemple :

edit -> revise
abstract -> summary
assent -> agreement
adjust -> modify

2. Syntagmatique. Les deux mots apparaissent souvent dans un même contexte. Ce sont des collocations.

edit -> film
team -> sport
abstract -> concept
occur -> phenomenon

3. Analytique. Le mot associé représente un aspect ou une composante du mot cible et fait probablement partie de la définition dans le dictionnaire.

team -> together
edit -> publishing
electron -> tiny
export -> oversea

Des problèmes de développement se sont posés. Selon Read (2000), il n'est pas toujours possible d'être constant dans la sélection des mots associés, d'un mot cible à l'autre. Les verbes et les adjectifs entretiennent des relations différentes avec les autres mots que les noms, particulièrement dans le cas des collocations. De plus, certains adjectifs ont plus de portée que d'autres : *physical* comparativement à *military*, par exemple. Par ailleurs, que faire quand un mot fait partie de plusieurs parties du discours ? *Abstract* comme nom, verbe ou adjectif, par exemple. Par ailleurs, soulignons que les mots choisis pour mesurer une relation analytique peuvent également montrer une relation syntagmatique, comme dans *tiny electron* ou *export oversea*.

Pour contourner certains de ces problèmes, Read a développé une nouvelle version qui ne présente que des adjectifs comme mot cible. Ceux dont la définition était trop précise ont été mis de côté. Par ailleurs, pour éviter que des participants devinent les bonnes réponses au hasard, il utilise plusieurs mots leurre et les mots à associer sont présentés en deux groupes de quatre. À gauche, ce sont des mots de la même classe, pour mesurer la relation paradigmatische. À droite, ce sont des noms, pour mesurer la relation syntagmatique. Par exemple :

sudden

beautiful	quick	surprising	thirsty		change	doctor	noise	school
-----------	-------	------------	---------	--	--------	--------	-------	--------

Ainsi, avec cette dernière version, il est possible de couvrir un grand nombre de mots et plusieurs sens pour chacun de ces mots. La tâche est rapidement et facilement administrée, comme le sont les traitements et l'analyse des réponses des participants.

Schoonen et Verhallen (2008) ont développé leur propre version d'un outil de mesure de la connaissance en profondeur en s'inspirant du WaT de Read. Nous présentons l'étude de Schoonen et Verhallen (2008) dans la section 2.6. Les chercheurs ont procédé à une sélection de mots familiers (des noms, des verbes et des adjectifs) pour des enfants de 9 ans. Ils ont créé des items comprenant un mot cible, au centre, et des mots, autour, qu'il est possible d'associer au mot cible. Trois connexions devaient être établies par item, sur six possibles. Par exemple, le mot *banane* :

fruit		singe
joli	banane	peler
glisser		jaune

Il est possible d'établir des liens avec tous les mots (ils ont été obtenus à partir de réponses d'autres enfants lors d'une tâche d'association de mots). Dans cette tâche, certaines associations sont plus décontextualisées et scolaires (fruit, jaune, peler), d'autre sont plus facilement associées à un contexte particulier, familier (joli, singe, glisser).

Qian (Qian, 2002; Qian et Schedl, 2004) a validé une nouvelle version du WaT de Read, appelée mesure de la connaissance en profondeur du vocabulaire (*Depth of Vocabulary Knowledge Measure*, DVK). Le test comprend 40 items. Il mesure le sens (synonymie et polysémie) et la connaissance des collocations (relations syntagmatiques). Chaque item présente un mot cible et deux boîtes, chacune avec quatre mots. Il y a quatre bonnes réponses, mais elles ne sont pas nécessairement réparties également dans les boîtes. Ceci a pour but de mieux contrôler les chances de réussite dues au hasard. La boîte de gauche peut contenir une, deux ou trois bonnes réponses et celle de droite, respectivement trois, deux ou une bonne réponses. Par exemple :

powerful

potent	definite	influential	supportive		position	engine	repetition	price
--------	----------	-------------	------------	--	----------	--------	------------	-------

Par ailleurs, dans leur article, les auteurs présentent également le test de vocabulaire dans le cadre du *Test of English as a Foreign Language* (TOEFL Vocabulary Measure, TOEFL-VOC). Il s'agit d'un questionnaire à 40 items, à choix multiple. Par exemple :

Que signifie « powerful » dans :

Tung oil is a powerful drying agent in varnishes and paints.

- a. Pure
- b. Potent
- c. Poisonous
- d. Permanent

La comparaison entre les deux tests effectuée par Qian et Schedl (2004) montre que les résultats au DVK sont comparables à ceux obtenus au TOEFL-VOC. Cependant, lors d'entrevues passées auprès des participants, le DVK était perçu comme étant plus difficile. Les auteurs ont également souligné la difficulté de trouver des collocations indiscutables. Par contre, les résultats obtenus avec ce type de test ont montré une corrélation positive élevée avec une mesure de la compréhension (Qian, 2002).

2.5.3 Administration papier ou assistée par ordinateur

Nous avons vu que les tests de Meara et d'Alderson sont présentés par ordinateur. DIALANG établit même une connexion à une base de données sur Internet. Le développement d'outil sous forme informatique présente plusieurs avantages. Traiter les données se fait rapidement avec un plus faible risque d'erreur que lorsqu'on le fait à la main. Considérant les variations dans le climat des salles de classe, le test informatisé garantit une plus grande uniformité dans les conditions de passation, particulièrement en ce qui concerne l'uniformité des explications et de la consigne. Par ailleurs, comme les items peuvent être présentés de façon aléatoire, le test peut facilement être administré à nouveau. Finalement, des modifications et corrections apportées aux tests ont un effet immédiat, sans avoir besoin de tout réimprimer.

Dans les premiers temps de leur utilisation, la validité de tels tests a été remise en question parce que les usagers n'étaient pas tous familiers avec le support. Depuis, selon Bernhardt,

Rivera et Kamil (2004), la majorité des études portant sur l'administration de tests à l'aide de l'ordinateur montre que les résultats obtenus ainsi sont comparables à ceux obtenus à l'aide de tâches papier-crayon. Maintenant, il semble que l'utilisation de l'ordinateur à des fins de mesure et d'évaluation ne pose pas de problème en ce qui concerne la validité des résultats (Bernhardt *et al.*, 2004).

2.5.4 Conclusion

A la suite de la présentation des différents outils de mesure de l'étendue du vocabulaire, nous choisissons de développer une tâche d'items à cocher selon sa connaissance comme celle de Meara ou d'Alderson. Ce type de tâche présente plusieurs avantages : de grandes quantités de mots peuvent être présentées et traitées; le test est rapide à administrer et les résultats sont rapidement analysés (Read, 2000). Par contre, il présente quelques inconvénients : on ne sait pas jusqu'à quel point un mot est connu ni quel est le degré de connaissance d'un mot déclaré connu comparativement à un autre (Read, 2000). Pour circonvenir ces inconvénients, nous proposons de reprendre une partie des mots présentés dans une telle tâche dans une tâche de mesure de la connaissance en profondeur du vocabulaire. Cette tâche, nous la développons en nous inspirant de celles développées par Read (le DVK et le WaT).

Cependant, contrairement aux outils développés par Read, nos tâches porteront sur les noms communs identifiés dans le corpus du vocabulaire scolaire que nous avons créé (voir la partie Méthodologie). Comme les relations syntagmatiques ne peuvent pas facilement être comparées entre mots de classe grammaticale différente (Read, 2000, 2004), il vaut mieux analyser les relations syntagmatiques au sein d'une même classe grammaticale de mots cibles. Les noms du vocabulaire scolaire sont nombreux comparativement aux autres mots des diverses disciplines scolaires et ce, de façon importante (voir l'annexe C). Les résultats concernant la connaissance de ces noms devraient être un meilleur indicateur de la connaissance du vocabulaire scolaire, considérant cette importance en nombre. Par ailleurs, considérant la contrainte de temps dans la passation de tests dans le contexte scolaire, il nous faut limiter le nombre d'items qui seront présentés. Nous voulons déjà analyser divers facteurs, comme le degré d'abstraction. Il vaut mieux nous en tenir à une seule catégorie

grammaticale pour ne pas multiplier le nombre de facteurs à contrôler. S'il y avait des mots de plusieurs catégories grammaticales, cela nous obligerait à en ajouter suffisamment pour pouvoir tenir compte du facteur de la catégorie grammaticale dans nos analyses.

Par ailleurs, nous avons vu que le sens des mots dépendait du contexte dans lequel ils apparaissaient. Nous avons proposé de présenter les mots en identifiant la discipline dans laquelle ils étaient utilisés. De plus, nous choisirons des réponses qui sont en lien avec le sens du mot dans sa discipline. Cette procédure circonscrira le sens des mots dans le seul contexte de leur utilisation dans la discipline et évitera le problème de devoir tenir compte de la polysémie.

2.6 Recension des écrits : la connaissance du vocabulaire chez les élèves allophones

Les éléments présentés dans la section 2.4 expliquent comment un apprenant d'une langue seconde peut développer son vocabulaire dans cette langue. Cependant, la situation pédagogique dans laquelle évoluent les élèves allophones ne correspond pas entièrement à celles d'apprenants de classes de langue seconde. En particulier, ces derniers sont déjà scolarisés avant d'apprendre la langue cible et ne sont pas plongés dans une situation de submersion dans le système scolaire. Dans cette section, nous présentons des études susceptibles d'éclairer notre compréhension du développement langagier et plus précisément de la connaissance du vocabulaire chez des élèves allophones ou des locuteurs d'une langue seconde en submersion dans une société ou dans un système scolaire. La plupart de ces études ont été menées auprès d'élèves du primaire, sauf une, menée auprès d'élèves du secondaire (Hacquebord, 1994) et une, auprès d'adultes (Hellman, 2011). Dans la plupart de ces études, il n'est pas toujours question explicitement de mesure de l'étendue et de la profondeur de la connaissance du vocabulaire. Nous proposons quelques interprétations de ces études en relation avec l'étendue et la profondeur de la connaissance dans la conclusion de cette section.

2.6.1 Hacquebord (1994)

L'étude de Hacquebord (1994) est une recherche longitudinale menée aux Pays-Bas auprès d'élèves d'origine turque et d'élèves dont le néerlandais, langue d'enseignement, est la langue première. L'objet de l'étude est l'évaluation de la compréhension de texte à l'écrit pendant les trois premières années du secondaire. Hacquebord a comparé les performances des élèves de langue seconde à celles des élèves de langue première en lecture, en tenant compte des habiletés verbales (connaissance du vocabulaire et de la grammaire en néerlandais), des habiletés non-verbales (connaissances générales et QI non-verbal) des participants des deux groupes. Ces participants étaient au nombre de 177, dont 88 élèves d'origine turque et 89 élèves de la majorité linguistique néerlandaise des systèmes scolaires de niveau inférieur et moyen (ces systèmes étant fréquentés par la plupart des élèves d'origine turque).

Dans cette étude, les résultats au test de vocabulaire ont montré des différences entre les élèves d'origine turque et les élèves locuteurs natifs. En général, au sein d'une même école, les premiers étaient plus faibles que les derniers dans la mesure de la connaissance des mots. Le vocabulaire a progressé au long des trois ans, mais en lecture, il n'a pas été corrélé avec la performance dans le traitement de la microstructure. Il n'y a pas eu d'analyse sur la relation entre le vocabulaire et les traitements de la structure moyenne et de la macrostructure. Finalement, au bout de trois ans, si les participants d'origine turque ont amélioré leur vocabulaire, ils n'ont pas rattrapé les locuteurs natifs.

2.6.2 Verhallen et Schoonen (Verhallen et Shoonen, 1998)

Les auteurs ont vérifié si les connaissances lexicales en néerlandais d'élèves d'origine turque diffèrent en quantité et en qualité de celles d'élèves monolingues. L'étude porte sur la connaissance en profondeur du vocabulaire d'élèves d'origine turque âgés entre 9 et 11 ans. Ces élèves sont nés en Hollande. Leur compétence langagière et leurs résultats scolaires n'étaient pas exagérément hors normes. Chaque élève est passé deux fois en entrevue. À chaque fois, l'élève voyait six mots cibles supposés connus des élèves, les mêmes en

néerlandais comme en turc. Les questions de l'entrevue amenaient l'élève à raconter tout ce qu'il savait au sujet du mot cible. La première entrevue avait lieu en néerlandais. La seconde, en turc. Deux semaines en moyenne séparaient les entrevues. Les réponses ont été analysées à l'aide d'un modèle de classification sémantique élaboré par les auteurs. Les éléments de réponse étaient distribués selon le type de définition donné par les élèves, c'est-à-dire selon que les éléments établissaient une relation paradigmique ou syntagmatique avec le mot cible.

Les auteurs ont catégorisé les réponses données par les élèves le long de deux axes possibles : les réponses d'ordre paradigmique et celles d'ordre syntagmatique. Les résultats ont montré des effets de la langue (selon qu'il s'agissait de mots en turc ou en néerlandais), de la compétence dans la langue et de l'âge sur le nombre de réponses par catégorie : les mots étaient décrits plus abondamment en turc, quand l'élève était plus âgé et, dans l'une ou l'autre langue, lorsqu'il était plus compétent dans cette langue. En termes de qualité, les élèves les plus vieux favorisaient des descriptions plus paradigmiques (taxonomiques ou de méronymie) que les plus jeunes. Ces derniers employaient des descriptions plus épisodiques, c'est-à-dire en établissant des relations par association d'idée et selon des opinions personnelles (par exemple, « un nez, c'est rigolo »).

Par ailleurs, les élèves employaient des descriptions plus paradigmiques quand ils s'exprimaient en néerlandais que lorsqu'ils s'exprimaient en turc. Les descriptions à caractère épisodique ou par association se produisaient surtout pour les mots en turc. Selon les auteurs, la connaissance en profondeur du vocabulaire en langue seconde des élèves était supérieure à celle en langue première. Cependant, cette connaissance était insuffisante, comparativement à celle des élèves monolingue néerlandais. Selon les auteurs, il y a une fracture entre le développement lexical des élèves d'origine turque entre la L1 et la L2 puisque la profondeur de la connaissance des mots en L1 est en retard comparativement à la profondeur de la connaissance des mots en L2.

2.6.3 Ordonez, Carlo, Snow et McLaughlin (Ordóñez *et al.*, 2002)

Les auteurs ont vérifié si les habiletés de locuteurs de l'espagnol à définir des mots de leur L1 et à donner des informations au sujet de ces mots ressemblaient à celles destinées à traiter des mots équivalents en anglais. Il s'agit d'une comparaison entre la profondeur de la connaissance du vocabulaire en L1 et celle en L2. Pour les auteurs, les réponses qui donnent des informations sur l'apparence d'un objet, son emplacement ou son utilisation sont *syntagmatiques* : une enveloppe sert à envoyer des lettres; est rectangulaire; est faite de papier. Alors qu'une réponse *paradigmatique* donne de l'information sur la place d'un objet dans une taxonomie : le nez est une partie du corps.

Les auteurs ont interrogé des 88 enfants de 4ème et de 5ème année d'une école publique pour qu'ils définissent oralement des noms concrets connus en L1 (espagnol) et en L2 (anglais). Les auteurs s'attendaient à des réponses *paradigmatiques* dans les deux langues, mais moins de réponses *syntagmatiques* dans la L2. Les auteurs ont vérifié l'étendue du vocabulaire en L1 et en L2 avec *le Peabody Picture Vocabulary Task (PPVT)* dans sa version espagnole et anglaise.

Dans la tâche pour mesurer la profondeur de la connaissance, les enfants devaient définir six noms concrets fréquents désignant un objet ou un animal, tirés du *Peabody (PPVT)*. Les réponses ont été analysées selon les informations *paradigmatiques*, la qualité de la communication de l'enfant (comment quelqu'un pourrait reconnaître l'objet décrit par l'enfant sans voir l'objet) et les informations *syntagmatiques* (les aspects non-hiéarchiques de la définition) : taille, quantité, forme, couleur, composition, partie, emplacement, fonction, utilisation, expérience avec l'objet, usage métaphorique du mot cible.

Les résultats ont montré une certaine relation entre la profondeur de la connaissance de noms en espagnol et celle de noms en anglais. Les connaissances *paradigmatiques* en espagnol sont corrélées aux connaissances *paradigmatiques* en anglais pour les élèves avec une plus grande étendue de vocabulaire en espagnol. Il y a un effet possible de facilitation pour les mots apparentés entre les deux langues. Les résultats ont également montré une corrélation entre

les connaissances syntagmatiques et l'habileté de communication en L1 et les connaissances syntagmatiques et l'habileté de communication en L2, pour ceux qui obtenaient les meilleures performances au PPVT en L1 et L2.

Selon les auteurs, les mots des deux langues sont représentés dans un lieu d'entreposage en mémoire des concepts partagés (Kroll et de Groot, 1997 ; voir les catégories partagées dans leur modèle dans la section précédente). Ce modèle prédit que le développement du vocabulaire en L2 sera plus efficace pour des concepts existants déjà en L1. Dans ce cas, acquérir de nouveaux mots de l'anglais consiste à apprendre de nouvelles formes et non pas des concepts qui existent déjà.

2.6.4 Droop et Verhoeven (Droop et Verhoeven, 2003)

L'étude constitue un apport important à la problématique des élèves allophones, prenant en compte de multiples facteurs et procédant à des analyses sophistiquées de leurs interactions. Il s'agit d'une étude longitudinale. Le but de l'étude de Droop et Verhoeven est d'explorer l'influence des différents aspects du développement de la performance langagière sur le développement de la lecture en langue seconde. Ces aspects sont les connaissances lexicales, les connaissances morphosyntaxiques, la compréhension à l'oral et le décodage des mots écrits. De plus, les auteurs ont pris en compte la langue première, la langue seconde et le statut socioéconomique (SSE). Les participants étaient 143 élèves locuteurs du néerlandais, langue d'enseignement, 62 élèves d'origine turque et 60 d'origine marocaine. Tous les élèves sont nés aux Pays-Bas ou ont commencé la maternelle aux Pays-Bas. L'âge moyen des élèves au début de l'étude était d'environ 8 ans 7 mois pour les élèves locuteurs natifs, de 8 ans 1 mois pour les élèves d'origine marocaine et de 8 ans 11 mois pour les élèves d'origine turque.

Au test de vocabulaire en lecture de l'étude, les résultats montrent des différences importantes entre les groupes. Au début de la troisième année, les enfants des minorités ont obtenu un résultat moyen équivalent au hasard. À la fin de la quatrième année, les élèves locuteurs natifs de SSÉ élevé obtiennent les meilleurs résultats, suivis par les élèves locuteurs natifs de SSÉ faibles, puis par les élèves d'origine marocaine et, finalement, par les élèves

d'origine turque. Il existe une grande variation à l'intérieur des groupes. Dans les tests de vocabulaire en lecture et en production, une différence croissante a été observée entre les élèves L1 de SSE faible et les élèves des minorités par rapport aux élèves L1 de SSE élevé. Les deux tests demandaient des connaissances de la langue décontextualisées et exigeantes cognitivement. Selon les auteurs, la question du déclin de la quatrième année semble se poser : « More generally, it can be assumed that any differences in the children's reading comprehension will increase when children are confronted with texts containing more complex and abstract language » (p. 97). Par ailleurs, pour les élèves issus de minorité, la performance aux tests de compréhension à l'oral était principalement expliquée par le vocabulaire et les habiletés syntaxiques à la fin de la troisième et de la quatrième année du primaire. Cela suggère que la compréhension des élèves à l'oral dépend davantage de la maîtrise du vocabulaire que celle des élèves issus de la majorité. En somme, en fin de troisième et de quatrième année, pour les élèves issus des minorités, la connaissance du vocabulaire à l'oral et à l'écrit influence de façon importante la compréhension à l'écrit.

2.6.5 Lefrançois et Armand (Lefrançois et Armand, 2003)

Bien qu'elle ne porte pas sur le vocabulaire à proprement dit, cette étude apporte une contribution importante : elle est la seule au Québec, à notre connaissance, qui prend en considération le facteur de la littératie développée dans la langue maternelle d'élèves allophones. L'objectif de l'étude était d'explorer le rôle de la conscience phonologique et de la conscience syntaxique en langue seconde à travers quatre composantes de la lecture : le décodage, la reconnaissance des mots, la compréhension de phrase et la compréhension de texte. Les participants étaient 38 locuteurs natifs de l'espagnol qui devaient avoir commencé en classe d'accueil depuis sept mois, avoir fréquenté l'école du pays d'origine pendant au moins trois ans, être âgés entre 9 et 11 ans et ne pas connaître d'autres langues que le français et l'espagnol. L'étude a pris en compte l'habileté de lecture en langue première chez des élèves allophones. L'analyse des résultats a montré que cette habileté était plus corrélée à l'habileté de lecture en français que la compétence à l'oral en français. Bien qu'il ne s'agisse pas de résultats relatifs au vocabulaire, l'étude de Lefrançois et Armand (2003) montre qu'il

faut prendre en compte les habiletés de littératie dans la L1 dans la mesure des habiletés de littératie dans une L2.

2.6.6 Morris et Labelle (Morris et Labelle, 2008)

L'étude de Morris et Labelle porte sur une population scolaire de la région de Montréal. Elle a pour but l'élaboration d'un protocole d'évaluation de la compétence langagière à l'écrit et à l'oral en français auprès d'élèves allophones. Il s'agit d'une étude transversale effectuée auprès de 2 248 élèves allophones, autochtones et francophones de tous les cycles du primaire. Nous ne tenons compte ici que des données relatives aux élèves allophones et francophones. Il est fait mention de plus de 50 langues maternelles différentes parlée par les élèves allophones. Ces élèves allophones fréquentaient des écoles dans des milieux de statut socioéconomique plutôt faible, alors que les cohortes d'élèves francophones proviennent de milieux plus favorisés.

Les mesures du vocabulaire de l'étude de Morris et Labelle ont révélé que les élèves allophones obtenaient des résultats équivalents à ceux des francophones en première, troisième, quatrième et sixième année. Pour expliquer des résultats aussi élevés chez les élèves allophones, il faut préciser que la mesure utilisée dans l'étude reposait sur des mots parmi les plus fréquents en français. Les élèves allophones sont plus susceptibles d'avoir appris ces mots. En ce qui concerne les résultats des élèves francophones de la sixième année, plutôt faibles dans l'étude, les auteures rappellent leur manque de motivation à répondre aux tests de l'étude en fin d'année scolaire.

2.6.7 Schoonen et Verhellen (Schoonen et Verhellen, 2008)

Cette étude a spécifiquement mesuré le vocabulaire des élèves allophones. Le but était de développer et de valider un test de vocabulaire mesurant la connaissance en profondeur des mots. Les participants étaient des élèves néerlandais de la deuxième moitié du primaire, âgés entre 9 et 11 ans, de troisième ou cinquième année. Ils étaient au nombre de 795, de langue maternelle néerlandaise, turque, marocaine et quelques autres.

Les auteurs ont voulu mesurer les différentes relations établies entre les mots auprès des élèves allophones néerlandais. Les chercheurs rappelaient à quel point les capacités à catégoriser, à généraliser et à concevoir de façon abstraite sont importantes pour le développement du lexique, particulièrement pour donner du sens lorsque les mots sont présentés de façon décontextualisée. Les chercheurs ont procédé à une sélection de mots familiers pour des enfants de 9 ans. Ils ont choisi des noms, des verbes et des adjectifs. Ils ont créé des items formés à la suite d'une étude pilote auprès d'élèves de 3ème à 5ème année qui devaient expliquer les associations de mots qu'ils faisaient avec un mot cible. Le test consistait à dessiner une ligne entre un mot cible, au centre, et des mots auxquels il peut être associé, autour. Trois connexions devaient être établies sur six possibles.

Les résultats ont montré des différences dans la connaissance de la profondeur des mots entre les élèves de 9 ans et 11 ans et entre les élèves allophones et les élèves issus de la majorité, les élèves allophones étant plus faibles que les élèves issus de la majorité. Par ailleurs, les analyses ont montré un effet d'interaction entre l'âge et la langue maternelle : les progrès liés à l'âge sont plus importants que ceux liés à la langue maternelle. L'interaction pouvait cependant s'expliquer par un effet de plafonnement à l'épreuve chez les élèves locuteurs natifs du néerlandais de cinquième année.

2.6.8 Jean et Geva (Jean et Geva, 2008)

L'étude de Jean et Geva (2008) porte sur le développement du vocabulaire en anglais langue seconde et son rôle pour prédire l'habileté de reconnaissance des mots. Deux buts sont énoncés : déterminer s'il existe des différences dans la connaissance du vocabulaire scolaire entre des élèves dont l'anglais est la langue première et des élèves pour qui l'anglais est une langue seconde; vérifier l'existence d'une relation entre la reconnaissance des mots et la connaissance du vocabulaire chez ces élèves. Il s'agit d'une étude longitudinale auprès de différents groupes d'élèves de cinquième et sixième année de différentes écoles d'une grande ville canadienne. Les écoles sont situées dans des quartiers dont le statut socio-économique des habitants est relativement bas. L'étude a porté sur 207 élèves, c'est-à-dire 61 dont l'anglais est la L1 et 146 dont l'anglais est une L2. L'âge moyen est de 10 ans et 10 mois. Les

élèves dont l'anglais est une L2 évoluent en classe régulière depuis qu'ils ont commencé l'école, mais les enseignants adaptent leur enseignement à ces élèves. De plus, pendant les deux premières années, les élèves ont pu recevoir quotidiennement de 30 à 40 minutes d'enseignement en anglais langue seconde.

Les différentes mesures ont porté sur les habiletés cognitives non verbales, sur la mémoire de travail, sur l'habileté d'énumération rapide (*Rapid automatized naming* - RAN, qui consiste à nommer des séries de cinq lettres les plus fréquentes de l'anglais présentées aléatoirement), sur la conscience phonologique et sur le vocabulaire. Dans ce dernier cas, les chercheurs ont eu recours au *Peabody Picture Vocabulary Test* (revu) pour mesurer le vocabulaire réceptif : il faut pointer l'image représentant un mot entendu parmi quatre images. Elles ont également eu recours au *Written root word vocabulary task* de Biemiller et Slonim (2001) qu'elles ont modifié : l'enfant voit une phrase qui contient un mot cible ; il doit écrire une phrase qui explique le sens du mot cible. (Le test original propose des choix de réponses, mais il a la réputation de surestimer la connaissance des élèves de ce groupe d'âge.) La reconnaissance des mots a également été mesurée à l'aide d'un test de lecture de mots monosyllabiques et polysyllabiques isolés et d'un test de lecture de mots d'une liste de mots racines.

Les résultats ont montré que les élèves des deux groupes linguistiques ne se distinguent pas en ce qui concerne la conscience phonologique. Les élèves dont l'anglais est une L2 ont connu une plus grande amélioration de la cinquième à la sixième année pour les habiletés cognitives non verbales que les élèves dont l'anglais est une L1. Il n'y a pas eu de différence dans les résultats au test sur la mémoire de travail. Il y a eu une différence significative entre les deux groupes dans le test sur le vocabulaire réceptif, de même qu'il y a eu une différence entre les groupes et une interaction avec le temps pour la tâche sur l'habileté à nommer rapidement : les élèves dont l'anglais est une L1 réussissaient mieux que ceux dont l'anglais est une L2 ; mais les deux groupes se sont améliorés avec le temps. Les élèves de sixième année dont l'anglais est une L1 ont mieux réussi la tâche portant sur les mots racines. Les deux groupes d'élèves ont amélioré leur habileté de reconnaissance des mots avec le temps pour les deux groupes d'élèves. Pour résumer, les résultats des deux groupes ne se sont pas distingués pour les mesures de conscience phonologique, d'énumération rapide, de mémoire

de travail. Par contre, les élèves dont l'anglais est une langue première ont obtenu de meilleurs résultats que leurs pairs pour qui l'anglais est une langue seconde en ce qui concerne les mesures sur le vocabulaire.

À la lumière de ces résultats, les auteures concluent que, relativement au vocabulaire, à la fin des années du primaire, les élèves L2 maîtrisent moins le vocabulaire que les élèves L1 même si les élèves des deux groupes continuent à développer leurs habiletés lexicales au cours de la scolarité. L'analyse en profondeur des résultats de la tâche sur les mots racines (*Written root word vocabulary task*) permet de mieux comprendre les différences en les deux groupes en ce qui concerne le vocabulaire scolaire. En cinquième année, les élèves des deux groupes connaissent les mots censés être connus par les enfants de deuxième et de quatrième année. Par contre, aucun des groupes n'a atteint la norme établie pour ces mots. Bien que ces mots doivent être maîtrisés au cours des années du primaire, ça n'est pas le cas pour une partie importante des deux groupes linguistiques. Les auteures expliquent ces résultats par un effet du statut socio-économique faible des élèves. En ce qui concerne les élèves L2, en cinquième année, ils ont un taux de connaissance du vocabulaire scolaire semblable aux pairs monolingues de la deuxième année. Au même moment, les élèves L1 poursuivent le développement de leur vocabulaire scolaire à un rythme plus rapide que les élèves L2, creusant ainsi l'écart entre les deux groupes.

2.6.9 Bialystok, Luk, Peets et Yang (Bialystok, Luk, Peets et Yang, 2009)

Les auteures rappellent les études qui montrent que les élèves bilingues du primaire connaissent moins de mots dans leurs deux langues que les élèves monolingues dans la leur. Ces résultats ont été obtenus à l'aide du test *Peabody Picture Vocabulary Test* (PPVT). Elles expliquent ces résultats par le fait que les élèves bilingues doivent diviser le temps disponible à l'apprentissage de la langue entre deux langues plutôt qu'une seule. Par contre, elles proposent que les élèves bilingues n'en sont pas pénalisés pour autant : certains mots seraient appris dans une langue dans un contexte donné et d'autres, dans l'autre langue, le seraient dans un autre contexte. Un contexte dans lequel il est important de développer le vocabulaire est le contexte scolaire. Les auteurs rappellent que la taille du vocabulaire est un facteur de

réussite scolaire et de développement de la littératie. Si les élèves bilingues ont un plus petit vocabulaire que les élèves monolingues, ils risquent de moins bien réussir à l'école. Cependant, selon les auteures, il est difficile de généraliser à partir des études à ce sujet, parce que chaque étude s'est concentrée sur un petit nombre d'élèves, chacune d'un groupe d'âge particulier. Elles ont voulu vérifier dans quelle mesure les différences dans le vocabulaire scolaire trouvées dans différentes études entre les élèves bilingues et monolingues sont le reflet de la situation générale vécue par un grand nombre d'élèves de tous âges. Elles ont procédé à une analyse par agrégation des données de différentes études. Les données obtenues au test PPVT ont été combinées et soumises à ce type d'analyse.

Les scores de 1 738 élèves de 3 à 10 ans ont été inclus dans l'analyse. Ces données proviennent d'études ayant eu lieu au cours d'une période de cinq ans. L'échantillon est composé de 772 enfants anglophones monolingues et 966 enfants bilingues. L'analyse a montré que les élèves monolingues obtiennent de meilleurs scores que les élèves bilingues à tous âges. Cependant, les auteures ont analysé les données d'une partie des élèves (161 élèves âgés entre 6 et 11 ans) en fonction des mots du test qui correspondent au vocabulaire de la conversation courante et à ceux qui correspondent davantage au vocabulaire scolaire. L'analyse a montré que les élèves monolingues réussissaient mieux le PPVT en général, mais qu'il y avait peu de différence dans la connaissance du vocabulaire scolaire entre les élèves bilingues et les élèves monolingues. Les auteures concluent que les élèves bilingues ne semblent pas particulièrement désavantagés en ce qui concerne la connaissance du vocabulaire scolaire tout en reconnaissance que de nouvelles études devraient être entreprises en considérant un plus grand éventail de catégories de mots.

2.6.10 Hellman (2011)

Cette étude n'a pas été menée auprès d'élèves, mais auprès de personnes qui ont commencé à apprendre une langue seconde à l'âge adulte. Elle nous intéresse parce qu'elle porte sur la mesure de l'étendue et de la profondeur de la connaissance du vocabulaire d'une langue seconde et parce qu'elle vérifie si des apprenants tardifs d'une langue seconde peuvent développer un vocabulaire comparable à celui de locuteurs natifs.

D'après l'auteure, en général, ceux qui ont commencé à apprendre une langue seconde à l'âge adulte n'atteignent pas le niveau de compétence des locuteurs natifs. La compétence sur le plan phonologique est rarement atteinte; les résultats sur la compétence morphosyntaxique sont matière à débat, mais les adultes L2 n'atteignent pas non plus le niveau des L1, à part quelques-uns. Par contre, l'auteur rappelle qu'il y a peu de données sur le niveau de vocabulaire atteint, en termes d'étendue et de profondeur de la connaissance. Selon elle, contrairement aux compétences phonologiques et morphosyntaxiques, les personnes qui apprennent une L2 après l'âge de 20 ans seraient plus susceptibles de développer des connaissances lexicales qui ressemblent à celles des locuteurs natifs si on se fie aux résultats de recherches en neurolinguistique lors de tâches morphosyntaxiques et lexicales. Ces résultats et ceux issus d'études comportementales suggèrent que le lexique mental peut s'adapter aux nouvelles informations au cours de toute la vie, dont celles apprises en L2. À la lumière de ces études, les apprenants d'une L2 à l'âge adulte pourraient développer leurs compétences lexicales à un niveau près de celles des locuteurs natifs.

Le but de l'étude est donc de vérifier les limites de la compétence lexicale atteinte par des personnes qui ont commencé à apprendre l'anglais à l'âge adulte et qui ont été exposées à l'anglais tout au long de leur vie adulte. Deux dimensions du développement du vocabulaire sont mesurées : l'étendue et la profondeur de la connaissance. L'étude a porté auprès sur 33 participants adultes dont la L1 est le hongrois (langue sans lien avec l'anglais) et apprenants de l'anglais L2, de 30 locuteurs natifs de l'anglais et de 30 locuteurs bilingues hongrois/anglais.

L'étendue du vocabulaire a été mesurée à l'oral, avec le *Peabody Picture Vocabulary Test* (PPVT) et à l'écrit, avec un *Self-Rated Vocabulary Test* portant sur deux listes de mots tirés du dictionnaire *Oxford American Dictionary*. La profondeur a été mesurée avec le *Word Associates Test*, version révisée de 1998.

Les résultats ont montré que le groupe anglais L2 était plus faible que le groupe de locuteurs natifs, à cause principalement des résultats au test PPVT. Par contre, les plus forts du groupe anglais L2 atteignent le niveau des locuteurs natifs pour les mesures SRVT et WAT après 20

ans d'exposition à l'anglais. Le groupe bilingue obtient des résultats semblables aux locuteurs natifs. Et dans le groupe anglais L2, 76 % ont des scores qui approchent « raisonnablement » (non précisé) celui des L1.

Selon l'auteure, le niveau de locuteur natif peut être atteint chez des apprenants d'une L2 ayant débuté l'apprentissage à l'âge adulte après une immersion à 20 ans et plus. Dans ce groupe, 76 % ont atteint un niveau proche du locuteur L1. Par contre, d'autres facteurs ont pu favoriser la performance, puisque les participants du groupe anglais L2 ont été sélectionnés pour leur scolarité élevée et leur performance générale en apprentissage de la L2. Selon l'auteure, la taille du vocabulaire en L2 n'est pas limitée par le vieillissement, probablement parce que le vocabulaire continue de s'enrichir toute la vie. Par contre, certains aspects sont peut-être plus contraints par le vieillissement, comme le traitement phonologique, la vitesse du traitement, l'apprentissage des collocations, des expressions et des locutions.

2.6.11 Conclusion

Dans la problématique, nous avons souligné que la recherche en langue première montre que ce ne sont pas tous les élèves qui parviennent à apprendre le très grand nombre de mots qu'il est possible de rencontrer dans le cadre scolaire. Des facteurs comme la culture de la littératie à la maison et les attitudes vis-à-vis des activités de littératie peuvent aider ou nuire à l'acquisition du vocabulaire. Par ailleurs, il semble qu'il faille tenir compte du statut socioéconomique bien qu'il ne soit pas clair jusqu'à quel point il constitue en soi un facteur d'apprentissage. Les travaux résumés dans cette section nous permettent de mieux appréhender l'état de la connaissance du vocabulaire scolaire et d'autres facteurs qui y contribuent chez les élèves allophones. Contrairement aux élèves locuteurs natifs, les élèves allophones ne connaissent pas la langue d'enseignement quand ils commencent l'école, à l'exception de ceux qui y sont initiés dans des classes d'accueil. Les élèves allophones développent-ils leur vocabulaire de façon comparable à celle des élèves locuteurs natifs ? La recension permet de tirer quelques conclusions à ce sujet.

En général, les élèves allophones développent plusieurs habiletés langagières dans la langue d'enseignement au point d'obtenir des résultats semblables à ceux d'élèves pour qui la langue d'enseignement est une L1 (Lefrançois et Armand, 2003; Morris et Labelle, 2008). Cependant, la plupart des études montrent que les élèves allophones n'ont pas développé une connaissance du vocabulaire comparable à celle des élèves locuteurs natifs (Bialystok *et al.*, 2009; Droop et Verhoeven, 2003; Hacquebord, 1994; Jean et Geva, 2008; Schoonen et Verhellen, 2008; Verhellen et Shoonen, 1998). L'étendue du vocabulaire n'a pas été explicitement mesurée dans les recherches recensées dans ce travail. La profondeur de la connaissance a davantage été prise en compte. Les résultats des recherches à ce sujet montrent que la profondeur de la connaissance du vocabulaire en L2 est moins développée que celles des locuteurs natifs (Schoonen et Verhellen, 2008; Verhellen et Shoonen, 1998). Elle est par ailleurs positivement corrélée avec des habiletés langagières développées dans la langue maternelle (Hellman, 2011; Ordóñez *et al.*, 2002). Par contre, les élèves allophones dans certaines études évoluent dans des milieux socioéconomiques défavorisés (Droop et Verhoeven, 2003; Hacquebord, 1994; Schoonen et Verhellen, 2008). Nous observons que des résultats des élèves allophones de l'étude de Droop et Verhoeven (2003) rejoignent les résultats des élèves locuteurs natifs de faible statut socioéconomique. Jusqu'à quel point le statut socioéconomique n'est-il pas un facteur plus important que la connaissance de la langue maternelle ?

En ce qui concerne le vocabulaire scolaire, l'analyse effectuée par Bialystok *et al.* (2009) suggère que ce vocabulaire puisse être développé quand la langue d'enseignement est une langue seconde, même si le vocabulaire de la conversation quotidienne ne l'est pas. De plus, la profondeur de la connaissance du vocabulaire peut être mieux développée dans la L2 que dans la L1 (Verhellen et Shoonen, 1998). Par ailleurs, les mots les plus fréquents de la langue d'enseignement peuvent être connus de façons comparables par les élèves allophones et les élèves locuteurs natifs (Morris et Labelle, 2008). Finalement, les habiletés de littératie développées en L1 peuvent avoir une influence sur les habiletés développées en L2 (Lefrançois et Armand, 2003; Ordóñez *et al.*, 2002).

À la lumière des résultats des études présentées plus tôt, ce ne sont pas tous les élèves allophones qui ne développent pas des connaissances lexicales suffisantes dans la L2. Plusieurs sont susceptibles d'y parvenir, s'ils ont des pratiques de littératie à la maison, si celles-ci sont valorisées ou si ces pratiques se déroulent dans la L1. De plus, le temps, ou l'expérience scolaire, semble être un facteur plus important d'acquisition que la langue maternelle (Droop et Verhoeven, 2003; Jean et Geva, 2008; Schoonen et Verhellen, 2008). Par ailleurs, en langue première comme en langue seconde, la recherche sur le développement du vocabulaire dans le contexte scolaire a essentiellement été réalisée auprès d'élèves du primaire. L'étude de Hacquebord (1994) a bien montré une différence entre des élèves L1 et des élèves allophones au secondaire, mais elle a été réalisée il y a maintenant près de 20 ans. Depuis, il s'est produit une grande sensibilisation aux problèmes langagiers auxquels sont confrontés les élèves allophones et des mesures ont été adoptées. Jusqu'à quel point la situation a-t-elle changé ? Au Québec, la question se pose avec d'autant plus d'acuité que le vocabulaire et la population des élèves du secondaire sont rarement l'objet d'études.

2.7 Objectifs de recherche

Dans la première partie, nous avons vu que les élèves du secondaire avaient besoin de connaître les mots du vocabulaire scolaire pour comprendre les contenus scolaires. Nous avons rappelé que, parmi ces élèves, ceux qui ont commencé leur scolarité sans connaître la langue d'enseignement étaient susceptibles de ne pas développer une connaissance suffisante de ce vocabulaire. Au début de la deuxième partie, nous avons montré le lien qui existait entre la connaissance du vocabulaire et la lecture, soulignant à quel point cette connaissance est importante pour lire couramment. Ensuite nous avons défini ce que nous entendions par vocabulaire scolaire. Nous avons défini celui-ci comme étant les mots des différents domaines d'apprentissages. Puis nous avons défini ce que connaître un mot veut dire. Nous avons expliqué que le nombre de mots autant qu'une connaissance en profondeur des mots importaient dans la compréhension de textes. Nous avons montré qu'au fur et à mesure que le nombre de mots connus augmentait, des liens entre les mots se tissaient, approfondissant la connaissance du vocabulaire. Ces liens organisés semblent favoriser une récupération des mots pour les comprendre ou les produire efficacement et rapidement. Nous avons mis

l'accent sur les liens syntagmatiques et les liens paradigmatiques, les premiers indiquant une connaissance de l'usage des mots en contexte, les seconds indiquant une connaissance du sens du mot au sein d'un système organisé, tel qu'il est présenté dans le monde scolaire. De plus, nous avons souligné l'importance de prendre en considération le degré d'abstraction des mots comme facteur d'apprentissage du vocabulaire. Par la suite, nous avons exposé différents modèles du lexique du locuteur bilingue. Dans ces modèles, les mots sont stockés séparément dans les deux langues tout en partageant certains aspects dont certaines catégories conceptuelles. D'autres catégories conceptuelles sont spécifiques à une seule des deux langues. Comme les travaux présentés dans cette section ont porté auprès de locuteurs qui maîtrisaient leur L1 avant d'apprendre une L2, nous nous sommes interrogés sur le développement du lexique des allophones quand ceux-ci ne développent pas des habiletés de littératie dans leur propre langue. Nous nous sommes également demandé comment les élèves allophones pouvaient développer la connaissance des mots plus rares et plus abstraits du langage scolaire.

La recension des écrits a révélé que les attitudes et habitudes développées à la maison concernant les activités de littératie pouvaient influencer le développement du lexique. C'est le cas en particulier des habitudes de lecture. En ce qui concerne la connaissance du vocabulaire scolaire chez les élèves allophones, nous avons vu que ceux-ci ne semblent pas avoir rattrapé les locuteurs pour qui la langue d'enseignement est la langue première. Plus particulièrement, les élèves allophones ne semblent pas avoir développé une connaissance en profondeur des mots de la langue seconde comme elle est développée chez les locuteurs natifs. Cependant, certains résultats montrent que les élèves allophones du primaire peuvent développer une connaissance des mots scolaires sans développer pour autant une connaissance des mots plus domestiques. Finalement, nous n'avons que très peu de données pour les élèves au secondaire, alors qu'ils évoluent dans le système scolaire francophone depuis plusieurs années.

Nous avons décrit différents outils pour mesurer, dans le cadre de notre travail, l'étendue et la profondeur du vocabulaire. Par ailleurs, nous avons souligné le besoin d'outils adaptés pour mesurer la connaissance du vocabulaire scolaire en français au Québec. Pour mesurer le

vocabulaire scolaire chez les élèves allophones, nous choisissons de développer deux tests informatisés, l'un pour mesurer l'étendue du vocabulaire, l'autre pour mesurer la connaissance en profondeur du vocabulaire. Les deux tests sont des épreuves basées sur la reconnaissance et la compréhension.

À la lumière du contenu présenté jusqu'à présent, nous précisons le but et les objectifs de notre étude. Le but de l'étude est de dresser un portrait de la connaissance du vocabulaire scolaire des élèves allophones après plusieurs années dans le système francophone. Nous avons terminé la première partie en posant une grande question de recherche, que nous répétons ici :

Quel est le degré d'acquisition du vocabulaire scolaire chez les élèves allophones du système scolaire francophone québécois après plusieurs années dans ce système ?

Plus spécifiquement, compte tenu de tous les éléments apportés jusqu'ici, nous proposons de réaliser les objectifs de recherche qui suivent.

1. Évaluer l'étendue du vocabulaire scolaire d'élèves allophones du secondaire.

La connaissance préalable des mots étant un facteur de la compréhension en lecture et les mots à connaître étant particulièrement nombreux, nous voulons vérifier si les élèves allophones ont développé leur vocabulaire en étendue. Leur résultat sera comparé à celui d'élèves francophones du même âge, le résultat de ces derniers constituant une balise de la normalité attendue de l'étendue.

2. Évaluer la profondeur de la connaissance du vocabulaire scolaire d'élèves allophones du secondaire, c'est-à-dire plus précisément évaluer les connaissances paradigmatisques et syntagmatiques de noms des domaines d'apprentissages du premier cycle du secondaire.

Pour comprendre les textes qu'ils lisent et réussir dans les différents domaines de connaissance scolaires, les élèves doivent maîtriser le vocabulaire de ces domaines. Après

plusieurs années dans le système scolaire francophone, les élèves allophones, qui ne connaissent pas la langue d'enseignement au moment d'entreprendre leur scolarité, doivent à la fois apprendre la langue et apprendre le contenu des domaines d'apprentissages. Nous nous demandons s'ils sont parvenus à le faire, particulièrement en ce qui concerne la connaissance du vocabulaire. Pour vérifier si les élèves allophones ont développé une maîtrise suffisante du vocabulaire des domaines d'apprentissages, nous mesurons deux dimensions de la profondeur de la connaissance du vocabulaire : les connaissances paradigmatisques et les connaissances syntagmatiques. D'une part, les connaissances paradigmatisques sont particulièrement favorisées dans le monde scolaire et demandent l'établissement de relations hiérarchiques entre les mots. D'autre part, les mots du vocabulaire scolaire apparaissent fréquemment dans des syntagmes complexes avec d'autres noms, verbes ou adjectifs spécialisés. Leur compréhension demande d'établir des relations syntaxiques et sémantiques avec les mots qui les composent. Les connaissances syntagmatiques permettent à l'élève de comprendre ces syntagmes, soit en faisant appel à ses connaissances syntaxiques, soit en récupérant le sens de blocs de mots en mémoire. Pour mieux situer le niveau de connaissance en profondeur atteint par les élèves allophones, nous comparons leurs résultats aux deux mesures à ceux d'élèves francophones du même âge.

3. Mesurer la relation entre le degré d'abstraction des mots et la profondeur de la connaissance du vocabulaire des élèves.

Les mots des domaines d'apprentissages sont souvent abstraits et appris de façon décontextualisée (c'est-à-dire appris de façon verbale, sans expérience personnelle directe avec le référent). Pour comprendre les mots plus abstraits, il faut avoir recours à des connaissances préalables, sémantiques et syntaxiques, en faisant appel particulièrement à son système de représentation verbale. Celui-ci risque d'être sous-développé chez les élèves allophones, du moins dans la langue d'enseignement. Il peut également l'être chez les élèves francophones. Voilà pourquoi nous mesurons l'effet du degré d'abstraction des mots : les mots plus abstraits devraient être moins bien réussis que les mots plus concrets, particulièrement chez les élèves allophones. Par ailleurs, dans les analyses, nous prenons en considération la longueur et la fréquence des mots parce qu'ils peuvent covarier avec le degré

d'abstraction (Ferrand, 2001). Nous voulons distinguer l'effet éventuel de ces variables de celui du degré d'abstraction pour mieux mesurer l'effet de celui-ci sur les résultats.

Dans la prochaine partie, nous exposons la procédure que nous avons adoptée pour atteindre ces différents objectifs.

CHAPITRE III

MÉTHODOLOGIE

3.1 Introduction

Notre étude a pour but de répondre à la question : quel est le degré d'acquisition du vocabulaire scolaire chez les élèves allophones du système scolaire francophone québécois après plusieurs années dans ce système ? Nous avons rappelé qu'à partir de la quatrième année du primaire, l'élève, peu importe sa langue maternelle, doit apprendre de plus en plus de mots dans le cadre de sa scolarité. Ces mots sont souvent abstraits et retrouvés principalement à l'écrit. Nous avons souligné que l'apprentissage de ces mots posait problème pour nombre d'élèves. Plus particulièrement, nous avons suggéré que ce pouvait être le cas des élèves pour qui la langue d'enseignement n'est pas la langue première. Nous avons souligné le manque d'études à ce sujet et proposé plusieurs objectifs de recherche pour mieux connaître la situation. Nous voulons mesurer l'étendue du vocabulaire scolaire des élèves allophones du secondaire et la profondeur de la connaissance de ce vocabulaire. En plus, nous avons vu que le degré d'abstraction des mots pouvaient affecter l'apprentissage. Nous voulons vérifier les effets de ce facteur sur la connaissance du vocabulaire des élèves. Dans cette partie, nous décrivons la procédure mis en œuvre et les instruments utilisés pour atteindre ces objectifs. Nous expliquons d'abord comment nous avons pris en considération des principes éthiques qui ont encadré ce travail. Nous décrivons ensuite les participants et la procédure mise en œuvre pour recueillir les données. Nous décrivons également la démarche entreprise pour identifier les mots du vocabulaire scolaire à considérer dans nos épreuves et celles pour développer les différents items des tests de connaissance du vocabulaire. Au cours de ces descriptions, nous montrons comment les différents outils de mesure sont intégrés

dans la suite informatique que nous avons développée. Nous terminons cette partie par une description des données et du traitement que nous en ferons.

3.2 Éthique

Comme notre étude implique la participation d'êtres humains, nous nous sommes assuré de respecter les principes et règles suggérés par le cadre normatif pour l'éthique de la recherche avec des êtres humains de l'UQAM.¹⁷ Toutes les instances participant de près ou de loin à notre projet ont été informées concernant le but de notre étude. Nous avons obtenu le consentement de la direction de l'école pour procéder à notre étude au sein de l'institution. De même, nous avons obtenu de tous les enseignants des groupes participant à notre étude leur consentement écrit. Les élèves et leurs parents ont pris connaissance du projet et ont pu faire un choix éclairé quant à leur participation à ce projet. Pour participer à notre étude, les élèves devaient avoir remis une feuille de consentement dûment remplie par les parents. Nous avons informé tous les élèves qu'ils pouvaient se retirer en tout temps, que leur anonymat serait respecté, leurs réponses associées à un code attribué par l'application et récupérables seulement par le chercheur. Nous avons également informé les parents et les élèves que les risques encourus par les élèves qui participaient à l'étude étaient comparables à ceux présents lors d'une journée scolaire typique. Un exemplaire des feuilles de consentement et le certificat d'éthique complété ont été remis au Département de linguistique de l'UQAM (voir un exemplaire à l'annexe A).

3.3 Les participants

3.3.1 Description

Les élèves qui ont participé à notre étude fréquentent une école secondaire privée de Montréal. L'école est située dans un quartier multiethnique et est fréquentée par une telle

¹⁷ <http://www.recherche.uqam.ca/ethique/humains/cadre-normatif-ethique-humains.html#3-principes-ethiques-directeurs>

population. Les places sont limitées et il y a une procédure de sélection des élèves lors d'une période d'examens d'admission. Les élèves proviennent surtout de milieu socioéconomique moyen à élevé. Au total, 627 élèves ont répondu aux questionnaires et procédé aux tests de vocabulaire. Le tableau 3.1 détaille la distribution des élèves en fonction du niveau de scolarité et du sexe. Les groupes d'élèves se répartissent en quatre classes de secondaire 1, dix classes de secondaire 2 et sept classes de secondaire 3. En secondaire 1, ils sont 93, en secondaire 2, 247 et en secondaire 3, 287. Le nombre de garçons s'élève à 392, le nombre de filles à 235. L'âge des élèves se situe entre 12 et 18 ans.

Tableau 3-1 Distribution des élèves de l'échantillon en fonction du niveau de scolarité et du sexe.

	Nombre de groupes	Nombre de garçons	Nombre de filles	Nombre total	Âge moyen
Secondaire 1	4	58	35	93	13,3
Secondaire 2	10	154	93	247	14,3
Secondaire 3	7	180	107	287	15,2
Total	21	392	235	627	

3.3.2 Détermination des groupes d'élèves allophones et francophones

Le premier traitement des données auquel nous avons procédé a été de distinguer les élèves allophones des élèves francophones. Le groupe d'élèves francophones sert de point de comparaison, de norme pour évaluer le degré de connaissance du vocabulaire des élèves allophones. Dans le questionnaire sociodémographique (voir section 3.6.2.2), les élèves pouvaient indiquer jusqu'à trois langues parlées à la maison, la fréquence de l'usage de ces langues et les personnes avec qui ils la parlaient, dont la mère. Les données obtenues suggèrent une situation complexe où plusieurs langues, dont le français et l'anglais, sont parlées à la maison. Pour distinguer les élèves allophones des élèves francophones, nous avons considéré la langue parlée principalement avec la mère (Armand, 2005; Morris et Labelle, 2008). Même dans ce cas, nous observons que plusieurs élèves font usage de plus d'une langue avec leur mère. Les réponses des élèves indiquent que 312 élèves parlent uniquement le français avec leur mère et que 69 élèves parlent le français et l'anglais avec elle. La fréquence indiquée pour l'usage du français parlé était « Fréquemment » pour tous ces élèves. Ces groupes deviennent le groupe contrôle d'élèves francophones dans notre étude.

que nous appelons LM1 (LM pour Langue Maternelle). Par ailleurs, nous observons que 38 élèves parlent uniquement l'anglais avec leur mère : comme nous ne pouvons les considérer ni francophones, ni allophones, nous les excluons de notre étude, quitte à procéder à des analyses ultérieures dans le cadre d'un autre travail. Finalement, il reste 208 élèves, qui parlent une autre langue avec leur mère que le français ou l'anglais, et qui ont indiqué la parler fréquemment. Ces élèves sont le groupe d'allophones, appelé LM2 dans les analyses à venir.

3.3.3 Description générale des élèves

Nous avons circonscrit les deux groupes d'élèves dont les données sont analysées dans notre étude. Nous décrivons ces élèves en fonction du groupe, de l'année de naissance, du sexe et du niveau scolaire.

Tableau 3-2 Distribution des élèves en fonction de la langue maternelle (LM), du niveau scolaire (1 à 3), de l'étendue des années de naissance et du sexe (filles, garçons).

Groupes	Niveaux scolaires	Années de naissance min. max.	Filles	Garçons	Total
LM1	Sec. 1	1998-2000	20	31	51
	Sec. 2	1997-1999	55	98	153
	Sec. 3	1997-1998	70	107	177
					Total 381
LM2	Sec. 1	1998-1999	11	25	36
	Sec. 2	1994-1999	31	48	79
	Sec. 3	1996-1997	32	61	93
					Total 208
					Total tous élèves 589

Le tableau 3-2 présente 12 groupes d'élèves, distingués par la langue maternelle, le niveau scolaire et le sexe. Au total, nous avons 208 élèves pour qui le français n'est pas la langue maternelle (LM2) et 381 élèves pour qui c'est le cas (LM1). Le plus petit groupe comprend 11 filles de secondaire 1 pour qui le français n'est pas la langue maternelle. Le plus grand groupe comprend 107 garçons francophones de secondaire 3.

3.3.4 Description des élèves allophones (groupe LM2)

Dans la problématique, nous avons déjà évoqué l'hétérogénéité des classes montréalaises sur le plan ethnolinguistique. La description des élèves allophones de notre échantillon qui suit montre à quel point c'est le cas. D'abord, le tableau 3-3 détaille les pays d'origine des élèves allophones.

Tableau 3-3 Pays de naissance des élèves allophones. Nombre de pays différents : 32.

Pays	n	Pays	n	Pays	n	Pays	n
Canada	120	Russie	3	Ukraine	2	Irlande	1
Chine	16	Colombie	2	Albanie	1	Koweït	1
Roumanie	11	Corée Du Sud	2	Argentine	1	Liban	1
Algérie	5	Égypte	2	Arménie	1	Pérou	1
Israël	5	États Unis	2	Émirats Arabes Unis	1	Philippines	1
Brésil	3	Japon	2	Espagne	1	Pologne	1
France	3	Kazakhstan	2	Grèce	1	Suisse	1
Iran	3	Moldavie	2	Haïti	1	Viêt Nam	1

La plupart des élèves sont nés au Canada (n=120). Le lieu de naissance des autres élèves est distribué dans 31 pays différents, les plus importants étant la Chine (n=16), suivi de la Roumanie (n=11). Les langues parlées à la maison reflètent également l'hétérogénéité ethnolinguistique des classes. Le tableau 3-4 dénombre les élèves en fonction de la langue parlée à la maison. (Rappelons que plusieurs langues sont souvent parlées à la maison.)

Tableau 3-4 Langues parlées à la maison chez les élèves allophones. Nombre de langues parlées différentes : 31.

Langue parlée	n	Langue parlée	n	Langue parlée	n	Langue parlée	n
Français	199	Créole	12	Albanais	3	Iranien	1
Anglais	183	Vietnamien	12	Farsi	3	Kabyle	1
Chinois (mandarin, cantonais)	36	Italien	11	Perse	3	Laotien	1
Espagnol	23	Portugais	10	Polonais	3	Libanais	1
Arabe	18	Arménien	5	Hébreux	2	Tamoul	1
Roumain	17	Coréen	5	Japonais	2	Tchèque	1
Russe	16	Allemand	4	Alsacien	1	Turc	1
Grec	14	Tagalog	4	Hongrois	1		

Le français (n=199) et l'anglais (n=183) sont les langues le plus parlées à la maison. Le chinois (« mandarin », « cantonais », « chinois », tels que rapporté par les élèves) suit (n=36), puis l'espagnol (n=23), l'arabe (18), le roumain (n=17), etc. Au total, 33 langues sont parlées à domicile. En ce qui concerne l'écrit, plusieurs élèves déclarent écrire dans d'autres langues que le français et l'anglais. Le tableau 3-5 détaille ces différentes langues.

Tableau 3-5 Langues écrites par les élèves. Nombre de langues différentes: 24.

Langue écrite	n	Langue écrite	n	Langue écrite	n	Langue écrite	n
Espagnol	25	Portugais	8	Albanais	2	Polonais	1
Chinois (mandarin, cantonais)	25	Vietnamien	7	Farsi	2	Tagalog	1
Russe	16	Créole	6	Japonais	2	Laotien	1
Roumain	15	Arabe	5	Allemand	1	Perse	1
Grec	12	Arménien	5	Géorgien	1	Tchèque	1
Italien	9	Coréen	3	Hébreux	1	Turc	1

En plus du français et de l'anglais, ce sont 24 langues qui sont pratiquées à l'écrit parmi les élèves allophones. Si nous comparons les données des tableaux 3-5 et 3-6, nous observons que ce ne sont pas tous les élèves qui écrivent dans la langue maternelle. Par exemple, sur 36 élèves qui parlent l'une ou l'autre variété du chinois, 25 disent l'écrire. Une analyse des résultats aux tests de vocabulaire de notre étude en fonction de variables l'écriture ou non de la langue maternelle à la maison sera réalisée éventuellement. Dans le cadre de ce travail, nous nous contentons de présenter ces données pour décrire notre échantillon. Par ailleurs, nous remarquons que, dans certains cas, plus d'élèves déclarent écrire certaines langues qu'il y en a qui déclarent les parler à la maison. Cela reflète probablement le fait que des élèves suivent ou ont suivi des cours de langue seconde (comme l'espagnol, par exemple).

Pour terminer cette description des élèves allophones, nous présentons le nombre d'années du primaire au cours desquelles ces élèves ont étudié en français. Nous avons décrit l'importance de cette donnée en ce qui concerne la réussite scolaire et la performance en littératie dans la problématique. Le tableau 3-6 détaille le nombre d'élèves ayant étudié en français au primaire.

Tableau 3-6 Nombre d'années du primaire en français chez les élèves allophones.

Nombre d'années	n
0	3
1	5
2	8
3	7
4	11
5	47
6	127
Total :	208

Dans ce tableau, nous observons que plus de la moitié (n=127) ont complété leurs six années du primaire en français. Certains n'ont complété aucune année (n=3); cependant, il est probable qu'ils n'aient simplement pas répondu à cette section du questionnaire (leurs résultats étant aussi élevés que ceux des autres élèves) et que la valeur 0 (par défaut) a été enregistrée dans la base de donnée. Par ailleurs, cinq élèves ont étudié pendant un an en français, huit élèves, deux ans et sept élèves, trois ans.

3.3.5 Déroulement

Au début d'un cours de français, les élèves se présentent à la salle d'ordinateur de la bibliothèque avec leur enseignant. L'expérimentateur est présent à chaque séance. Il accueille les élèves qui s'installent chacun devant un ordinateur. Les élèves qui ne participent pas à l'étude travaillent dans une autre partie de la bibliothèque. Dans la salle avec les participants, l'expérimentateur présente les grandes lignes de son projet. Ensuite, il remercie les élèves et leur rappelle que leur anonymat sera préservé. Il leur rappelle également qu'il s'agit d'une étude, pas d'un examen, et que les résultats n'affectent d'aucune façon leurs résultats scolaires. Finalement, il leur dit que la séance dure de 25 à 45 minutes et qu'il peut clarifier des items des questionnaires à la demande.

La consigne pour se connecter à l'application en ligne étant projetée au mur, chaque élève procède ensuite à la séance proprement dite. Il doit visiter le site (microbe.ca/evocel), cliquer sur le lien « Participer à une étude » puis attendre que la connexion de l'application avec la

base de données se produise. Ensuite, il doit sélectionner son groupe dans une liste, sélectionner son nom dans la liste d'élèves et entrer son mot de passe. Une fois le mot de passe confirmé, l'élève voit la consigne générale de l'application (chaque fenêtre de l'application est représentée dans l'annexe B).

Dans l'ordre, l'élève répond à un questionnaire sociodémographique et linguistique de deux pages, à un questionnaire sur les attitudes et habitudes familiales, au premier test de vocabulaire et au second test de vocabulaire (voir section 3.6, pour une description générale, et l'annexe B, pour le détail du questionnaire sociodémographique et des exemples de fenêtre des tests). Quand l'élève a terminé ce dernier, une fenêtre présentant sa connaissance du vocabulaire dans les différentes disciplines s'affiche. Il peut ensuite fermer la fenêtre, fermer sa session à l'ordinateur puis aller travailler dans une autre section de la bibliothèque.

3.4 Le vocabulaire scolaire

À notre connaissance, il n'existe pas de liste des mots du vocabulaire scolaire québécois du secondaire. Et il n'existe pas d'outil d'évaluation de ce vocabulaire. Il nous a donc fallu d'abord constituer une telle liste puis développer un tel outil. Dans cette section, nous décrivons comment nous avons constitué la liste des mots du vocabulaire scolaire du premier cycle du secondaire. Dans la section 3.6, nous décrivons comment nous avons développé les outils de mesure.

Pour établir la liste des mots et nous assurer qu'ils répondaient bien à notre critère de vocabulaire d'une discipline (voir Section 2.2.2), nous avons numérisé les glossaires et index de manuels scolaires de différentes éditions pour chaque matière scolaire. L'annexe C donne les références de manuels utilisés dans notre étude. Sur les cinq domaines d'apprentissages du premier cycle du secondaire, nous avons retenu cinq disciplines pour lesquelles nous avions à notre disposition plusieurs éditions de manuels : 1. français; 2. géographie; 3. histoire; 4. mathématiques; 5. science et technologie.

Pour identifier les mots propres à chaque discipline, nous avons analysé les mots des fichiers numérisés d'une même discipline à l'aide des outils *Vocabprofile* et *Range* sur le site

Compleat Lexical Tutor (lextutor.ca) et l'outil Lexique (lexique.org). L'outil *Range* permet d'identifier les mots qui sont communs à différents textes. Nous avons comparé les fichiers numérisés des différents manuels d'une même discipline et avons retenu la liste des mots communs à tous les manuels. Parmi ces mots, il y avait des mots de haute fréquence qui n'étaient pas spécifiques à une discipline en particulier. Nous les avons exclus à l'aide de l'outil *Vocabprofile*. Cet outil permet de créer une liste des mots d'un texte par type (c'est-à-dire peu importe combien de fois un mot apparaît dans le texte) et de les regrouper par fréquence. Pour identifier les mots spécifiques à une discipline, nous n'avons retenu que les mots les moins fréquents (au-delà de l'indice de fréquence des 2000 mots les plus fréquents).

Pour mieux distinguer les résultats concernant la connaissance du vocabulaire scolaire par rapport aux autres mots de la langue, nous avons également établi une liste de mots choisis aléatoirement parmi les mots les plus fréquents de la base de données Frantext. Cette base est construite à partir de 218 textes littéraires (des romans) publiés entre 1950 et 2000 (New et Pallier, 2005). Il s'agit d'un corpus de 14,7 millions d'items. Nous avons choisi aléatoirement dix noms parmi les plus fréquents de la base (voir Tableau 3-7 plus loin) pour les besoins de nos tests sur l'étendue du vocabulaire et sur la connaissance en profondeur du vocabulaire. Ces noms ne devaient pas faire partie des noms des disciplines scolaires identifiés préalablement.

Comme nous avons choisi d'étudier la connaissance des noms, nous avons ensuite analysé les listes de mots obtenus par discipline et ceux de la base Frantext à l'aide de l'outil Lexique (New et Pallier, s. d.). Cet outil permet une description linguistique des mots en fonction de divers paramètres : catégorie grammaticale, nombre de lettres, nombre de morphèmes, etc. Cette analyse nous a permis d'identifier les noms dans les listes et les caractéristiques linguistiques de ces noms. Le Tableau 3-7 montre le nombre total de mots et le nombre de noms, pour chaque discipline.

Tableau 3-7 – Nombre de mots total et nombre de noms pour chaque domaine

Domaine	Nombre de mots	Nombre de noms
Français	98	47
Géographie	454	281
Histoire	165	125
Maths	210	139
Sc. technologie	172	113
Très Fréquents	151	72

Considérant le grand nombre de noms dans les listes et considérant qu'il n'est pas possible de mesurer la connaissance de tous ces noms, nous avons dû procéder à un échantillonnage des noms du vocabulaire scolaire du premier cycle du secondaire. Comme une analyse fine des mots de la liste constituerait un travail de thèse en soit et que le but de notre étude est d'abord de vérifier si les élèves allophones développent leur vocabulaire scolaire, nous avons choisi de sélectionner aléatoirement des noms pour constituer les items de nos tests de vocabulaire. Par ailleurs, pour éviter des complications dans l'interprétation de nos résultats dues à la polysémie de mots appartenant à plusieurs disciplines (Hyland et Tse, 2007), nous nous sommes assuré que les noms retenus pour une discipline n'apparaissaient pas dans les listes des autres disciplines.

Nous avons établi à une dizaine de noms par discipline le nombre de mots à retenir pour le test sur l'étendue du vocabulaire et à six noms le nombre de mots à retenir pour le test sur la profondeur de la connaissance (voir la liste à l'annexe C.) Pour déterminer le nombre de mots à retenir, nous avons grossièrement suivi la procédure utilisée par Meara (Meara, 1992), qui sélectionne un échantillon d'un mot pour vingt-cinq (pour chaque catégorie de fréquence ; ce qui fait 40 mots pour mesurer les 1000 mots les plus fréquents, par exemple). Pour notre test sur la profondeur de la connaissance, les rapports de proportion varient donc d'un mot sur huit (6/47) en français à un mot sur quarante-sept (6/281) en géographie, les autres gravitant autour d'un mot sur vingt-cinq.

Pour le test sur l'étendue du vocabulaire, nous avions également besoin de pseudomots (P. Meara, 1992). Nous savons que la reconnaissance de pseudomots peut être facilitée par la

fréquence du mot de base (le mot qui a servi à construire le pseudomot) (Perea, Rosa et Gómez, 2005). Pour ne pas que les élèves identifient facilement les pseudomots (les élèves savent qu'ils travaillent sur des mots du vocabulaire scolaire ; des suites de lettres qu'ils n'ont jamais rencontrées pourraient plus facilement être identifiées comme des pseudomots), nous avons décidé d'utiliser des noms du vocabulaire scolaire. Nous avons d'abord sélectionné aléatoirement des noms dans nos listes de vocabulaire des disciplines scolaires (parmi les mots qui n'étaient pas déjà sélectionnés pour nos tests). Ensuite, nous avons modifié ces mots en remplaçant une consonne ou une voyelle par une autre ou un morphème par un autre (Keuleers et Brysbaert, 2010). La liste des pseudomots apparaît dans l'annexe C.

Dans la section 2.3.7, nous avons rappelé que le degré d'abstraction d'un mot était un facteur important d'apprentissage, parmi d'autres facteurs lexicaux. Nous prenons en compte ce facteur dans notre étude. Pour cette tâche de jugement sur le degré d'abstraction, nous nous sommes inspiré des tâches résumées dans Ferrand (Ferrand, 2001) et de celle développée par Altarriba, Bauer et Benvenuto (Altarriba *et al.*, 1999). Nous avons demandé à quatre personnes ayant étudié en linguistique de juger les mots des tests sur la profondeur de la connaissance. À chacun des juges, nous avons remis un fichier Excel avec la consigne et les mots sur lesquels ils devaient porter leur jugement. Voici la consigne :

Classer chaque mot selon qu'il est plus concret (1) ou plus abstrait (6). Un mot est concret si son référent peut être perçu par un ou plusieurs sens. Un mot est abstrait si ce à quoi il réfère ne peut être défini qu'à l'aide du langage ou de connaissances. Analyser les mots en tenant compte de la matière scolaire. Par exemple, le mot 'angle' en mathématiques est plus concret que le même mot en histoire, de même que le mot 'champ' en français sera plus abstrait que le même mot en géographie. Répondre le plus spontanément possible; ne pas passer plus de dix secondes par mot.

(Voir l'annexe D pour de plus amples détails sur toute la procédure et le calcul de l'accord inter-juge.)

Après avoir recueilli les réponses des linguistes, nous avons calculé l'accord inter-juge. Nous avons d'abord procédé à une analyse à l'aide du coefficient kappa de Cohen. Le coefficient a montré un accord faible. Un examen des données (voir annexe D) et une discussion avec les

différents juges révèlent que ceux-ci n'ont pas interprété la valeur des degrés d'abstraction de la même manière.¹⁸ Nous avons alors vérifié l'accord inter-juge en mesurant la corrélation entre les scores attribués par les quatre différents juges pour les 36 noms. Selon nous, il y aurait accord s'il y avait corrélation positive entre les scores donnés aux noms par les juges et la moyenne des scores, de même que s'il y avait corrélation positive entre les scores donnés par les juges. Le Tableau 3-8 présente l'analyse des scores attribués aux noms par les juges effectuée à l'aide du coefficient de corrélation de Spearman.

Tableau 3-8 Coefficients de corrélation de Spearman pour l'accord inter-juge (quatre juges : J1, J2, J3 et J4) sur le degré d'abstraction des noms.

Nombre de noms = 36

	moyenne	J1	J2	J3	J4
moyenne	1.00000	0.91274	0.86996	0.83047	0.89342
		<.0001***	<.0001***	<.0001***	<.0001***
J1	0.91274	1.00000	0.79040	0.67237	0.79132
		<.0001***		<.0001***	<.0001***
J2	0.86996	0.79040	1.00000	0.59992	0.75840
		<.0001***	<.0001***		<.0001***
J3	0.83047	0.67237	0.59992	1.00000	0.64881
		<.0001***	<.0001***	=0.001**	
J4	0.89342	0.79132	0.75840	0.64881	1.00000
		<.0001***	<.0001***	<.0001***	<.0001***

** p<0,01; *** p<0,0001

L'analyse montre que les degrés d'abstraction attribués par chacun des juges sont corrélés avec la moyenne des degrés d'abstraction (tous les p < 0,001). En d'autres mots, plus un nom est jugé abstrait en moyenne, plus il est susceptible de l'être par chacun des juges. De plus, les degrés d'abstraction attribués par chacun des juges sont corrélés avec chacun des autres juges (tous les p < ou = à 0,001). Ainsi, même si les juges attribuent un degré d'abstraction différent pour un mot donné (par exemple, le nom *comparaison* obtient les scores 5, 3, 3 et 6 et le nom *point* obtient les scores 2, 2, 1 et 4), il y a corrélation positive entre les scores

¹⁸ Comme certains juges étaient moins familiers avec certains domaines, ils étaient plus susceptibles d'attribuer des scores d'abstraction élevés aux mots de ces domaines alors qu'ils pouvaient attribuer des scores moins élevés pour les mots des domaines avec lesquels ils étaient plus familiers.

attribués par les juges d'un mot à l'autre (par exemple, le nom *comparaison* est jugé plus abstrait que le nom *point*).

Comme la longueur du mot et sa fréquence sont des indices lexicaux qui peuvent interagir avec celui du degré d'abstraction (voir Section 2.3.7), nous avons récupéré les données relatives à ces indices pour chacun des mots cibles. Pour le nombre de lettres, nous avons utilisé l'application Lexique (lexique.org). Pour la fréquence, nous avons utilisé les indices de fréquence du logiciel Antidote. Le Tableau 3-9 présente la moyenne des indices lexicaux pour chacun des mots en fonction du domaine d'apprentissages et pour les mots les plus fréquents.

Tableau 3-9 Moyennes des indices lexicaux pour chaque domaine d'apprentissages et pour les mots les plus fréquents

Domaine	Moyenne Abstraction	Moyenne Fréquence	Moyenne Longueur
Français	3,9	59,8	7,0
Géographie	3,3	52,3	6,7
Histoire	4,0	61,5	7,0
Maths	3,3	52,7	8,2
Sc. Tech.	3,7	56,8	8,3
Mots fréquents	2,5	66,8	5,3

Dans ce tableau, nous observons que les mots des domaines d'apprentissages sont en moyenne plus abstraits que les mots les plus fréquents. Parmi les domaines, ce sont les mots en français (3,9), en histoire (4,0) et en science et technologie (3,7) qui sont les plus abstraits. Les mots des domaines scolaires sont également en moyenne moins fréquents et plus longs que les mots les plus fréquents.

3.5 La suite d'outils informatiques ÉVOCEL

Nous avons développé les outils de mesure de notre étude dans un format informatique. Il s'agit d'une suite d'applications que nous avons nommée ÉVOCEL, pour Évaluation du VOCabulaire En Ligne (microbe.ca/evocel). Cette suite comprend une application en Flash

(*Adobe Flash Player*¹⁹) qui, téléchargée côté client, présente une interface avec laquelle interagit le participant (voir l'annexe B ou la section 3.6 pour un aperçu des diverses fenêtres de l'application en Flash). L'application transfère les données recueillies par des scripts PHP (*Hypertext Preprocessor* ; *php.net*) fonctionnant côté serveur. Les scripts PHP enregistrent les données dans une base de données *MySQL* sur le serveur. Une page Web dynamique permet de récupérer les données des sujets et de les télécharger pour pouvoir les analyser dans un ou l'autre logiciel (Excel ou SPSS, par exemple).

3.6 Les outils de mesure

3.6.1 Mesure de l'étendue du vocabulaire

Nous mesurons l'étendue du vocabulaire scolaire des élèves à l'aide d'un test inspiré de ceux développés par Alderson (2005), Cobb (lextutor.ca) et Meara et Buxton (1987). L'élève voit d'abord une fenêtre avec la consigne (voir annexe B). Quand il est prêt, il passe à la fenêtre du test lui-même. Celle-ci présente 16 items (mots ou pseudomots) à la fois. Pour chacun, l'élève doit dire s'il connaît l'item (en cochant « Oui ») ou s'il ne le connaît pas (en cochant « Non »). Il faut avoir coché une réponse vis-à-vis chaque item pour pouvoir passer à la prochaine série. Il y a en tout 80 items, à raison de 10 mots par discipline (français, géographie, histoire, mathématique et science et technologie), 10 mots parmi les plus fréquents (Frantext) et 20 pseudomots, sur 5 pages. L'image 3-1 montre une fenêtre du test sur l'étendue du vocabulaire.

¹⁹ get.adobe.com/fr/flashplayer/

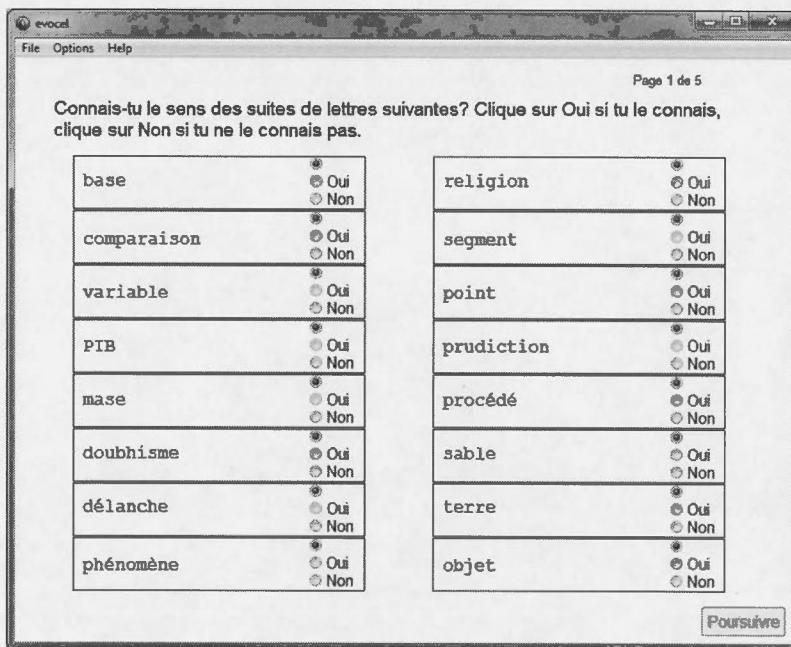


Image 3-1 Exemple de fenêtre du test sur l'étendue du vocabulaire

Le score de l'étendue du vocabulaire se calcule comme suit : un point est attribué à chaque mot pour lequel l'élève a répondu « Oui » et trois points sont soustraits quand un élève a répondu « Oui » à un pseudomot. Ainsi, un élève qui répondrait « Oui » à tous les items obtiendrait un score de zéro. Le score d'un élève varie donc entre -60 (dans le cas improbable où un élève aurait coché « oui » à tous les pseudomots et « non » à tous les mots) et 60 (« oui à tous les mots et « non » à tous les pseudomots). Ce score est converti en pourcentage et indique le pourcentage de mots connus du vocabulaire scolaire du premier cycle et des mots les plus fréquents.

3.6.2 Mesure de la profondeur de la connaissance

3.6.2.1 Développement

La profondeur de la connaissance du vocabulaire est mesurée à l'aide d'un test inspiré du *Depth of Vocabulary Knowledge Measure* (DVK) de Qian (Qian, 1999) et du *Word-associates Test* (WaT) de Read (Read, 2000). Le répondant doit associer des relations

paradigmatiques et syntagmatiques avec chacun des noms. Pour mesurer les relations paradigmatisques, nous avons recherché deux noms qui sont synonymes, hyperonymes ou hyponymes du nom cible tout en étant de la même discipline que celui-ci. Pour la mesure des relations syntagmatiques, nous avons recherché des mots avec lesquels les noms cibles étaient le plus susceptibles d'apparaître en contexte. Nous avons utilisé l'outil de cooccurrences du logiciel Antidote pour ce faire. Quand Antidote n'offrait pas de cooccurrence pertinente (valable dans le contexte de la discipline scolaire), nous avons recherché la cooccurrence dans le corpus de manuels scolaires utilisé pour établir la liste du vocabulaire scolaire. Pour chaque mot cible, l'élève doit choisir deux mots parmi quatre avec lesquels le mot cible partage une partie de son sens (relation paradigmatique) et deux mots parmi quatre avec lesquels le mot cible est le plus susceptible d'apparaître dans une phrase (relation syntagmatique).

En plus des deux bonnes réponses dans la mesure de chacune des relations, nous avons choisi des leurre pour compléter le choix de réponses. Les leurre dans la mesure de la relation paradigmatique étaient des noms de la discipline qui n'étaient ni synonyme, ni hyperonyme, ni hyponyme du nom cible. Les leurre dans la mesure de la relation syntagmatique étaient des mots de la même discipline choisis au hasard mais peu susceptibles de se retrouver fréquemment dans une relation syntagmatique avec le mot cible. Pour ne pas induire les bonnes réponses chez les élèves qui pourraient se servir de leur connaissance de la fréquence des mots, nous nous sommes assuré que la fréquence moyenne des leurre (trouvée dans Antidote) ne se distinguait pas de la fréquence moyenne des réponses.²⁰

Voici deux exemples, un pour les relations paradigmatisques et un autre pour les relations syntagmatiques (voir la liste complète en annexe C) :

²⁰ Par exemple, si les leurre sont plus fréquents que les mots cibles, l'élève pourrait déduire que la bonne réponse n'est pas le leurre, trop fréquent, mais le mot plus rare, sans vraiment connaître le mot cible. L'inverse est aussi vrai, des leurre d'une faible fréquence permettraient de déduire que les mots cibles, plus fréquents, sont la bonne réponse.

- Relations paradigmatisques pour le nom COMPARAISON. Les réponses sont JUGEMENT et RAPPROCLEMENT et les leurres sont RACINE et CAUSE.
- Relations syntagmatiques pour le nom VÉGÉTAUX. Les réponses sont CROISSANCE et CULTIVER et les leurres sont ENTENTE et INDUSTRIEL.

Pour chaque réponse à la mesure de la relation syntagmatique, nous avons pris en note la catégorie grammaticale du mot (verbe, nom ou adjectif) et le type de relation (complément du nom, avec épithète, avec complément nominal, etc.). L'annexe C présente une description des différents items du test et des relations établies avec les mots cibles.

L'application présente un item à la fois, c'est-à-dire les consignes, la discipline, le mot cible, le choix de réponses pour les relations paradigmatisques et le choix de réponses pour les relations syntagmatiques, comme le montre l'image 3-2.

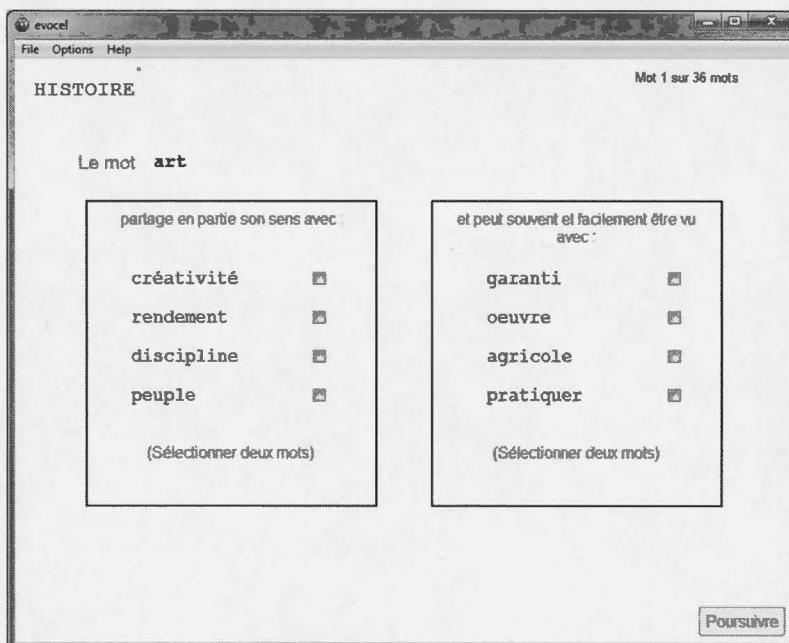


Image 3-2 Exemple d'une fenêtre du test sur la profondeur de la connaissance

L'élève doit choisir deux réponses dans chacune des boîtes. Le bouton « Poursuivre » n'est activé que si c'est le cas. L'application enregistre les réponses de l'élève et le temps en millisecondes mis à répondre à l'item.²¹ Trois données sont donc obtenues pour chaque item :

- le score pour la relation paradigmique. L'élève doit avoir choisi les deux bonnes réponses pour obtenir un point, sinon, il n'en a aucun.
- le score pour la relation syntagmatique. L'élève doit également avoir choisi les deux bonnes réponses pour obtenir le point, sinon il n'en a aucun.
- le temps de réponse en millisecondes.

Nous décrivons le traitement que nous faisons de ces données dans la section 3.7.

3.6.2.2 Sélection des items pour le test sur la profondeur de la connaissance : étude pilote

Dans une première version du test, nous avons préparé 15 items par domaine d'apprentissage. Nous les avons testés auprès de lecteurs experts et très éduqués (c'est-à-dire ayant une scolarité universitaire). Nous voulions nous assurer que les items ne posaient pas de problème à cause d'ambiguités dans les choix de réponses, démontrant ainsi qu'un lecteur expert éduqué avait toutes les connaissances lexicales pour répondre au test avec succès. Nous avions également besoin des commentaires et de suggestions de correction pour améliorer le test (dans les consignes et l'interface) pour sa version définitive. Nous avons fait appel à 16 participants de formation universitaire (baccalauréat ou plus) dont le français est la langue première et à 7 participants de formation universitaire (ayant tous un doctorat) pour qui le français est une langue seconde. En plus d'obtenir divers commentaires pour améliorer l'interface et apporter des corrections aux items qui posaient problème (parce que plus de deux réponses étaient possibles, par exemple), l'étude pilote a permis d'identifier six mots cibles par domaine qui ont été réussis à au moins 95% par les participants pour qui le français

²¹ Nous ne procédons pas à une analyse des temps de réponse dans le cadre de ce travail. Ils seront l'objet d'un travail ultérieur.

est une langue première. Ainsi, nous nous assurons que, pour les items retenus, les lecteurs experts francophones éduqués ont toutes les connaissances lexicales requises pour réussir le test. La version définitive du test pour notre étude porte donc sur 36 noms. Il y a six noms par discipline et six noms parmi les plus fréquents de la base Frantext. La Figure 3-1 montre la moyenne des résultats pour les mots retenus pour chacun des groupes de lecteurs experts.

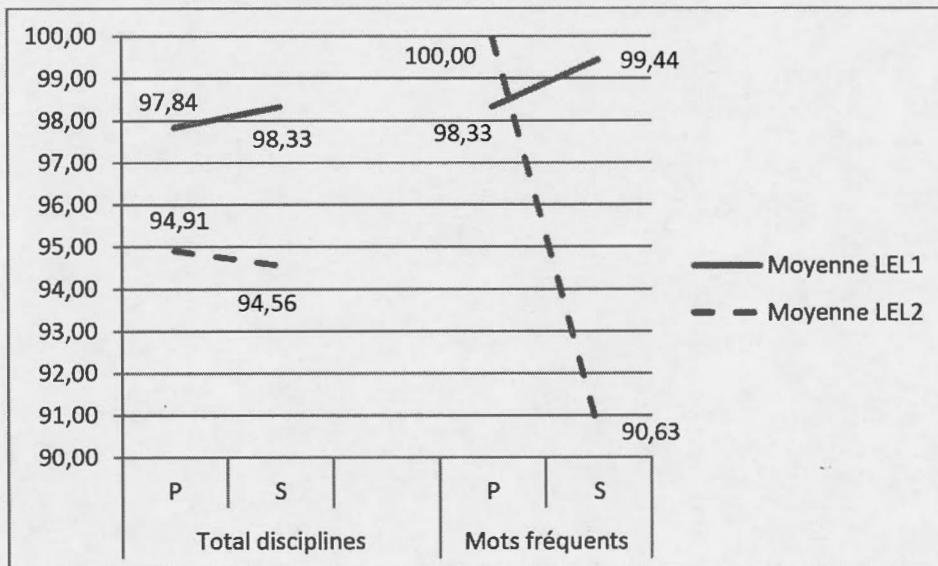


Figure 3-1 - Moyennes des résultats en pourcentage pour les mots retenus en fonction du domaine (disciplines - mots fréquents), des connaissances paradigmatiques (P) et syntagmatiques (S) et du type de locuteur du français (LEL1 = lecteurs experts pour qui le français est une langue première ; LEL2 = lecteurs experts pour qui le français est une langue seconde)

La figure montre que les lecteurs experts éduqués francophones obtiennent en moyenne des résultats très près de 100%, pour les connaissances paradigmatiques ou syntagmatiques, que ce soit pour le vocabulaire scolaire ou les mots les plus fréquents du français. Les lecteurs experts et éduqués pour qui le français est une langue seconde obtiennent des résultats légèrement inférieurs (non vérifiés statistiquement), sauf pour les connaissances paradigmatiques des mots les plus fréquents où ils obtiennent un résultat de 100%. Par contre, leurs résultats pour les connaissances syntagmatiques pour les mots les plus fréquents du français sont nettement inférieurs, suggérant des habitudes de lecture comprenant ces mots un peu moindre que les habitudes de lecture comprenant des mots du vocabulaire scolaire.

3.6.3 Le questionnaire sociodémographique

Le but de ce questionnaire est de recueillir des informations sur l'origine de l'élève et de ses parents ainsi que des informations sur les langues parlées, lues et les années d'étude. Peu de ces données sont utilisées dans le cadre de notre étude, mais elles seront utiles dans le cadre de travaux ultérieurs plus approfondis. Nous avons construit ce questionnaire en nous inspirant des questionnaires développés par Boivin et Pinsonneault (Boivin et Pinsonneault, 2011) et Simard (Simard, 2008). Le tableau suivant montre comment les données relativement à la naissance sont obtenues.

Tableau 3-10 Procédures de cueillette d'informations relativement à la naissance.

Information recherchée	Format	Détail
Date de naissance	Listes déroulantes	Trois listes, une pour le jour, une pour le mois et une pour l'année.
Sexe	Liste déroulante	Femme, homme
Le pays de naissance	Liste déroulante	Tous les pays actuels par ordre alphabétique, sauf le Canada, qui apparaît en premier.
Si né au Canada, la province de naissance	Liste déroulante	Les provinces et territoires, par ordre alphabétique
Si né en dehors du Canada, âge d'arrivée au Canada	Liste déroulante	Avant 2 ans, puis de 3 à 14, puis 15 ans et plus

En ce qui concerne les langues parlées par l'élève, celui-ci peut indiquer jusqu'à trois langues. Pour chacune, il doit spécifier la fréquence avec laquelle il la parle (liste déroulante : Rarement; Une ou deux fois par mois; Chaque semaine) puis avec qui (boîtes à cocher : mère; autres membres de la famille; amis; à l'école). L'Image 3-3 montre la fenêtre de cette partie du questionnaire.

Questionnaire

Ta date de naissance : Jour Mois Année Sexe : Sélectionner...

Le pays où tu es né(e) :

Est-ce que ta mère est née au Canada ? ?

Est-ce que ton père est né au Canada ? ?

Quelle(s) langue(s) parles-tu ? (En écrire au moins une.) Précise la fréquence et avec qui tu parles cette langue.

1. [Text input] Sélectionner fréquence... Ma mère D'autres membres de ma famille
 Ami(e)s À l'école

2. [Text input] Sélectionner fréquence... Ma mère D'autres membres de ma famille
 Ami(e)s À l'école

3. [Text input] Sélectionner fréquence... Ma mère D'autres membres de ma famille
 Ami(e)s À l'école

Poursuivre

Image 3-3 Interface pour que l'élève indique quelles langues il parle et à quelle fréquence

Pour les besoins de notre étude, la langue parlée avec la mère est considérée comme la langue maternelle de l'élève. Cette décision repose sur le fait qu'il est difficile d'établir la langue première des élèves allophones. Plusieurs langues sont souvent parlées à la maison. L'indice le plus fiable est la langue parlée fréquemment avec la mère. Si seul le français est parlé avec la mère, l'élève est considéré comme francophone. Si une langue autre que le français ou l'anglais est parlée avec la mère et qu'elle est parlée fréquemment, l'élève est considéré comme allophone. Si l'élève déclare parler l'anglais avec sa mère mais aucune langue autre que le français, il est considéré anglophone. Dans ce dernier cas, nous ne présentons pas les résultats de ces élèves dans notre étude.

Des informations sur les langues lues sont également demandées. L'élève peut indiquer jusqu'à trois langues. Pour chacune des langues, il doit spécifier la fréquence (Rarement; Une ou deux fois par mois; Chaque semaine).

L'élève dont la langue maternelle n'est pas le français doit indiquer s'il a suivi des cours de français en classe d'accueil ou en classe de francisation, des cours de langue d'origine ou des cours dans d'autres langues. Finalement, il doit indiquer les années au cours desquelles il a étudié en français, de la maternelle à la deuxième secondaire. L'Image 3-4 montre la fenêtre de cette deuxième partie du questionnaire.

Image 3-4 Interface pour que l'élève indique dans quelles langues il lit et à quelle fréquence, de même que sa scolarité en français au primaire et au secondaire

3.7 Traitements des données, objectifs, variables et analyses

3.7.1 Objectif de recherche 1 : évaluer l'étendue du vocabulaire scolaire d'élèves allophones du secondaire

Nous avons identifié 705 noms dans le corpus tiré des manuels scolaires. Nous en avons extrait un échantillon de 60, à raison de 10 par discipline (50 mots) puis un échantillon de 10 parmi les plus fréquents de la base Frantext. Le résultat au test sur l'étendue du vocabulaire est sur 60 : l'élève obtient un point s'il répond « Oui » à un mot, mais il en perd trois s'il

répond « Oui » à un pseudomot. Comme il y a 60 mots et 20 pseudomots, le score maximum qu'il est possible d'obtenir est 60 et le minimum, 0 (jusqu'à -60, en fait, mais nous changeons les résultats négatifs par le résultat 0). Le résultat sur 60 est ensuite transformé en pourcentage. La transformation indique le pourcentage de noms du vocabulaire scolaire de premier cycle du secondaire et des mots fréquents connus par les élèves (Vermeer, 2001).

Nous voulons vérifier si les élèves allophones connaissent autant de mots du vocabulaire scolaire et de mots fréquents que les élèves francophones. Pour ce faire, nous comparons les résultats des élèves allophones à ceux des élèves francophones. Nous prenons également en compte les variables contrôle (VC) suivantes : le niveau scolaire (secondaire 1 à 3) et le sexe puisque les résultats des filles sont souvent plus élevés que ceux des garçons à l'école et que les scores en général devraient s'améliorer de la première secondaire à la troisième. Nous vérifions d'abord s'il existe une différence entre l'étendue du vocabulaire des élèves allophones et celle des élèves francophones à l'aide d'un test t. Le cas échéant, nous procédons à une ANOVA en prenant en considération les variables contrôles.

3.7.2 Objectif de recherche 2 : évaluer la profondeur de la connaissance du vocabulaire scolaire d'élèves allophones du secondaire, c'est-à-dire plus précisément évaluer les connaissances paradigmatisques et syntagmatiques de noms des domaines d'apprentissages du premier cycle du secondaire.

Avec la mesure de la profondeur de la connaissance du vocabulaire, nous obtenons des résultats sur la connaissance du sens du mot cible (relation paradigmatique) et de sa compréhension en contexte (relation syntagmatique), par domaine de connaissance. Chaque résultat est calculé comme suit : un point est donné si les deux bonnes réponses sont choisies par l'élève, sinon, il n'y a pas de point. La raison pour laquelle nous n'attribuons pas de point pour le choix d'une bonne réponse sur deux est que les probabilités de choisir une bonne réponse sont de 50 %. Les probabilités d'en choisir deux sont de 25 %. Nous préférons

éliminer le plus possible l'obtention de points de façon simplement aléatoire.²² Par ailleurs, le temps de réponse en millisecondes pour chaque mot cible est enregistré (mais nous ne procérons pas à l'analyse de cette donnée dans le cadre de ce travail). Le tableau 3-11 présente les variables dépendantes obtenues avec le test.

Tableau 3-11 Compilation des résultats en fonction du domaine des mots et des variables dépendantes

Domaines	Nombre de mots cibles	Variables dépendantes	
Français	6	Score connaissance du sens par discipline (sur 6)	Score compréhension en contexte par discipline (sur 6)
Géographie	6		
Histoire	6		
Mathématiques	6		
Sciences et techniques	6		
Tous les domaines	30		
Mots fréquents	6	Score c. du sens (sur 30)	Score c. en contexte (sur 30)
		Score c. du sens (sur 6)	Score c. en contexte (sur 6)

Le tableau 3-11 montre que les résultats seront compilés par item (mot cible), par domaines d'apprentissages (et les mots les plus fréquents) et pour l'ensemble des domaines (somme des résultats de chaque discipline, sauf les mots les plus fréquents).

Nous vérifions si les élèves allophones ont un même niveau de connaissance du vocabulaire scolaire que les élèves francophones à l'aide d'un test t. Nous comparons les résultats des élèves allophones à ceux des élèves francophones pour les variables dépendantes de la connaissance du sens et de l'utilisation en contexte. Si la différence est statistique, nous comparons les résultats en fonction de la variable indépendante et de variables contrôles. La variable indépendante est la langue maternelle (VI). Nous prenons également en compte les variables contrôles (VC) suivantes : le niveau scolaire (secondaire 1 à 3) et le sexe puisque les résultats des filles sont souvent plus élevés que ceux des garçons à l'école et que les scores en général devraient s'améliorer de la première secondaire à la troisième. Pour chaque

²² Un problème qui a été soulevé par Read (2000) et Qian (1999).

analyse entre les variables VI et VC et une des variables dépendantes, le test utilisé est une analyse de variance (ANOVA). Le tableau qui suit résume ces analyses.

Tableau 3-12 Les variables contrôle (VC), indépendantes (VI) et dépendantes prises en considération dans les analyse de variance pour l'objectif de recherche 2

VC	VC	VI	Variables dépendantes		
Niveau	Sexe	Langue maternelle	étendue	connaissance du sens	compréhension en contexte
Secondaire 1	Féminin	Français L1			
Secondaire 2	Masculin	allophone			
Secondaire 3					

Si nous observons des différences statistiques, nous vérifions si ces différences sont maintenues par domaine d'apprentissages, en procédant avec des ANOVA.

3.7.3 Objectif de recherche 3 : Mesurer la relation entre le degré d'abstraction des mots et la profondeur de la connaissance du vocabulaire des élèves

La facilité avec laquelle un mot peut être appris est affectée, entre autres, par sa longueur, sa fréquence et son degré d'abstraction (Ferrand, 2001). Dans le cadre de notre étude, l'effet du degré d'abstraction nous intéresse particulièrement, puisque les mots du vocabulaire scolaire sont estimés plus abstraits que les mots plus courants. À travers ce quatrième objectif, nous voulons explorer l'influence de ce facteur sur la profondeur de la connaissance des élèves allophones et francophones tout en contrôlant les effets d'autres facteurs lexicaux reconnus comme le nombre de morphèmes dans le mot, sa longueur et sa fréquence. Ainsi, pour chaque mot cible, le score obtenu sera soit « 1 » (deux bons choix de réponse) ou « 0 » (un ou aucun bon choix de réponse). En conséquence, nous avons recours à une modélisation logistique, dans laquelle nous cherchons à prédire la probabilité d'obtenir un « 1 » en fonction de différents facteurs (les facteurs lexicaux) (Bressoux, 2010). De plus, comme plusieurs résultats sont obtenus d'un même élève, nous devons prendre en compte que les résultats proviennent d'une mesure répétée. Pour procéder à l'analyse, nous avons recours à une modélisation linéaire appelée « Équations d'estimation généralisées » (en anglais, *Generalized Estimating Equations* ou GEE). Cette modélisation est appropriée pour la vérification de notre objectif puisqu'elle permet en effet l'analyse de mesures répétées ou

d'autres observations corrélées, comme les données agglutinées (Carrière et Bouyer, 2002); SPSS, aide en ligne), quand les variables sont, entre autre, dichotomiques.

Dans la prochaine partie, nous décrivons et analysons les données obtenues à la lumière des objectifs et procédures d'analyses précisés ici.

CHAPITRE IV

ANALYSE DES RÉSULTATS

4.1 Introduction

Dans cette partie, nous présentons les diverses analyses qui permettent d'atteindre les objectifs de notre étude. Nous avons fait ces analyses à l'aide de la version 20 (française) du logiciel SPSS (IBM) et certains calculs et figures à l'aide du logiciel Excel (Microsoft). Pour les objectifs 1 et 2, mesures de l'étendue et de la profondeur de la connaissance du vocabulaire, nous présentons les statistiques descriptives. Nous vérifions la normalité des distributions et l'homogénéité des variances. Quand les distributions et les variances le permettent, nous comparons ensuite les moyennes entre les groupes, à l'aide de test t ou d'ANOVA. Pour mesurer l'ampleur des effets, nous avons recours au R carré ou à l'eta carré. En ce qui concerne l'objectif 3, relatif aux relations entre les variables lexicales et la profondeur de la connaissance du vocabulaire, nous illustrons la distribution des résultats en fonction des différents facteurs lexicaux. Ensuite, nous analysons les données à l'aide d'équations d'estimation généralisées, en insérant dans les modèles différents facteurs. Nous décrivons le modèle le plus vraisemblable. Nous terminons cette partie en produisant une synthèse des résultats les plus importants.

4.2 Objectif de recherche 1 : évaluer l'étendue du vocabulaire scolaire des élèves allophones

Nous présentons dans cette section les résultats des élèves au test sur l'étendue du vocabulaire. Nous cherchons à savoir si les résultats des élèves allophones se distinguent ou non de ceux des élèves francophones. Nous voulons également connaître les effets éventuels de facteurs comme le niveau de scolarité ou le sexe. Au cours du test sur l'étendue du

vocabulaire, les élèves devaient indiquer s'ils connaissaient l'item présenté en cochant « oui » ou « non ». Le test présentait 60 mots et 20 pseudomots. Un élève a pu obtenir un score entre -60 (s'il a répondu « oui » à tous les pseudomots et « non » à tous les mots); notons qu'un tel résultat est peu probable) et 60 (s'il a répondu « oui » à tous les mots et « non » à tous les pseudomots). Plus le score est élevé, plus grande est l'étendue du vocabulaire de l'élève. Pour mieux représenter cette étendue, nous avons transformé le résultat des élèves en pourcentage. Par exemple, nous estimons qu'un élève qui obtient un score de 60 % reconnaît, dans le cadre de notre étude, 60 % des noms du vocabulaire scolaire du premier cycle du secondaire et des mots les plus fréquents du français (Read, 2000; Vermeer, 2001).

Après avoir transformé les résultats en pourcentage, nous avons calculé les différentes statistiques descriptives et vérifié la normalité de la distribution de ces résultats. Le détail de cette analyse est exposé à l'annexe E. Nous présentons ici un résumé de cette analyse. Le Tableau 4-1 montre les différentes statistiques descriptives des résultats des élèves en fonction de la langue maternelle. Pour chaque niveau de la variable, il y a deux lignes. La première présente les statistiques descriptives pour tous les participants. La seconde présente les statistiques descriptives après que nous ayons retiré les valeurs extrêmes (voir l'annexe E pour comprendre comment nous justifions le retrait de ces valeurs).

Tableau 4-1 Statistiques descriptives des moyennes des résultats en pourcentage au test sur l'étendue du vocabulaire, en fonction de la langue maternelle (LM1= élèves francophones; LM2=élèves allophones).

	n	Moyenne (%)	Erreur standard	Médiane	Variance	Écart-type	Min.	Max.	Asym.	Aplat.
LM1	381	86.54	.77	90.0	225.46	15.02	0	100.0	-2.060	5.825
	321	91.75	.41	93.30	52.97	7.28	58.3	100.0	-.994	.858
LM2	208	87.99	1.00	93.30	208.74	14.45	23.3	100.0	-2.278	5.541
	187	92.02	.50	93.30	46.78	6.84	71.7	100.0	-.961	.373

Première ligne (en gris) : tous les élèves; deuxième ligne : sans les valeurs extrêmes.

Dans ce tableau, nous observons que les élèves francophones (LM1) connaissent en moyenne 91.75 % du vocabulaire scolaire, comparativement à 92.02 % chez les élèves allophones (LM2). Pour les deux groupes, la variance est relativement grande, les scores variant de 58.3

à 100 % pour le groupe LM1 et de 71.7 à 100 % pour le groupe LM2. La Figure 4-1 suivante illustre le profil des résultats pour les deux groupes.

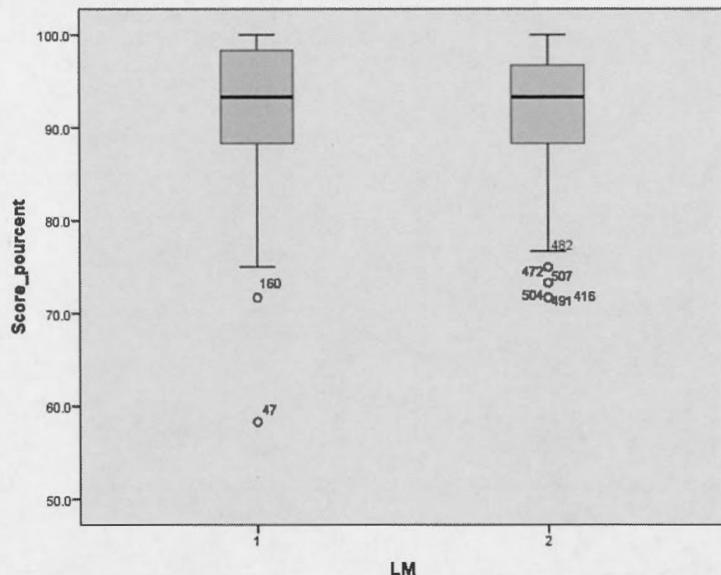


Figure 4-1 Diagramme en boîte des profils des résultats (scores en pourcentage) au test sur l'étendue du vocabulaire des élèves francophone (LM1) et allophones (LM2).

À la lumière de ce diagramme et des diagrammes tiges et feuilles à l'annexe E, nous constatons que la plupart des élèves obtiennent des scores élevés, autour ou au-dessus de 90 %. Nous n'avons pas de distribution statistiquement normale pour aucun des groupes : LM1, $D(321)=.152$, $p<0,001$; LM2, $D(187)=.179$, $p<0,001$ (voir détails à l'annexe E). Les résultats suggèrent un plafonnement dans la performance à ce test. Par ailleurs, le calcul de l'homogénéité de la variance de Levene basé sur la moyenne montre que les variances des deux groupes sont égales, $F(1,506)=.815$, ns. Comme la moyenne des résultats et leur distribution sont semblables, nous ne procéderons pas à une comparaison des moyennes, la différence éventuelle entre les groupes étant peu importante.

Nous poursuivons la description des résultats des élèves au test sur l'étendue du vocabulaire en fonction de la langue maternelle, du niveau scolaire et du sexe. Ce sont douze groupes que nous obtenons après avoir réparti les élèves selon ces facteurs. Les valeurs extrêmes sont

exclues. Le Tableau 4-2 détaille les statistiques descriptives des résultats pour ces douze groupes.

Tableau 4-2 Statistiques descriptives des moyennes des résultats en pourcentage au test sur l'étendue du vocabulaire, en fonction de la langue maternelle (LM1= élèves francophones; LM2=élèves allophones), du niveau scolaire (N1, N2, N3) et du sexe (F, M).

LM	Niv.	S	n	Moy.	Err. Stand.	Méd.	Var.	Écart-type	Min.	Max.	Asym.	Aplat.	
1	1	f	18	90.37	2.34	92.50	98.39	9.92	58.3	100.0	-2.204	5.984	
		m	27	90.36	1.27	93.30	43.73	6.61	75.0	100.0	-.700	-.598	
	2	f	53	92.92	.93	95.00	46.16	6.79	80.0	100.0	-.710	-.731	
		m	76	91.90	.81	93.30	49.48	7.03	71.7	100.0	-1.058	.733	
	3	f	63	92.35	.81	93.30	41.03	6.41	75.0	100.0	-.618	-.307	
		m	84	91.17	.87	95.0	63.52	7.98	75.0	100.0	-.743	-.588	
	2	1	f	11	89.99	1.89	93.3	39.32	6.27	76.7	96.7	-1.377	1.087
			m	19	89.21	1.68	91.70	53.70	7.33	71.7	98.3	-1.086	.432
		2	f	29	91.38	1.28	95.0	47.24	6.87	80.0	100.0	-.487	-1.214
			m	43	91.24	1.08	93.30	50.0	7.07	71.7	100.0	-1.091	1.021
		3	f	29	93.79	1.17	95.0	39.93	6.32	78.3	100.0	-1.103	.356
			m	56	93.39	.88	95.0	43.13	6.57	73.3	100.0	-1.165	1.192

Abréviations : n=nombre; Moy.=moyenne; Err. Stand.=erreur standard; Méd.=médiane; Var.=variance; Min.=résultat le plus petit; Max.=résultat le plus grand; Asym.=asymétrie de la pente; Aplat.=aplatissement de la pente

Dans ce tableau, nous observons que les élèves obtiennent en moyenne des résultats très proches : de 89,21 % pour les élèves allophones de sexe masculin de première secondaire à 93,79 % pour les élèves allophones de sexe féminin de troisième secondaire. Aucune des distributions n'est normale (voir annexe E pour la description des résultats aux tests de normalité), ce qui s'explique en partie par des résultats très élevés, suggérant un plafonnement des élèves au test sur l'étendue. La Figure 4-2 montre le profil des résultats pour les douze groupes.

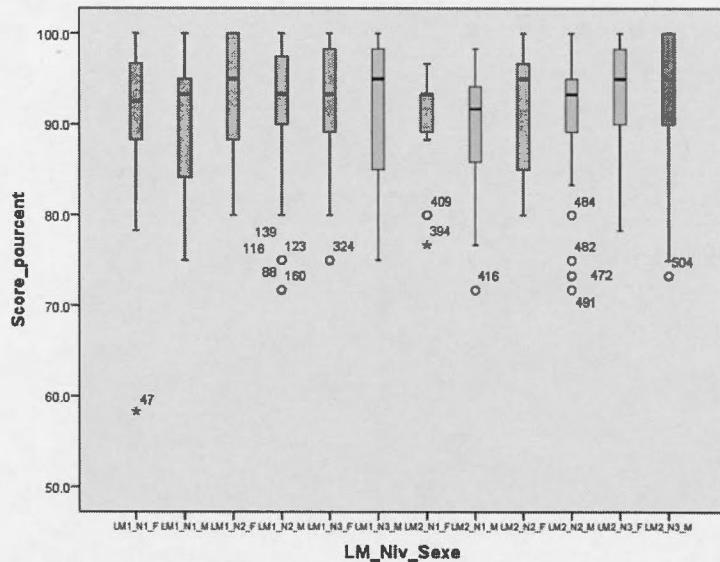


Figure 4-2 Diagramme en boîte des profils des résultats au test sur l'étendue du vocabulaire des élèves en fonction de la langue maternelle (LM1, LM2), du niveau scolaire (N1, N2, N3) et du sexe (F, M).

La figure illustre le score élevé de tous les élèves et la ressemblance entre les groupes. D'ailleurs, le calcul de l'homogénéité de la variance de Levene basé sur la moyenne montre que les variances sont comparables, $F(11,496)=.866$, ns. Considérant le plafonnement des élèves au test sur l'étendue du vocabulaire et le peu de différence entre les groupes, nous ne procéderons pas à une comparaison entre moyennes.

Pour conclure cette section sur l'analyse des résultats au test sur l'étendue du vocabulaire scolaire, nous constatons que les élèves de notre étude reconnaissent la plupart des mots du vocabulaire scolaire et des mots les plus fréquents du français (autour de 90 % des mots), peu importe la langue maternelle, le niveau, le sexe ou la scolarité en français au primaire. L'homogénéité de la variance entre les groupes LM1 (élèves francophones) et LM2 (élèves allophones) suggère, avec des moyennes semblables, que les résultats sont équivalents. Par reconnaissance, nous entendons la capacité à distinguer un mot du français au moins sur le plan de la forme écrite (voir Cadre théorique, section 2.3.8). Ces résultats ne montrent pas jusqu'à quel point chacun des mots est connu. Cette profondeur de la connaissance des mots chez les élèves est analysée dans les sections qui suivent.

4.3 Objectif de recherche 2 : évaluer la profondeur de la connaissance du vocabulaire scolaire des élèves allophones

Dans cette section, nous analysons les résultats des élèves au test sur la profondeur de la connaissance du vocabulaire. Nous cherchons à savoir si les élèves allophones ont développé une connaissance en profondeur comparable ou non à celles d'élèves francophones au cours de leur scolarité. Nous avons retenu deux dimensions de cette profondeur dans le cadre de notre étude : l'établissement de relations paradigmatisques avec les mots cibles par association du mot avec des synonymes, hyperonymes ou hyponymes (dimension appelée plus brièvement « SENS » dans les analyses qui suivent); et l'établissement de relations syntagmatiques avec les mots cibles, par traitement ou reconnaissance de cooccurrences fréquentes et pertinentes (dimension appelée plus brièvement « CONTEXTE » ou « CONT » dans les analyses qui suivent). Le test présentait successivement 36 mots aux participants, à raison de 6 mots par domaine d'apprentissages (cinq domaines) et 6 mots pour les mots les plus fréquents de la base Frantext. Pour chaque mot, un participant pouvait obtenir un score de 1 pour la dimension SENS et 1 pour la dimension CONTEXTE s'il choisissait les deux bonnes réponses pour chaque dimension. Nous avons calculé le total des points obtenus pour chacune des dimensions pour l'ensemble des domaines, pour chaque domaine et pour les mots les plus fréquents (6 points par dimension par domaine, donc). Nous avons transformé ces scores en pourcentage. Les analyses portent sur ces pourcentages. L'annexe F décrit l'analyse des statistiques descriptives dans le détail.

Nous présentons ici un résumé des informations les plus pertinentes et concentrons notre propos sur la description des résultats sans les valeurs extrêmes : nous avons retiré des analyses les résultats des élèves dont les valeurs étaient extrêmes par rapport à la moyenne, tel que suggéré par les sorties SPSS. Nous avons procédé ainsi pour chacune des analyses par domaine et non une bonne fois pour toute : ainsi, les résultats de certains élèves qui sont retirés pour un domaine donné et une dimension donnée ne le sont pas nécessairement pour un autre domaine ou une autre dimension. Nous justifions notre choix par le fait qu'un élève n'obtient pas nécessairement des résultats comparables d'un domaine à l'autre. Par exemple, un élève peut avoir un résultat extrême dans la dimension CONTEXTE des mots en français,

mais pas dans la dimension SENS. Un autre peut démontrer des résultats inverses en géographie ou d'un domaine à l'autre. Seuls les résultats extrêmes de tels élèves seront retirés de l'analyse. Nous conservons ainsi le plus de données possible, sans prendre en considération des difficultés particulières éprouvées par un élève pour l'une ou l'autre dimension dans l'un ou l'autre domaine.²³

Dans les sous-sections qui suivent, nous décrivons d'abord les résultats généraux pour la dimension SENS, tous domaines d'apprentissage confondus, en fonction de la langue maternelle. Dans la mesure où nous observons des différences significatives entre les deux populations, nous procédons à des analyses en prenant en considération, en plus de la langue maternelle, les variables Niveau scolaire et Sexe, pour l'ensemble des domaines d'apprentissages puis chacun d'entre eux. Nous procédons ensuite de même pour la dimension CONTEXTE. Nous terminons en présentant des tableaux et figures synthèses illustrant les résultats les plus importants à la lumière de l'objectif de recherche.

4.3.1 Analyse des résultats, pour l'ensemble des domaines, selon la langue maternelle, pour la dimension SENS

Nous analysons les résultats pour l'ensemble des domaines pour la dimension SENS selon la langue maternelle (LM1=élèves francophones; LM2=élèves allophones). Dans le Tableau 4-3, nous présentons les statistiques descriptives des résultats des élèves pour la dimension SENS en fonction de la langue maternelle pour l'ensemble des domaines. Les lignes grisées montrent les données brutes, les lignes claires montrent les données sans les valeurs extrêmes.

²³ En d'autres mots, un élève peut avoir bien performé dans un domaine, mais pas dans un autre, peu importent les raisons. Nous n'avons pas de raison d'exclure ses résultats dans tous les domaines s'ils ne sont pas des valeurs extrêmes.

Tableau 4-3 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable SENS au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM1=élèves francophones; LM2=élèves allophones).

	n	Moyenne	Erreurs standard	Médiane	Variance	Écart-type	Min.	Max.	Asym.	Aplat.
LM1	381	86.00	.58	86.67	130.03	11.43	13.33	100.0	-2.466	10.221
	370	87.32	.42	90.0	66.79	8.17	60.0	100.0	-.771	.512
LM2	208	83.03	.79	86.67	130.18	11.41	30.0	100.0	-1.507	4.240
	203	84.04	.66	86.67	87.12	9.33	60.0	100.0	-.499	-.378

En gris, les données brutes; en blanc, les données dans les valeurs extrêmes.

Quand les valeurs extrêmes sont retirées, nous observons que les élèves francophones obtiennent en moyenne 87,32 % et les élèves allophones, 84.04 %, pour la dimension SENS dans l'ensemble des domaines. Le test de normalité ne montre pas de distribution normale, LM1, D (370) = .140, $p<0.001$; LM2, D(203) = .128, $p<0.001$. Cependant, la symétrie des pentes (LM1 = -.771; LM2=-.499), leur aplatissement (LM1 = .512; LM2=-.378), la forme des diagrammes tiges et feuilles et les valeurs des résidus observés dans les normogrammes (voir annexe F-B) suggèrent des distributions relativement normales. Nous procédons à une comparaison des moyennes à l'aide du test t pour échantillons indépendants. Le test de Levene sur l'égalité des variances montre que les variances ne sont pas équivalentes, $F(1, 571) = 6.903$, $p< 0.01$. Nous nous référions en conséquence au résultat du test t pour l'hypothèse de variances inégales. Le test montre une différence statistique entre les moyennes des élèves francophones (87.32 %) et allophones (84.04 %), $t(371.659)=4.195$, $p<0.001$. Le calcul de l'ampleur de l'effet donne un résultat ($\eta^2 = 0.21$) suggérant un petit effet, expliquant environ 4 % de la variance (voir annexe F-B-f).

Comme la différence entre les moyennes des deux populations est statistique, nous poursuivrons notre analyse de la variable SENS en fonction des variables Niveau scolaire et SEXE.

4.3.2 Analyse de la dimension SENS, tous domaines, selon la langue maternelle, le niveau scolaire et le sexe

Les résultats des deux populations étant différents pour la dimension SENS de la profondeur de la connaissance du vocabulaire, nous vérifions si des variables comme le niveau scolaire

(N., Niv. ou Niveau) et le sexe (Sexe ou S) interviennent dans les résultats. Comme nous prenons en compte trois variables indépendantes (langue maternelle, niveau scolaire, sexe), nous procédons à des analyses de variance à trois facteurs, si la normalité des distributions et l'homogénéité des variances le permettent. Nous analysons d'abord les résultats pour l'ensemble des domaines. Dans les sous-sections suivantes, nous décrivons les résultats pour chacun des domaines. Pour alléger les tableaux, nous ne présentons que les données sans les valeurs extrêmes (voir annexe F pour le détail des explorations). De plus, pour alléger le texte, nous résumons les tableaux des statistiques descriptives globalement, avec surtout le souci de décrire les distributions. Nous rappelons qu'une synthèse des résultats est présentée à la section 4.10.

Statistiques descriptives

Le Tableau 4-4 montre les statistiques décrivant les résultats obtenus par les élèves pour la variable SENS dans tous les domaines en fonction des variables langue maternelle, niveau scolaire et sexe.

Tableau 4-4 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable SENS au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).

LM	Niv.	S	n	Moy.	Err. Stand.	Méd.	Var.	Écart- type	Min.	Max.	Asym.	Aplat.
1	1	f	20	81.33	2.21	81.67	97.54	9.88	60.0	93.3	-.624	-.325
		m	31	84.52	1.68	86.67	87.07	9.33	66.67	100.0	-.254	-.257
	2	f	54	88.65	1.0	90.0	53.46	7.31	70.0	100.0	-.391	-.548
		m	94	87.13	.84	90.0	65.85	8.12	60.0	100.0	-.882	.964
		f	70	87.43	.90	86.67	56.09	7.49	66.67	100.0	-.452	-.021
		m	101	88.75	.78	90.0	60.86	7.80	63.33	100.0	-1.086	1.375
	2	f	11	78.18	3.09	80.0	105.25	10.26	63.33	93.33	-.248	-.974
		m	24	78.06	2.16	76.67	112.0	10.58	60.0	100.0	.516	-.353
		f	31	83.87	1.89	86.67	111.18	10.54	60.0	100.0	-.570	-.029
		m	46	84.78	1.28	86.67	75.88	8.71	63.33	96.67	-.673	-.192
		f	31	86.24	1.22	86.67	46.476	6.82	73.33	96.67	-.289	-.659
		m	60	85.89	1.09	86.67	71.325	8.45	63.33	100.0	-.537	-.308

Dans le tableau, nous observons des moyennes allant de 78,06 % pour les élèves allophones de sexe masculin de secondaire 1 à 88,75 % pour les élèves francophones de sexe masculin de secondaire 3. Les variances sont grandes, les résultats variant entre 60 et 100 % dans plusieurs groupes. Le calcul de la normalité des distributions montre que cinq distributions sur douze sont normales (voir annexe F). Les pentes sont relativement symétriques, à l'exception des élèves francophones de sexe masculin de secondaire 3 (-1,086).

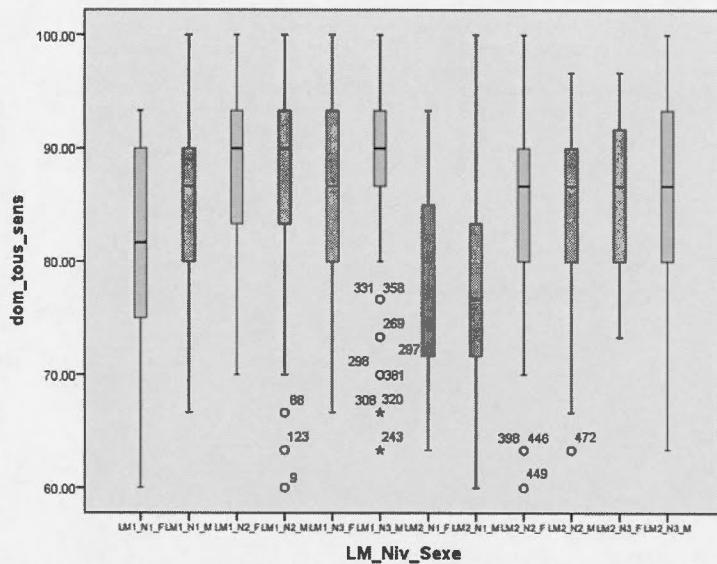


Figure 4-3 Diagramme en boîte des profils des résultats pour la variable SENS, tous domaines, en fonction de la langue maternelle (LM1, LM2), du niveau scolaire (N1, N2, N3) et du sexe (F, M).

Les observations qui précèdent, le diagramme ci-haut et l'examen des diagrammes à l'annexe F suggèrent des distributions relativement normales.

Analyse de variance

Considérant la relative normalité des distributions, nous avons procédé à une analyse de variance à trois facteurs. Nous voulons savoir si la moyenne pour la dimension SENS au test de profondeur de la connaissance du vocabulaire varie en fonction de la langue maternelle, du niveau scolaire et du sexe des élèves. Nous précisons que, dans toutes les analyses de variances, nous utilisons le modèle de la somme des carrés de type II (Larson-Hall, 2010). Dans les analyses de comparaisons multiples (*post hoc*), nous utilisons la correction de Bonferroni, qui a plus de pouvoir quand le nombre de moyennes à comparer est limité (Field, 2009; Larson-Hall, 2010). Dans notre étude, la variable Niveau de scolarité a trois niveaux, ce qui donne trois moyennes à comparer dans les tests *post hoc*. Cependant, considérant le nombre inégal d'élèves d'un groupe à l'autre, nous utilisons également la correction GT2 de

Hochberg, développée pour des tailles d'échantillon différentes (Field, 2000).²⁴ Le détail des analyses de variance se trouve à l'annexe G.

Le test d'égalité des variances des erreurs de Levene révèle que les variances des erreurs de la variable dépendante sont égales sur les différents groupes, $D(11, 561) = 1.377$, ns (voir annexe G-A). Nous poursuivons avec l'analyse de variance. Elle examine les effets du niveau de scolarité, de la langue maternelle et du sexe sur le pourcentage de réussite pour la variable SENS au test sur la profondeur de la connaissance. Des effets sont trouvés pour les variables Niveau de scolarité ($D_{2,566} = 17.501$, $p < 0,001$, Eta carré partiel = 0,059) et Langue maternelle ($D_{1,566} = 18.320$, $p < 0,001$, Eta carré partiel = 0,032). Le test *post hoc* de Bonferroni montre une différence statistique entre les résultats des élèves de secondaire 1 et 2 ($p < 0,001$), 1 et 3 ($p < 0,001$), mais pas 2 et 3 (ns). Le test GT2 propose la même interprétation (voir annexe G-A). Les résultats des élèves de secondaire 1 sont plus faibles statistiquement que ceux des élèves de secondaire 2 et 3. Les résultats des élèves allophones sont plus faibles statistiquement que ceux des élèves francophones. Il n'y a pas d'effet de la variable Sexe ($D_{1,566} = .266$, ns, Eta carré partiel = 0,000). L'ampleur des effets est petite (respectivement 6 % et 3 %). Il n'y a pas d'interaction entre les variables (voir annexe G). Le modèle explique $R^2 = 10\%$ de la variance dans les résultats pour la variable SENS.

La Figure 4-4 illustre les résultats en fonction du niveau et de la langue maternelle (mais pas en fonction du Sexe, cette variable n'ayant pas d'effet statistique).²⁵

²⁴ Field (2009) rappelle que le GT2 est fiable seulement si les variances sont homogènes. Sinon, il est possible d'utiliser le test de Gabriel, mais celui-ci augmente la probabilité d'une erreur de type II si la différence entre les échantillons est trop grande.

²⁵ La figure montre les coordonnées des moyennes marginales calculées par SPSS. Celles-ci sont utilisées quand le nombre de participants est inégal entre les groupes. Les analyses de variance effectuées par SPSS sont basées sur ces moyennes marginales. Avec la figure, nous voulons illustrer les profils des groupes pour mieux les comparer. Dans le texte, nous reproduisons les moyennes des statistiques descriptives, qui peuvent différer légèrement des moyennes marginales.

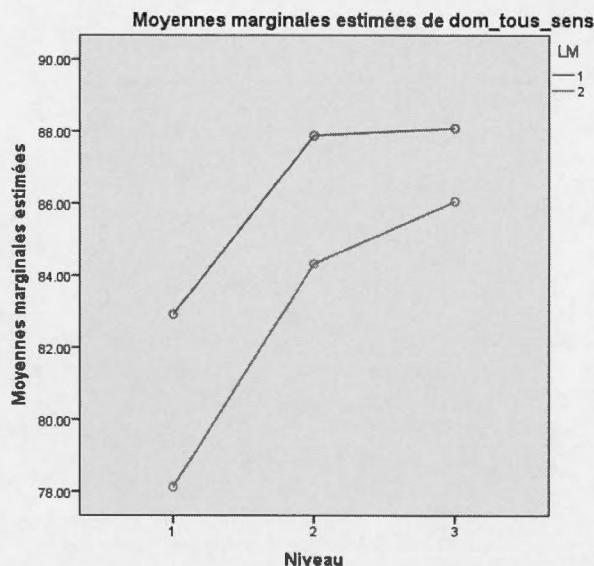


Figure 4-4 Illustration des moyennes des résultats pour la variable SENS en fonction des deux effets statistiques Niveau de scolarité (1 à 3) et Langue maternelle (1=francophones; 2=allophones).

La figure montre deux profils semblables entre les niveaux de scolarité chez les élèves francophones et les élèves allophones, les résultats de ces derniers étant cependant systématiquement plus faibles que ceux des élèves francophones. Les résultats des élèves de secondaire 1 se distinguent de ceux des élèves de secondaire 2 et 3. Les résultats entre les élèves de secondaire 2 et 3 ne se distinguent pas statistiquement. Nous remarquons que l'écart entre les élèves allophones et francophones s'amoindrit en secondaire 2 puis en secondaire 3, sans que les élèves allophones ne rattrapent les élèves francophones.

4.3.3 Analyse de la dimension SENS pour chacun des domaines d'apprentissages et les mots les plus fréquents selon la langue maternelle, le niveau de scolarité et le sexe

Après avoir vérifié s'il y avait une différence entre les résultats des groupes pour tous les domaines, nous examinons les données pour chaque domaine et pour les mots les plus fréquents. Comme les données sont nombreuses, nous ne présentons que les informations les plus pertinentes. Le détail des sorties SPSS peut être étudié dans les annexes 6 et 7.

Statistiques descriptives

Le Tableau 4-5 décrit les moyennes des résultats obtenus par les élèves pour la dimension SENS pour chacun des domaines d'apprentissages et les mots fréquents du français en fonction des variables Langue maternelle, Niveau scolaire et Sexe (nous rappelons que le nombre de sujets varie d'un groupe et d'un domaine à l'autre; voir le détail dans l'annexe F).

Tableau 4-5 Moyennes des résultats en pourcentage pour la variable SENS au test sur la profondeur du vocabulaire pour chacun des domaines et pour les mots fréquents, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).

			Français	Géographie	Histoire	Maths	Science et techno.	Mot fréquents
LM	Niv.	S	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne
1	1	f	90.35	75.00	77.50	88.60	79.17	98.33
		m	86.67	87.22	78.49	90.86	82.26	97.31
	2	f	89.20	90.57	84.29	93.33	89.42	96.36
		m	87.63	88.83	83.68	92.67	86.52	95.41
	3	f	91.19	90.05	84.54	94.05	80.92	97.62
		m	90.07	90.71	87.42	91.83	86.33	96.26
2	1	f	86.67	75.76	75.76	78.33	80.30	93.94
		m	87.88	78.79	70.14	88.66	78.47	92.00
	2	f	88.33	85.00	82.80	89.66	79.03	96.77
		m	86.23	90.91	77.78	88.30	80.21	96.53
	3	f	92.47	87.78	84.41	91.15	77.08	98.44
		m	90.48	84.17	86.26	92.90	80.28	96.99

Nous observons des moyennes variant de 70.14 % pour les garçons allophones de secondaire 1 en histoire à 98.44 % pour les filles allophones de secondaire 3 pour les mots les plus fréquents. En général, les moyennes sont élevées et les variances sont grandes. Le calcul de la normalité des distributions montre que peu de distributions sont normales (voir annexe F). Les écarts entre les moyennes et les médianes sont élevés. Les pentes sont relativement symétriques. Par contre, les résultats des élèves suggèrent un plafonnement à la dimension SENS de la profondeur de la connaissance pour les mots du français, de la géographie, des mathématiques et pour les mots les plus fréquents. Le test d'égalité des variances des erreurs

de Levene révèle que les variances des erreurs de la variable dépendante par domaine sont inégales (voir annexe G). Nous ne poursuivons pas avec une analyse de variance en fonction des domaines.

Par contre, l'observation des profils des résultats par domaine selon le niveau et la langue maternelle dans la Figure 4-5 révèlent des informations susceptibles d'enrichir la discussion.

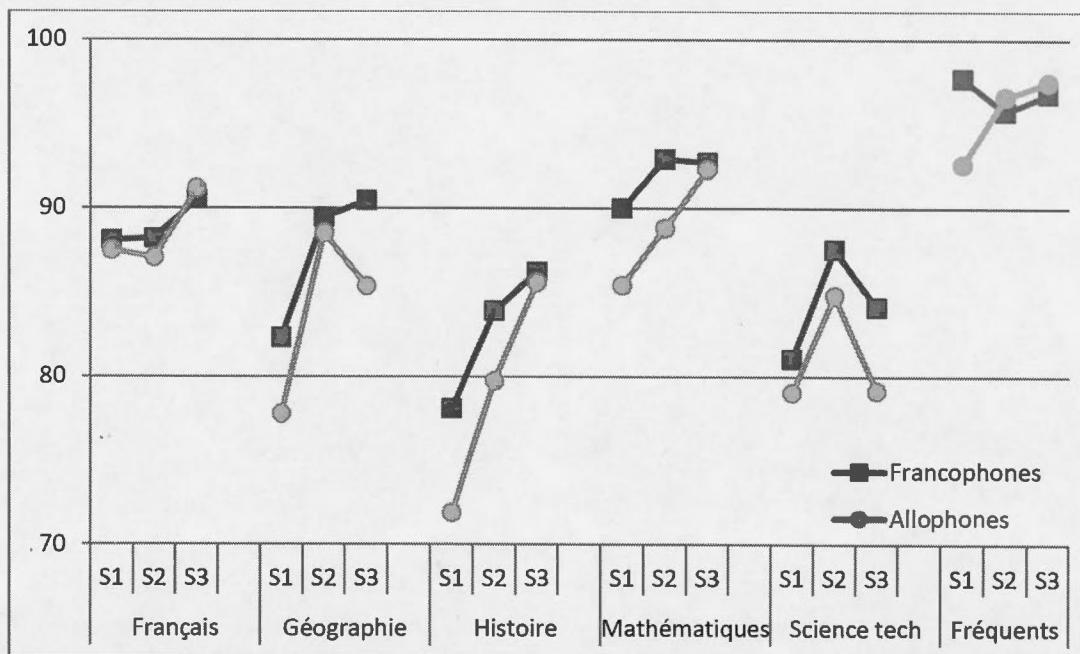


Figure 4-5 Illustration des profils des résultats des élèves pour la variable SENS (connaissances paradigmatiques) en fonction de la langue maternelle, du niveau scolaire et des domaines d'apprentissages.

Les connaissances paradigmatiques des mots des domaines français, mathématiques et, davantage, les mots les plus fréquents semblent maîtrisés d'emblée par la plupart des élèves. Rappelons que le français et les mathématiques sont des cours que les élèves suivent à tous les ans. En histoire, il semble y avoir une progression linéaire pour les élèves allophones, qui rejoignent les élèves francophones. Le phénomène semble se répéter en géographie et en science et technologie, mais les élèves allophones de troisième secondaire performent moins bien que les élèves de secondaire 2. C'est également le cas pour les élèves francophones en

science et technologie. Rappelons que les cours de géographie et de science et technologie ne sont donnés qu'en secondaire 2. Nous développons ce sujet dans la discussion.

Après avoir présenté notre analyse des résultats pour la dimension SENS, nous procémons à celles de la dimension CONTEXTE.

4.3.4 Analyse des résultats, pour l'ensemble des domaines, selon la langue maternelle, pour la dimension CONTEXTE

Nous analysons les résultats pour l'ensemble des domaines pour la dimension CONTEXTE selon la langue maternelle (LM1=élèves francophones; LM2=élèves allophones). Le Tableau 4-6 montre les statistiques descriptives des résultats des élèves pour la variable CONTEXTE en fonction de la langue maternelle pour l'ensemble des domaines d'apprentissages.

Tableau 4-6 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable CONTEXTE au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM1=élèves francophones; LM2=élèves allophones).

	n	Moyenne	Erreurs standard	Médiane	Variance	Écart-type	Min.	Max.	Asym.	Aplat.
LM1	381	68.92	.68	70.0	177.12	13.31	10.0	93.33	-1.033	2.023
	371	70.06	.59	70.0	129.10	11.36	40.0	93.33	-.358	-.343
LM2	208	65.63	.92	66.67	174.27	13.20	20.0	96.67	-.539	.359
	207	65.86	.89	66.67	164.96	12.84	30.0	96.67	-.420	.005

En gris, les données brutes; en blanc, les données sans les valeurs extrêmes.

Sans les valeurs extrêmes, nous observons que les élèves francophones obtiennent en moyenne 70,06 % et les élèves allophones, 64,86 %, dans la dimension CONTEXTE pour l'ensemble des domaines. Les résultats pour cette dimension sont nettement plus faibles que pour la dimension SENS, chez les deux populations. Le test de normalité ne montre pas de distribution normale, LM1, $D(371) = .107$, $p < 0,001$; LM2, $D(207) = .089$, $p < 0,01$. Cependant, le peu de différence entre la moyenne et la médiane, la symétrie des pentes (LM1 = -.358; LM2 = -.420), leur aplatissement (LM1 = -.343; LM2 = .005), les formes des diagrammes tiges et feuilles et les valeurs des résidus observés dans les normogrammes (voir annexe F) suggèrent des distributions normales. Nous procémons à une comparaison des

moyennes à l'aide du test t pour échantillons indépendants. Le test de Levene sur l'égalité des variances montre que les variances sont équivalentes, $F(1, 576) = 3.125$, ns. Nous nous référerons en conséquence au résultat du test t pour l'hypothèse de variances égales. Le test montre une différence statistique entre les moyennes des élèves francophones (70.06 %) et allophones (65.85 %), $t(576)=4.081$, $p<0,001$. Les résultats des élèves allophones sont statistiquement plus faibles que ceux des élèves francophones. Le calcul de l'ampleur de l'effet donne un résultat ($r = 0,17$) suggérant un petit effet, expliquant environ 3 % de la variance (voir annexe F).

Comme la différence entre les moyennes des deux populations est statistique, nous poursuivrons notre analyse de la dimension CONTEXTE en fonction des variables Niveau scolaire et Sexe.

4.3.5 Analyse de la dimension CONTEXTE, tous domaines, selon la langue maternelle, le niveau et le sexe

Comme les résultats des deux populations sont différents pour la dimension CONTEXTE de la profondeur de la connaissance du vocabulaire, nous vérifions si des variables comme le niveau scolaire (N., Niv. ou Niveau) et le sexe (Sexe ou S) interviennent dans les résultats. Comme nous prenons en compte trois variables indépendantes (Langue maternelle, Niveau scolaire, Sexe), nous procédons à des analyses de variance à trois facteurs, si la normalité des distributions et l'homogénéité des variances le permet. Nous analysons d'abord les résultats pour l'ensemble des domaines. Dans la sous-section suivante, nous poursuivons décrivons les résultats pour chacun des domaines.

Statistiques descriptives

Le Tableau 4-7 montre les statistiques décrivant les résultats obtenus par les élèves pour la variable CONTEXTE pour l'ensemble des domaines en fonction des variables langue maternelle, niveau scolaire et sexe.

Tableau 4-7 Statistiques descriptives des moyennes des résultats en pourcentage pour la variable CONTEXTE au test sur la profondeur du vocabulaire pour l'ensemble des domaines d'apprentissages, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).

LM	Niv.	S	n	Moy.	Err. Stand.	Méd.	Var.	Écart- type	Min.	Max.	Asym.	Aplat.	
1	1	f	20	61.17	2.47	60.0	121.96	11.04	40.0	80.0	-.217	-.578	
			31	61.72	2.02	60.0	126.57	11.25	43.33	90.0	.477	-.060	
	2	f	51	74.51	1.11	73.33	62.59	7.91	60.0	93.33	.130	.100	
			95	68.63	1.11	70.0	116.19	10.78	46.67	90.0	-.321	-.604	
	3	f	68	73.43	1.27	73.33	110.44	10.51	46.67	90.0	-.461	.024	
			101	72.41	1.04	73.33	109.14	10.45	43.33	93.33	-.231	-.479	
	2	1	f	11	53.64	3.44	56.67	129.90	11.34	33.33	63.33	-1.085	-0.46
			m	25	52.53	2.60	50.0	168.78	12.99	30.0	86.67	.747	.933
		2	f	31	66.24	1.82	66.67	102.77	10.138	43.33	80.0	-.751	-.119
			m	46	66.16	1.46	66.67	98.26	9.91	43.33	90.0	-.164	-.167
		3	f	32	71.46	1.84	73.33	108.92	10.44	46.67	96.67	-.206	.370
			m	60	71.28	1.38	71.67	114.91	10.72	46.67	93.33	-.034	-.780

Nous observons dans ce tableau des moyennes allant de 52,53 % pour les élèves allophones de sexe masculin de secondaire 1 à 74,51 % pour les élèves francophones de sexe féminin de secondaire 2. En moyenne, les résultats sont nettement plus faibles que les résultats pour la dimension SENS. Les variances sont très grandes, les résultats variant entre 30 et 100 %. Le calcul de la normalité des distributions montre que cinq distributions sur douze sont normales (voir annexe F). Les écarts entre les moyennes et les médianes sont peu élevés. Les pentes sont relativement symétriques. Pour mieux illustrer les résultats des groupes d'élèves, nous présentons la Figure 4-6 montrant le profil de ces résultats pour chacun des groupes du Tableau 4-7.

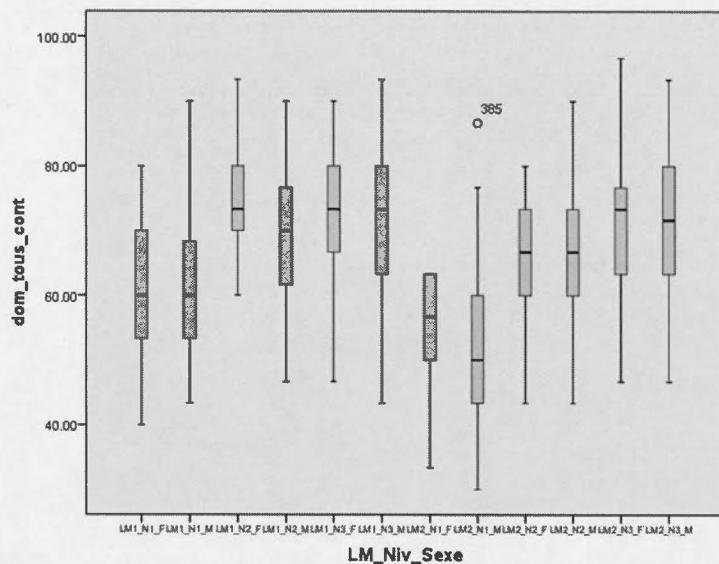


Figure 4-6 Diagramme en boite des profils des résultats pour la variable CONTEXTE, tous domaines, en fonction de la langue maternelle (LM1, LM2), du niveau scolaire (N1, N2, N3) et du sexe (F, M).

Les observations ci-dessus, la Figure 4-6 et l'examen des diagrammes en annexe F nous font conclure à une normalité des distributions. Nous poursuivons avec le test sur l'homogénéité des variances.

Analyse de variance

Le test d'égalité des variances des erreurs de Levene révèle que les variances des erreurs de la variable dépendante sont égales sur les différents groupes, $D(11, 559) = 1.091$, ns. Nous poursuivons avec l'analyse de variance. Il s'agit d'une analyse de variance à trois facteurs. Nous voulons savoir si la moyenne pour la dimension CONTEXTE au test de profondeur de la connaissance du vocabulaire varie en fonction de la langue maternelle, du niveau scolaire et du sexe des élèves. Nous avons recours aux tests post hoc de Bonferroni et au GT2 de Hochberg. Le détail de cette analyse se trouve à l'annexe G.

Des effets sont trouvés pour les variables Niveau de scolarité ($D_{2,559} = 59,009$, $p < 0,001$, Eta carré partiel = 0,174), Langue maternelle ($D_{1,559} = 17,557$, $p < 0,001$, Eta carré partiel = 0,030)

et Sexe ($D_{1,559} = 4,052$, $p < 0,05$, Eta carré partiel = 0,007). L'ampleur des effets est petite (respectivement 17 %, 3 % et moins de 1 %). Il y a interaction entre les variables Niveau et Langue maternelle ($D_{2,559} = 3,946$, $p < 0,05$, Eta carré partiel = 0,014) avec un petit effet (1 %). Le modèle explique $R^2 = 22$ % de la variance dans les résultats pour la variable CONTEXTE. Le test post hoc de Bonferroni montre une différence statistique dans la moyenne des résultats entre tous les niveaux de scolarité (secondaire 1 et 2, $p < 0,001$; secondaire 1 et 3, $p < 0,001$; secondaire 2 et 3, $p < 0,01$). Le test GT2 propose la même interprétation.

La Figure 4-7 illustre les résultats en fonction du niveau et de la langue maternelle pour la variable CONTEXTE, tous domaines d'apprentissages.

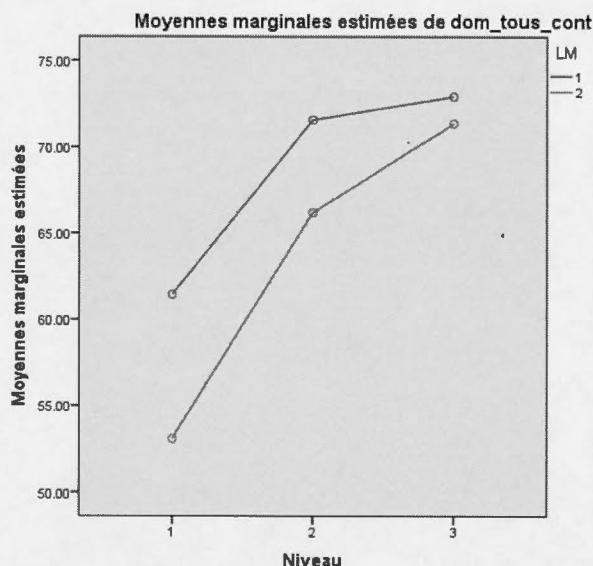


Figure 4-7 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Niveau de scolarité (1 à 3) et Langue maternelle (1=francophones; 2=allophones).

La figure montre des profils des élèves semblables selon le niveau de scolarité. Cependant, les élèves allophones obtiennent des résultats plus faibles en secondaire 1 et 2. Bien que les moyennes se distinguent généralement d'un niveau de scolarité à l'autre, l'interaction statistique entre les variables Niveau de scolarité et Langue maternelle suggère que les moyennes des élèves francophones et celles des élèves allophones de secondaire 3 sont

équivalentes. En d'autres mots, il n'y a plus d'écart entre les deux populations en secondaire 3.

Un effet Sexe a été montré. Pour mieux illustrer l'effet de cette variable sur la dimension CONTEXTE, nous présentons les figures 4-8 à 4-10 qui montrent les résultats en fonction du niveau de scolarité et du sexe.

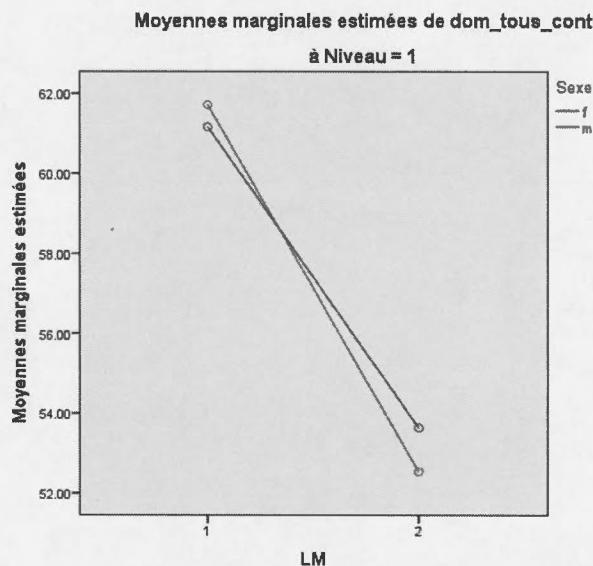


Figure 4-8 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 1.

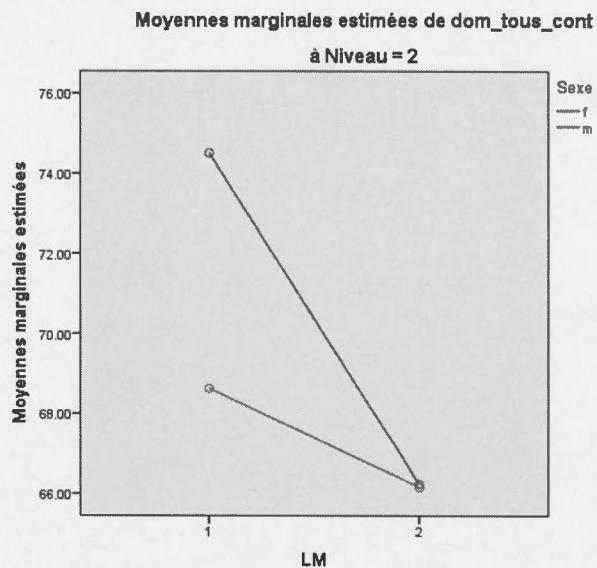


Figure 4-9 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 2.

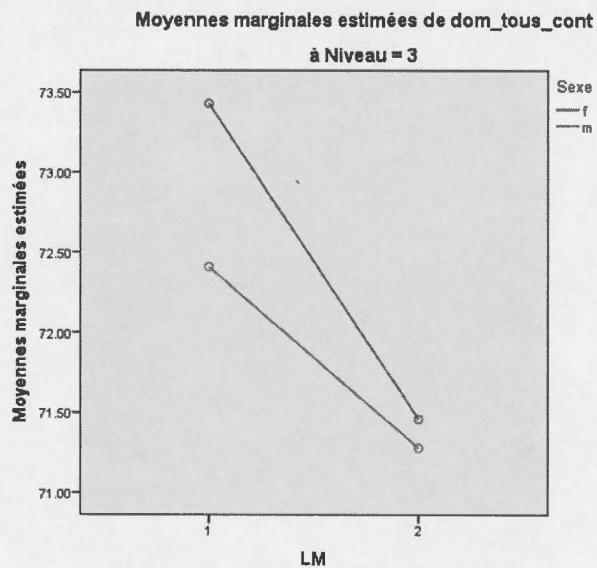


Figure 4-10 Illustration des moyennes des résultats pour la variable CONTEXTE en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 3.

La Figure 4-8 illustre l'écart entre la moyenne des élèves francophones et celle des élèves allophones de secondaire 1, mais suggère des résultats semblables entre les filles et les garçons de chaque population : la moyenne des filles francophones est de 61,17 %, celle des garçons, 61,72 %; la moyenne des filles allophones est de 53,64 %, celles des garçons, 52,53 %. Par contre, la Figure 4-9 montre un profil des populations différent. En secondaire 2, les moyennes des filles (74,51 %) et des garçons (68,63 %) francophones se distinguent, alors que les moyennes des filles (66,24 %) et des garçons (66,16 %) allophones restent semblables. Le même phénomène peut être remarqué à la Figure 4-10, où les moyennes des filles (73,43 %) et des garçons (72,40 %) francophones se distinguent, mais pas les moyennes des filles (71,46 %) et des garçons (71,28 %) allophones. La différence pour la variable Sexe reflète donc une différence entre les garçons et les filles francophones de secondaire 2 et 3 alors qu'il n'y a pas de différence entre les garçons et les filles allophones.

L'analyse qui précède montre des profils généraux de la dimension CONTEXTE pour l'ensemble des domaines d'apprentissages. Dans la section qui suit, nous procérons aux analyses par domaine, toujours en prenant en considération la langue maternelle, le niveau de scolarité et le sexe des élèves.

4.3.6 Analyse de la dimension CONTEXTE pour chacun des domaines d'apprentissages et les mots les plus fréquents selon la langue maternelle, le niveau de scolarité et le sexe

Après avoir vérifié s'il y avait une différence entre les résultats des groupes pour tous les domaines, nous examinons les données pour chaque domaine et pour les mots les plus fréquents. Comme les données sont nombreuses, nous ne présentons que les informations les plus pertinentes. Le détail des sorties SPSS peut être étudié dans les annexes 6 et 7.

Statistiques descriptives

Le Tableau 4-8 décrit les moyennes des résultats obtenus par les élèves pour la dimension CONTEXTE pour chacun des domaines d'apprentissages et pour les mots fréquents en fonction des variables Langue maternelle, Niveau scolaire et Sexe (le nombre de sujets varie d'un groupe et d'un domaine à l'autre; voir le détail dans l'annexe F).

Tableau 4-8 Statistiques descriptives des moyennes des résultats en pourcentage pour la dimension CONTEXTE au test sur la profondeur du vocabulaire pour chacun des domaines et pour les mots fréquents du français, en fonction de la langue maternelle (LM; 1=francophones; 2=allophones), du niveau scolaire (Niv.; 1 à 3) et du sexe (S; f, m).

			Français	Géographie	Histoire	Maths	Science et technico.	Mot fréquents
LM	Niv.	S	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne	Moyenne
1	1	f	44.17	75.44	57.41	70.00	68.42	75.83
		m	42.47	76.19	58.02	68.28	70.43	77.78
	2	f	54.24	88.78	67.65	86.54	70.61	79.81
		m	46.43	85.13	63.12	81.21	65.99	77.36
	3	f	49.29	86.72	72.14	88.06	70.95	76.81
		m	50.31	87.91	69.42	85.35	65.73	78.43
2	1	f	41.67	63.64	43.94	63.64	59.10	69.70
		m	36.00	57.02	50.0	60.42	57.33	68.67
	2	f	50.54	76.11	62.78	79.57	62.90	72.22
		m	39.24	78.72	68.15	79.51	61.46	71.63
	3	f	53.65	81.25	65.56	86.46	70.83	77.01
		m	48.91	84.20	67.50	84.44	72.68	75.42

Nous observons des moyennes variant de 36.00 % pour les garçons allophones de secondaire 1 en français à 88.78 % pour les filles francophones de secondaire 2 en géographie. Les moyennes sont faibles en français et en histoire, modérées dans les autres domaines et pour les mots les plus fréquents. Les variances sont grandes. Le calcul de la normalité des distributions montre que peu de distributions sont normales (voir annexe F). Les écarts entre les moyennes et les médianes sont élevés. Les pentes sont relativement symétriques. Par contre, les résultats des élèves suggèrent un plafonnement à la dimension CONTEXTE pour les domaines géographie et mathématiques. Pour les autres domaines, les distributions sont relativement normales et nous permettent de procéder à des analyses de variance.

Domaine Français

Pour le domaine Français, le test d'égalité des variances des erreurs de Levene révèle que les variances des erreurs de la variable dépendante sont égales sur les différents groupes, $D(11, 576) = .534$, ns (voir annexe G). Nous poursuivons avec l'analyse de variance. Elle examine

les effets du niveau de scolarité, de la langue maternelle et du sexe sur le pourcentage de réussite pour la variable CONTEXTE pour le domaine Français. Des effets sont trouvés pour les variables Niveau de scolarité ($D_{2,576} = 7,488$, $p<0,01$, Eta carré partiel = 0,025) et Sexe ($D_{1,576} = 7,730$, $p<0,01$, Eta carré partiel = 0,013). Il n'y a pas d'effet statistique pour la variable Langue maternelle, quoiqu'une tendance peut être observée ($D_{1,576} = 2,964$, $p=0,086$, Eta carré partiel = 0,005). L'ampleur des effets est petite (respectivement 3 %, 1 % et moins de 1 %). Il n'y a pas d'interaction statistique entre les variables mais une tendance se dessine entre Niveau et Sexe ($D_{2,576} = 2,798$, $p=0,062$, Eta carré partiel = 0,01) avec un petit effet (1 %). Le modèle explique $R^2 = 6$ % de la variance dans les résultats pour la variable CONTEXTE pour le domaine Français. Le test post hoc de Bonferroni montre une différence statistique dans la moyenne des résultats entre les niveaux de scolarité 1 et 2 ($p<0,05$), 1 et 3 ($p<0,001$), mais pas 2 et 3 (ns). Le test GT2 propose la même interprétation (voir annexe G-D).

La Figure 4-11 illustre les moyennes en fonction du niveau de scolarité et de la langue maternelle pour la variable CONTEXTE en français.

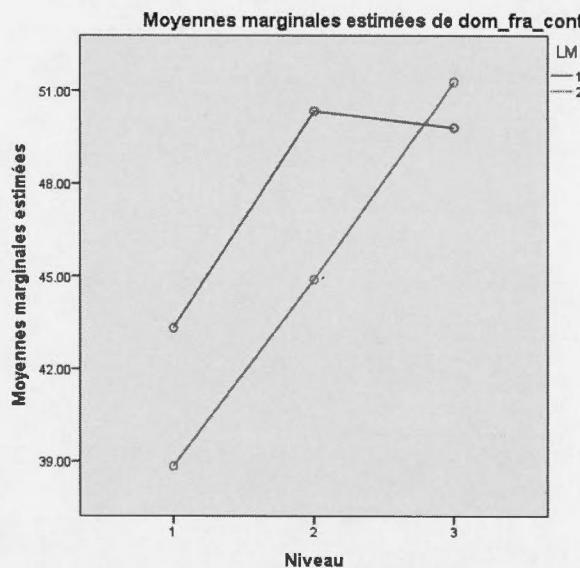


Figure 4-11 Illustration des moyennes des résultats pour la variable CONTEXTE en français en fonction des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2).

La figure illustre la progression d'une année de scolarité à l'autre, particulièrement chez les élèves allophones (37,62 % en sec. 1; 43,67 % en sec. 2 et 50,54 % en sec. 3). L'effet du niveau de scolarité est moins important chez les élèves francophones. Leurs résultats sont différents entre la première et la deuxième secondaire (respectivement 43,14 % et 49,24 %), mais pas entre la deuxième et la troisième secondaire (49,91 %). Les élèves allophones rattrapent les élèves francophones en secondaire 3, ce qui explique en partie le faible effet de la variable Langue maternelle.

Pour l'effet de la variable Sexe, les résultats des garçons sont plus faibles que ceux des filles, à l'exception des garçons francophones de secondaire 3, qui obtiennent des résultats semblables à ceux des filles. Ceci explique le petit effet d'interaction qui se dessine entre la variable Sexe et la variable Niveau.

Domaine Histoire

À la lumière des observations des données de l'annexe F, les distributions des différents groupes varient en termes de normalité. Plusieurs ne sont pas normales, en particulier celles des résultats des élèves francophones de sexe féminin de secondaire 2 et des résultats des élèves allophones de sexe féminin de secondaire 3. Par contre, le test d'égalité des variances des erreurs de Levene révèle que les variances des erreurs de la variable dépendante sont égales sur les différents groupes, $D(11, 549) = 1,681$, ns (voir annexe G). Nous poursuivons donc avec l'analyse de variance. Elle examine les effets du niveau de scolarité, de la langue maternelle et du sexe sur le pourcentage de réussite pour la variable CONTEXTE pour le domaine Histoire. Des effets sont trouvés pour les variables Niveau de scolarité ($D_{2,549} = 44,535$, $p < 0,001$, Eta carré partiel = 0,140) et Langue maternelle ($D_{1,549} = 5,571$, $p < 0,05$, Eta carré partiel = 0,010). Il n'y a pas d'effet statistique pour la variable Sexe ($D_{1,549} = 0,237$, ns). L'ampleur des effets est petite (respectivement 14 %, 1 %). Il y a une interaction statistique entre les variables Niveau et Langue maternelle ($D_{2,549} = 5,800$, $p < 0,01$, Eta carré partiel = 0,021) avec un petit effet (2 %) de même qu'entre les variables Langue maternelle et Sexe ($D_{2,549} = 8,320$, $p < 0,01$, Eta carré partiel = 0,015), avec un petit effet (2 %). Le modèle explique $R^2 = 18\%$ de la variance dans les résultats pour la variable CONTEXTE pour le domaine Histoire. Les comparaisons multiples avec le test post hoc de Bonferroni montrent une différence statistique dans la moyenne des résultats entre tous les niveaux de scolarité (1 et 2, $p < 0,001$; 1 et 3, $p < 0,001$; 2 et 3, $p < 0,01$). Le test GT2 propose la même interprétation (voir annexe G).

La Figure 4-12 illustre les effets des variables Niveau de scolarité et Langue maternelle et leur interaction sur la variable CONTEXTE en histoire.

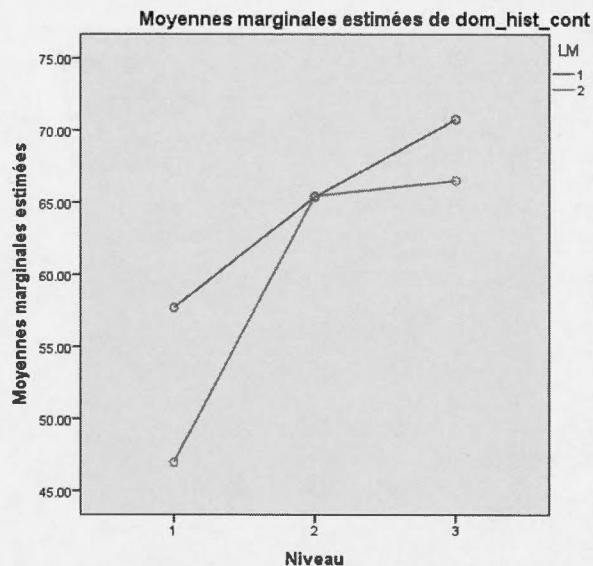


Figure 4-12 Illustration des moyennes des résultats pour la variable CONTEXTE en Histoire en fonction des effets des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2) et de leur interaction.

Les moyennes se distinguent d'un niveau scolaire à l'autre, les résultats augmentant progressivement (voir annexe G). Cependant, les moyennes des résultats en fonction de la langue maternelle ne se distinguent qu'en secondaire 1 (LM1=57,78 %; LM2=48,15 %) et en secondaire 2 (LM1=70,49 %; LM2=66,85 %), pas secondaire 2 (LM1=64,71 %; LM2=66,0 %).

Les figures 4-13 à 4-15 illustrent l'interaction entre les variables Langue maternelle et Sexe.

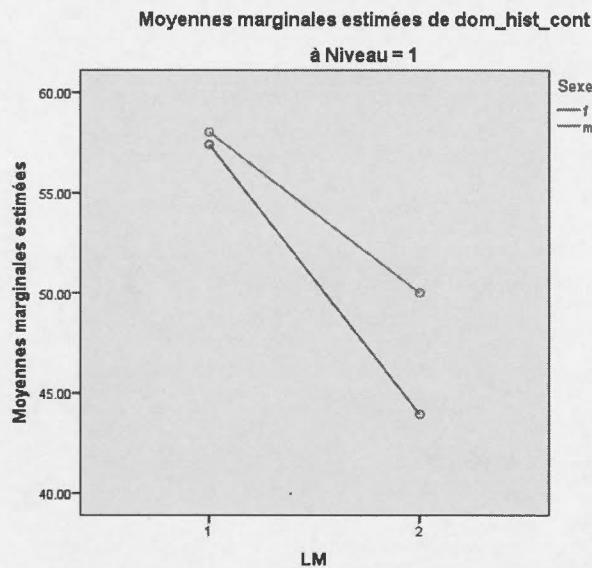


Figure 4-13 Illustration des moyennes des résultats pour la variable CONTEXTE en histoire en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 1.

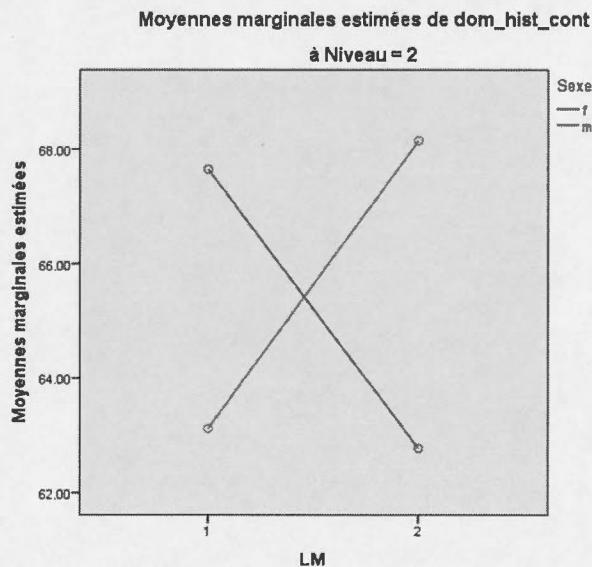


Figure 4-14 Illustration des moyennes des résultats pour la variable CONTEXTE en histoire en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 2.

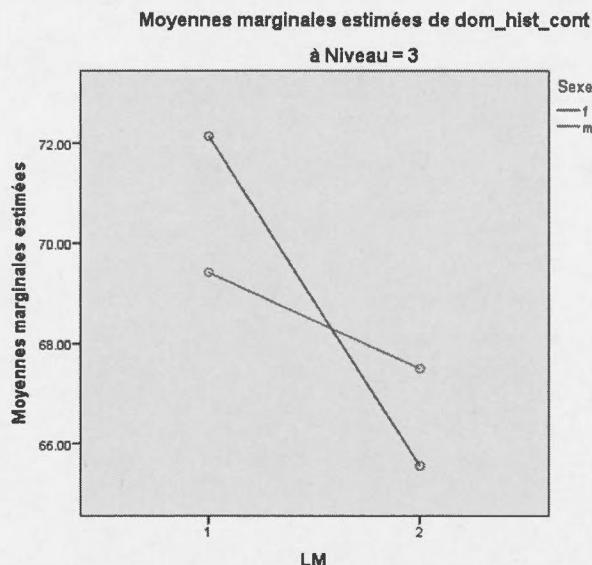


Figure 4-15 Illustration des moyennes des résultats pour la variable CONTEXTE en histoire en fonction des effets statistiques Sexe et Langue maternelle (1=francophones; 2=allophones) pour les élèves de secondaire 3.

Dans la figure 4-13, nous observons qu'en première secondaire, les garçons (LM1=58,02 %; LM2=50,0 %) obtiennent des résultats légèrement supérieurs aux filles (LM1=57,41 %; LM2=43,94 %) (mais nous rappelons qu'il n'y a pas d'effet statistique pour la variable Sexe). Par contre, pour les deux autres niveaux, les garçons francophones obtiennent des résultats plus faibles que les filles francophones (m, sec. 2=63,12 % et m, sec. 3=69,42 % contre f, sec 2=67,65 % et f, sec. 3=72,14 %), alors que les garçons allophones obtiennent des résultats plus élevés que les filles allophones (m, sec. 2=68,15 % et m, sec. 3=67,5 % contre f, sec 2=62,78 % et f, sec. 3=65,56 %), comme le montrent les Figures 4-14 et 4-15. Nous remarquons que le phénomène est diamétralement opposé et presque symétrique entre les garçons et les filles de secondaire 2 selon la langue maternelle, ce qui explique les performances semblables pour ce niveau de scolarité et l'interaction Niveau de scolarité et Langue maternelle illustrés dans la Figure 4-12.

Domaine Science et technologie

Les variances sont très grandes, les résultats variant entre 0 et 100 % (voir annexe F). Les différences entre les moyennes et les médianes sont grandes. Le calcul de la normalité des distributions montre qu'aucune distribution sur douze n'est normale. Par contre, les pentes sont symétriques. De plus, le test d'égalité des variances des erreurs de Levene révèle que les variances des erreurs de la variable dépendante sont égales sur les différents groupes, $D(11, 576) = 1,432$, ns (voir annexe G). Nous poursuivons donc avec l'analyse de variance. Elle examine les effets du niveau de scolarité, de la langue maternelle et du sexe sur le pourcentage de réussite pour la variable CONTEXTE pour le domaine Science et technologie. Aucun effet n'est trouvé pour les variables Niveau ($D_{2,576} = 2,399$, ns, Eta carré partiel = 0,008), Langue maternelle ($D_{1,576} = 1,364$, ns, Eta carré partiel = 0,002) ou Sexe ($D_{1,576} = 2,265$, ns, Eta carré partiel = 0,004). Par contre, il y a une interaction statistique entre les variables Niveau et Langue maternelle ($D_{2,576} = 6,195$, $p < 0,01$, Eta carré partiel = 0,021) avec un petit effet (2 %). Le modèle explique $R^2 = 4\%$ de la variance dans les résultats pour la variable CONTEXTE pour le domaine Science et technologie. Les comparaisons multiples avec le test post hoc de Bonferroni ne montrent aucune différence statistique dans la moyenne des résultats entre les niveaux de scolarité ($p > 0,05$ dans tous les cas). Le test GT2 propose la même interprétation (voir annexe G).

La Figure 4-16 illustre les variables Niveau de scolarité et Langue maternelle et leur interaction statistique sur la variable CONTEXTE en Science et technologie.

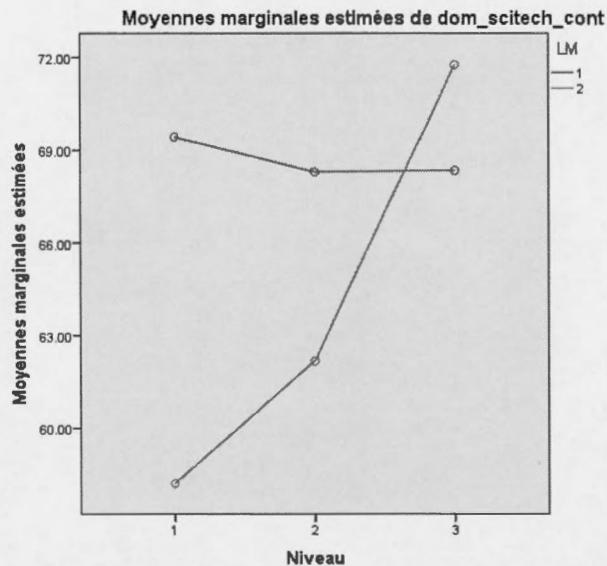


Figure 4-16 Illustration des moyennes des résultats pour la variable CONTEXTE en Science et technologie en fonction des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2).

Nous voyons que les résultats des élèves francophones sont semblables, peu importe le niveau (sec. 1=69,67 %; sec. 2=67,65 %; sec. 3=67,80 %) alors que les résultats des élèves allophones se distinguent d'un niveau à l'autre (sec. 1=57,87 %; sec. 2=62,03 %; sec. 3=72,04 %) (voir annexe G). Bien que les résultats des élèves allophones soient plus faibles que ceux des élèves francophones en secondaire 1 et 2, l'effet n'est pas statistique, probablement en partie parce que les élèves allophones performent mieux que les élèves francophones en troisième secondaire. Cette distinction entre les droites en fonction du niveau scolaire et de la langue maternelle explique l'effet statistique d'interaction entre les deux variables.

Les mots fréquents

Les variances sont grandes entre les moyennes des résultats pour la dimension CONTEXTE pour les mots fréquents (voir annexe F). Les résultats varient entre 50 et 100 %. Les différences entre les moyennes et les médianes sont grandes. Le calcul de la normalité des distributions montre qu'une distribution sur douze est normale (voir annexe F). Par contre, les

pendentes sont symétriques et le test d'égalité des variances des erreurs de Levene révèle que les variances des erreurs de la variable dépendante sont égales sur les différents groupes, $D(11, 554) = 1,799$, ns (mais $p=0,51$; voir annexe G-N). Malgré un résultat au test près du seuil statistique, nous poursuivons avec l'analyse de variance. Elle examine les effets du niveau de scolarité, de la langue maternelle et du sexe sur le pourcentage de réussite pour la variable CONTEXTE pour les mots fréquents. Aucun effet n'est trouvé pour les variables Niveau ($D_{2,554} = 1,622$, ns, Eta carré partiel = 0,006), et Sexe ($D_{1,554} = 0,064$, ns, Eta carré partiel = 0,000). Un petit effet (2 %) est trouvé pour la variable Langue maternelle ($D_{1,554} = 13,903$, $p<0,001$, Eta carré partiel = 0,024). Aucune interaction statistique n'a été trouvée. Le modèle explique $R^2 = 4\%$ de la variance dans les résultats pour la variable CONTEXTE pour les mots fréquents.

La Figure 4-17 illustre les variables Niveau de scolarité et Langue maternelle et l'effet statistique de cette dernière sur la variable CONTEXTE pour les mots fréquents.

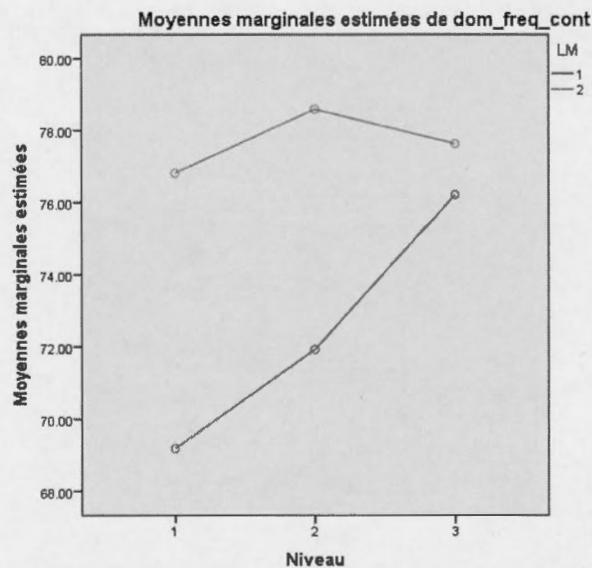


Figure 4-17 Illustration des moyennes des résultats pour la variable CONTEXTE pour les mots fréquents en fonction des variables Niveau de scolarité (1 à 3) et Langue maternelle (francophone=1; allophones=2).

Nous observons que les résultats des élèves allophones (sec. 1=68,98 %; sec. 2=71,86 %; sec. 3=75,95 %) sont plus faibles que ceux des élèves francophones (sec. 1=77,0 %; sec. 2=78,24 %; sec. 3=77,78 %), particulièrement en secondaire 1 et 2. Nous observons également une différence entre les résultats des élèves allophones d'une année scolaire à l'autre, alors que ça n'est pas le cas chez les élèves francophones. Cependant, cette dernière observation n'est pas soutenue par l'analyse statistique.

Profils des résultats par domaine pour la dimension CONTEXTE

Pour compléter cette section, nous présentons la Figure 4-18 pour illustrer les profils des moyennes des résultats par domaine, par Niveau et par Langue maternelle, pour la dimension CONTEXTE.

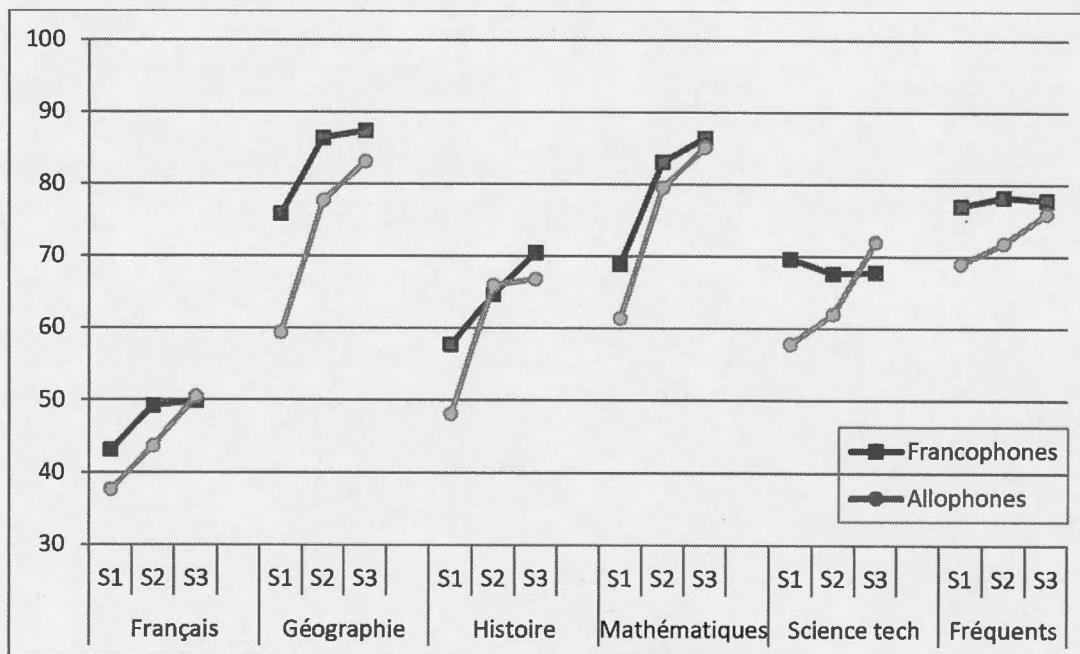


Figure 4-18 Illustration des profils des résultats des élèves pour la variable CONTEXTE (connaissances syntagmatiques) en fonction de la langue maternelle, du niveau scolaire et des domaines d'apprentissages.

Les profils montrent une progression chez les élèves allophones pour tous les domaines, les résultats de ces élèves étant plus faibles en secondaire 1 puis rejoignant et dépassant même

ceux des élèves francophones en histoire en secondaire 2 et en français, en mathématiques et en science et technologie en secondaire 3. Chez les élèves francophones, les performances ne se distinguent pas d'un niveau de scolarité à l'autre pour ce qui est des connaissances syntagmatiques en science et technologie et pour les mots fréquents, sans qu'il n'y ait manifestation de plafonnement. Les résultats s'améliorent d'un niveau à l'autre dans les autres domaines.

4.3.7 Conclusion

Pour résumer cette section, nous avons inscrit dans des tableaux les différents effets des variables sur les connaissances paradigmatisques (variable SENS) puis sur les connaissances syntagmatiques (variable CONTEXTE), pour l'ensemble des domaines d'apprentissages et pour chaque domaine. Nous avons également produit deux figures, l'une pour les connaissances paradigmatisques, l'autre pour les connaissances syntagmatiques, pour illustrer les profils des résultats en fonction de la langue maternelle et des domaines. Le Tableau 4-9 résume les effets des variables pour les connaissances paradigmatisques.

Tableau 4-9 Résumé des effets statistiques pour les variables Niveau de scolarité (1 à 3), Langue maternelle et Sexe sur les connaissances paradigmatisques.

Effets/Domaine	Tous domaines	Français	Géographie	Histoire	Maths	Science tech.	Fréquents
Niveau de scolarité	***	plaf., dv	plaf., dv	plaf., dv	plaf., dv	plaf., dv	plaf., dv
Sec 1 vs 2	***						
Sec 1 vs 3	***						
Sec 2 vs 3	ns						
Langue maternelle	***						
Sexe	ns						
Niveau x LM	ns						
Niveau x Sexe	ns						
LM x Sexe	ns						

Légende. *** : $p < 0,001$; ** : $p < 0,01$; * : $p < 0,05$; ns : non significatif; dv : distributions anormales et/ou variances inégales; plaf. : plafonnement.

Le Tableau 4-10 résume les effets des variables pour les connaissances syntagmatiques.

Tableau 4-10 Résumé des effets statistiques pour les variables Niveau de scolarité (1 à 3), Langue maternelle et Sexe sur les connaissances syntagmatiques.

Effets/Domaine	Tous domaines	Français	Géographie	Histoire	Maths	Science tech.	Fréquents
Niveau de scolarité	***	**	plaf., dv	***	plaf., dv	ns	ns
Sec 1 vs 2	***	*		***		ns	ns
Sec 1 vs 3	***	***		***		ns	ns
Sec 2 vs 3	**	ns		***		ns	ns
Langue maternelle	***	ns		*		ns	***
Sexe	ns	**		ns		ns	ns
Niveau x LM	*	ns		**		**	ns
Niveau x Sexe	ns	ns		ns		ns	ns
LM x Sexe	ns	ns		**		ns	ns

Légende. *** : p<0,001; ** : p<0,01; * : p<0,05; ns : non significatif; dv : distributions anormales et/ou variances inégales; plaf. : plafonnement.

À la page suivante, la Figure 4-19 montre les profils des résultats des élèves pour les connaissances paradigmiques et syntagmatiques en fonction de la langue maternelle et du niveau de scolarité.

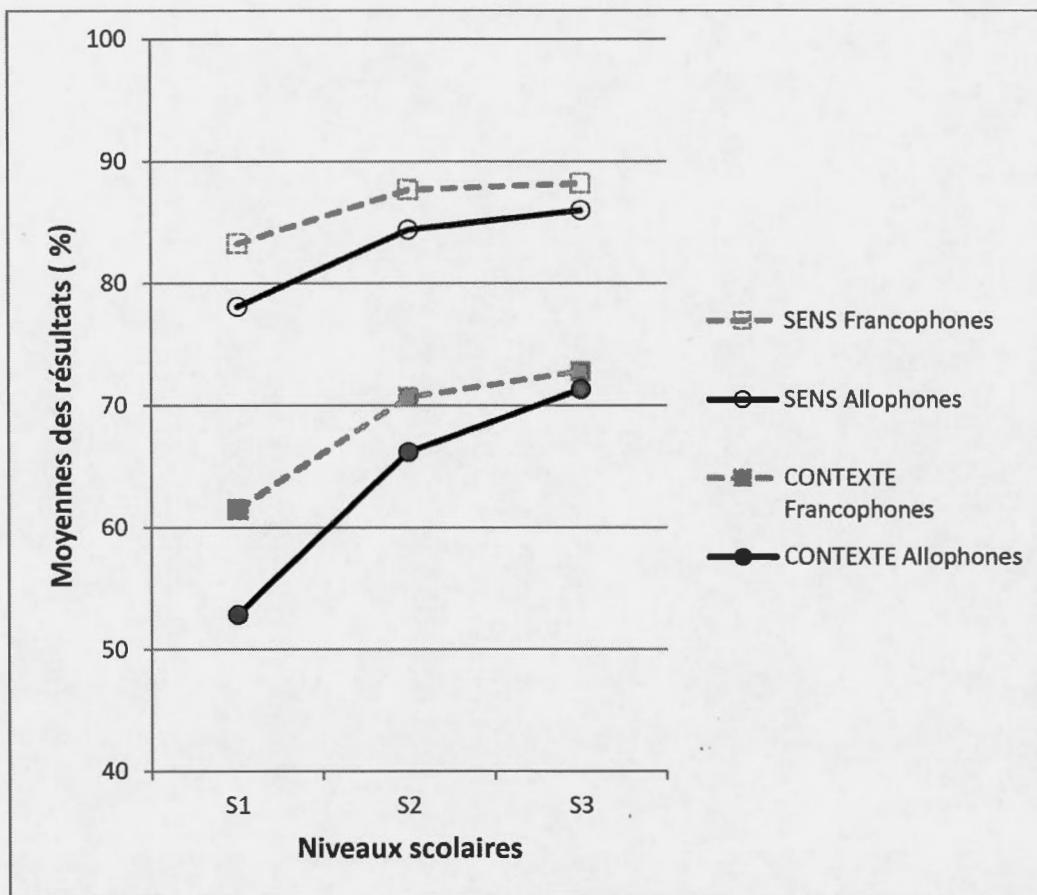


Figure 4-19 Illustration des profils des résultats des élèves pour les dimensions SENS (connaissances paradigmatiques) et CONTEXTE (connaissances syntagmatiques) en fonction de la langue maternelle et du niveau scolaire

Les illustrations montrent des profils semblables entre les élèves allophones et les élèves francophones, mais des résultats plus faibles pour les élèves allophones, particulièrement en secondaire 1. Nous soulignons cependant que les analyses montrent une grande variance d'un groupe à l'autre et une faible ampleur des effets. Les analyses ont montré peu d'effet entre les sexes. Par ailleurs, nous observons que les connaissances paradigmatiques sont systématiquement plus grandes que les connaissances syntagmatiques, peu importe le niveau de scolarité ou la langue maternelle.

4.4 Objectif de recherche 3 : mesurer la relation entre le degré d'abstraction des mots et la profondeur de la connaissance du vocabulaire des élèves

Dans cette section, nous analysons les scores des élèves à chaque mot en fonction du degré d'abstraction des mots (en plus des facteurs Langue maternelle et Niveau de scolarité).²⁶ Dans le cadre théorique, nous avons expliqué l'importance de tenir compte du degré d'abstraction des mots comme facteur d'apprentissage. L'apprentissage des mots les plus abstraits dépend de connaissances préalables, dont des connaissances syntaxiques. Or, les élèves allophones sont susceptibles de moins maîtriser les règles syntaxiques de la langue d'enseignement. Nous vérifions dans cette section si le degré d'abstraction d'un mot est un facteur dans les scores pour chacune des dimensions (SENS et CONTEXTE). Comme l'effet du degré d'abstraction peut interagir avec les effets de la fréquence et de la longueur d'un mot, nous vérifions également les effets éventuels de ces facteurs et leurs interactions (voir section 2.3.7).

4.4.1 Description des résultats en fonction des facteurs lexicaux

Avant de procéder à l'analyse statistique, nous décrivons les moyennes des résultats des élèves pour chaque dimension (SENS, CONTEXTE) en fonction des divers facteurs lexicaux. Le détail de ces moyennes se trouve à l'annexe J. Bien que les données soient traitées autrement dans notre analyse (par des calculs logistiques), nous jugeons que de décrire les données avec les moyennes facilite la compréhension des effets éventuels. Rappelons que nous avons marqué chacun des mots pour son degré d'abstraction, sa fréquence et sa longueur. Nous avons vérifié si ces indices étaient corrélés entre eux et c'est le cas : il y a une corrélation positive entre le degré d'abstraction et la longueur des mots ($P(35)=.499$, $p<0,01$) et une corrélation négative entre la fréquence et la longueur des mots ($P(35)=-.502$, $p<0,01$). La

²⁶ Nous n'incluons pas la variable Sexe dans notre analyse, les effets mesurés dans les analyses de la section précédente n'étant pas assez importants. De plus, nous voulons limiter le nombre de variables dans notre modélisation.

description de la procédure et le tableau détaillant les indices pour chaque mot se trouvent en annexe D.

Les figures Figure 4-20 et Figure 4-21 suivantes montrent la dispersion des résultats moyens en fonction du degré d'abstraction des mots pour la dimension SENS puis pour la dimension CONTEXTE. Le degré d'abstraction des mots varie de 1 à 6. Les moyennes se situent entre 0 et 1.

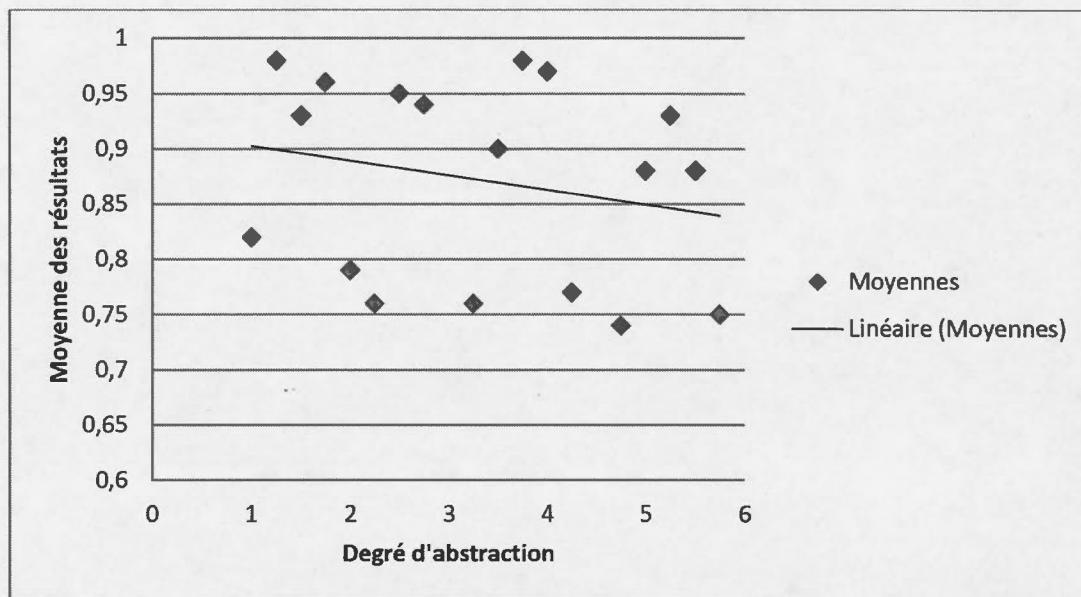


Figure 4-20 Dispersion des résultats moyens en fonction du degré d'abstraction des mots pour la dimension SENS

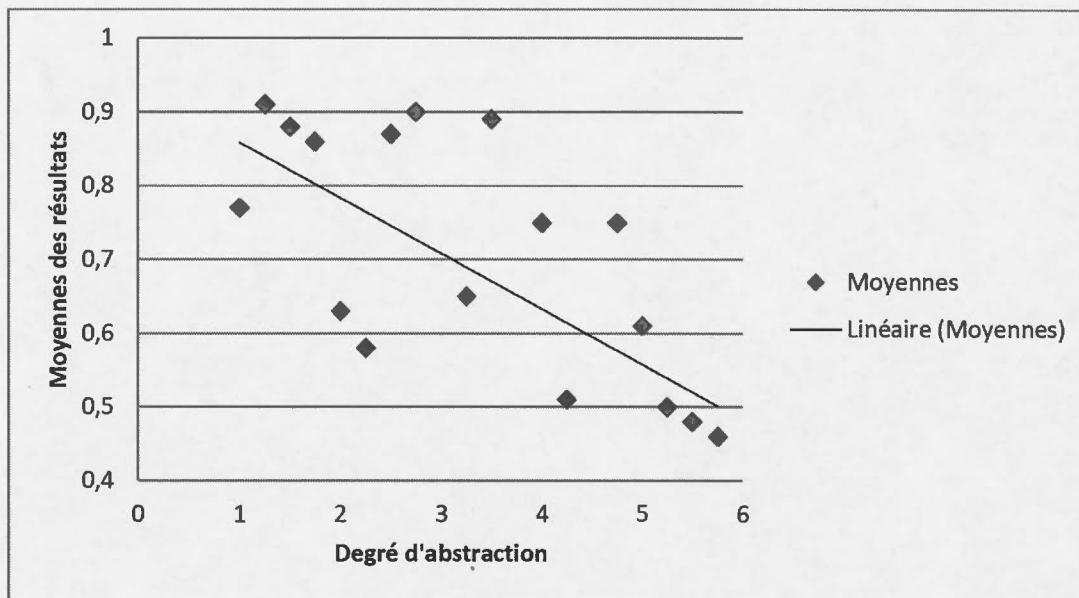


Figure 4-21 Dispersion des résultats moyens en fonction du degré d'abstraction des mots pour la dimension CONTEXTE

Les pentes suggèrent une corrélation négative entre les résultats moyens et le degré d'abstraction pour les dimensions SENS et CONTEXTE. Plus le degré d'abstraction est élevé, plus faible serait le score. De plus, nous observons que la pente est plus prononcée pour la dimension CONTEXTE.

Les figures Figure 4-22 et Figure 4-23 montrent la dispersion des résultats moyens en fonction de la fréquence des mots pour la dimension SENS puis pour la dimension CONTEXTE. La fréquence des mots varie de 34 à 72.

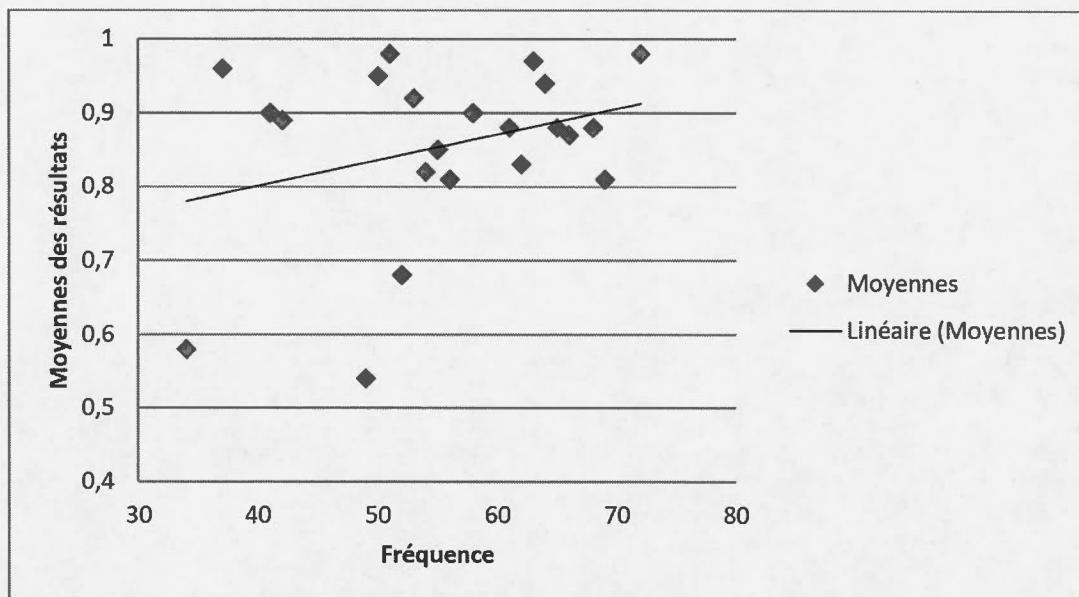


Figure 4-22 Dispersion des résultats moyens en fonction de la fréquence des mots pour la dimension SENS

La pente de la Figure 4-22 suggère une corrélation positive entre les résultats moyens et la fréquence pour la dimension SENS. Plus la fréquence est élevée, meilleur serait le score.

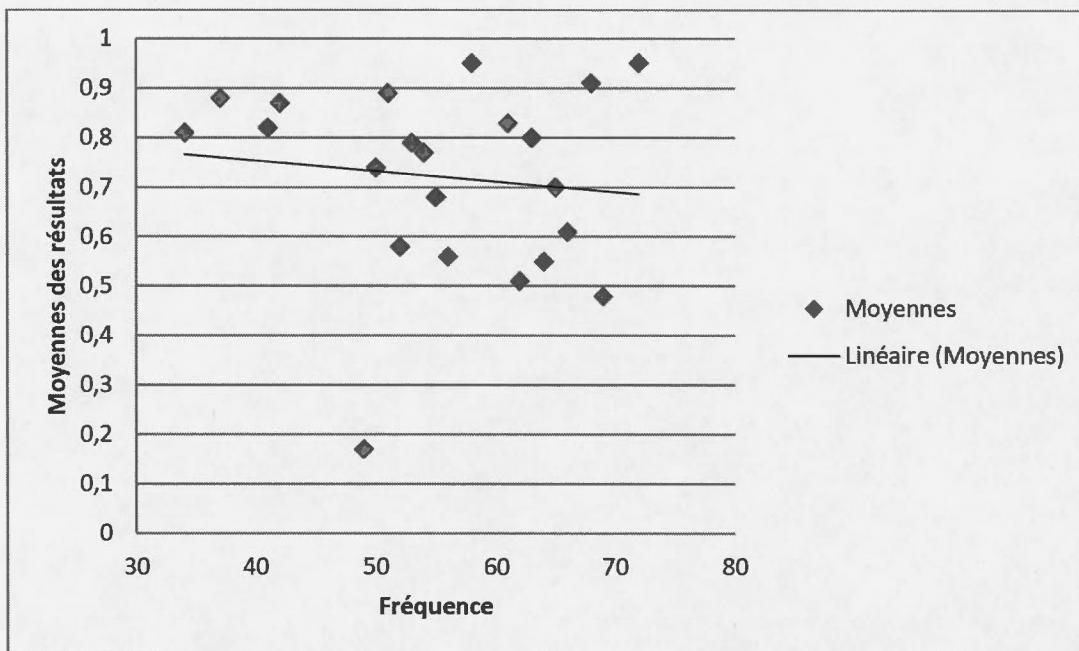


Figure 4-23 Dispersion des résultats moyens en fonction de la fréquence des mots pour la dimension CONTEXTE

Par contre, la pente de la Figure 4-23 suggère une corrélation négative entre la fréquence et le score pour la dimension CONTEXTE : plus la fréquence est élevée, plus faible serait le score, ce qui est paradoxal. Même en retirant la donnée extrême (fréquence=49), la pente reste négative. Nous revenons à ce phénomène dans la section suivante à la lumière de l'analyse statistique.

Les figures Figure 4-24 et Figure 4-25 suivantes montrent la dispersion des résultats moyens en fonction de la longueur des mots pour la dimension SENS puis pour la dimension CONTEXTE. La longueur des mots varie de 3 à 13 lettres.

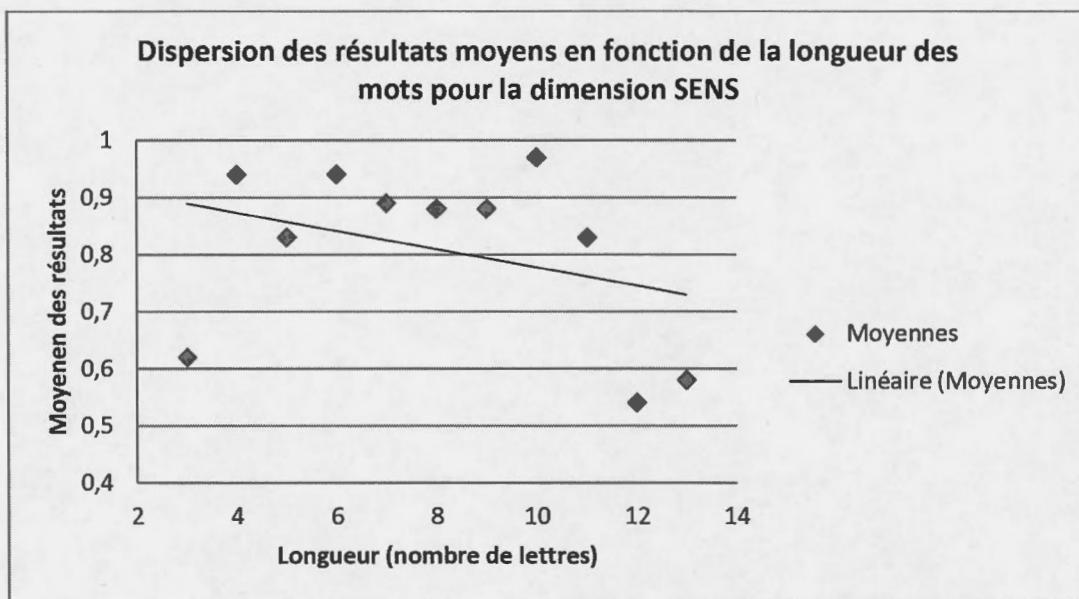


Figure 4-24 Dispersion des résultats moyens en fonction de la longueur des mots pour la dimension SENS

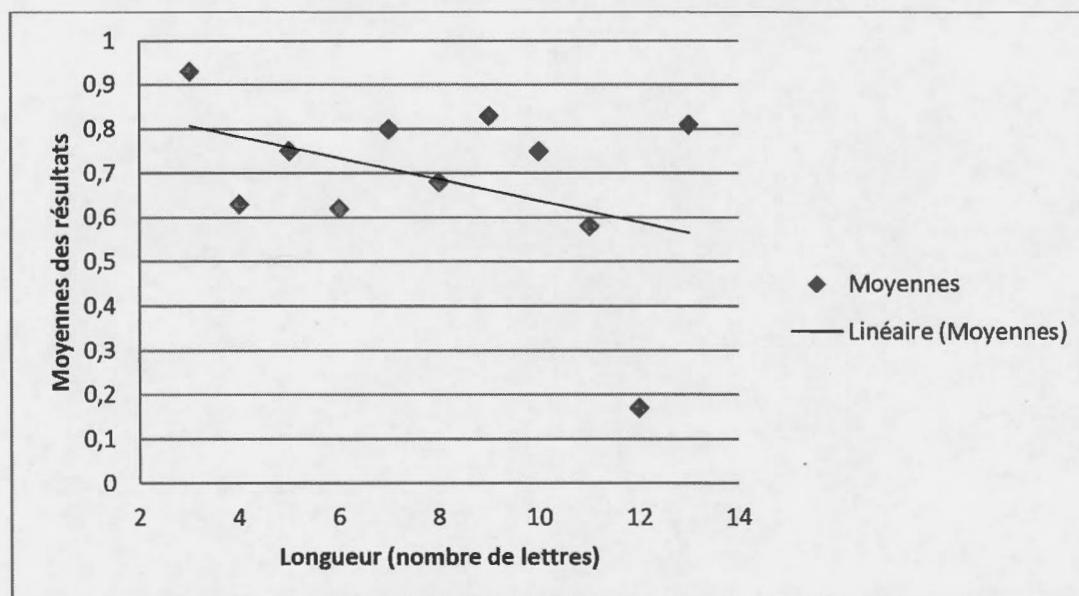


Figure 4-25 Dispersion des résultats moyens en fonction de la longueur des mots pour la dimension CONTEXTE

Les pentes suggèrent une corrélation négative entre les résultats moyens et la longueur des mots. Plus la longueur est élevée, plus faible serait le score, pour la dimension SENS autant que pour la dimension CONTEXTE.

Pour vérifier si les pentes observées dans les figures qui précèdent sont statistiques, nous avons besoin d'une méthode d'analyse qui puisse mesurer les relations entre les variables quand la variable dépendante est dichotomique (le résultat est 1 ou 0) et qu'il s'agit d'une mesure répétée (plusieurs réponses proviennent d'un participant et les réponses d'un participant sont probablement corrélées). Ces méthodes d'analyse de corrélation sont rares (Zorn, 2001). L'une d'elles est une modélisation linéaire appelée « Équations d'estimation généralisées » (en anglais, *Generalized Estimating Equations* ou GEE). Cette modélisation est appropriée pour la vérification de notre objectif puisqu'elle permet l'analyse de mesures répétées ou d'autres observations corrélées, comme les données agglutinées (comme les données corrélées d'un même élève ou selon la langue maternelle ou le niveau de scolarité), quand les variables sont, entre autres, dichotomiques (Carrière et Bouyer, 2002); SPSS, aide en ligne). Cette méthode permet d'estimer les coefficients des paramètres de régression (l'effet des facteurs lexicaux dans notre étude) de façon efficace et de produire des erreurs standards précises (Hanley, Negassa, Edwardes et Forrester, 2003).

Notre analyse consiste ainsi en une analyse linéaire à mesure répétée pour laquelle nous avons des réponses à 35 mots²⁷ par 589 élèves (381 élèves francophones; 208 élèves allophones). Comme les scores sont soit « 1 » (deux bons choix de réponse) ou « 0 » (un ou aucun bon choix de réponse), nous cherchons à modéliser la régression pour des variables dichotomiques. Nous avons recours à une modélisation logistique, dans laquelle nous cherchons à prédire la probabilité d'obtenir un « 1 » en fonction de différents facteurs (Bressoux, 2010).

²⁷ Nous retirons d'emblée de l'analyse le mot « PIB », parce qu'il est un acronyme, même s'il est lexicalisé, susceptible de biaiser la relation éventuelle entre les facteurs Longueur de mot et Degré d'abstraction.

Pour décrire cette probabilité d'occurrence d'un évènement, nous utilisons la notion d'odds, notée O (Bressoux, 2010). Par exemple, si nous avons une valeur $O_i = 2$, cela signifie qu'il y a deux fois plus d'occurrences attendues d'un évènement i que de non-occurrences, alors qu'une valeur $O_i = 0,25$ signifie que l'on attend quatre fois moins d'occurrences de l'évènement i que de non-occurrences (Bressoux, 2010, p. 229). Cependant, nous soulignons que la modélisation produite par SPSS et décrite plus loin calcule des coefficients de probabilité (valeurs B) fondés sur le log de l'odds ratio. Le log permet d'obtenir des effets de $-\infty$ à $+\infty$ (l'odds, lui, a une valeur minimale de 0). Comme la valeur du coefficient B est difficile à appréhender, nous récupérons l'odds ratio dans la conclusion de cette section, en calculant l'exponentielle des coefficients, c'est-à-dire $\exp(B)$. L'odds ratio est un rapport de chances d'occurrence d'un évènement entre deux catégories ou plus d'une variable nominale (par exemple, la langue maternelle : dans quelle mesure un élève du groupe LMx a-t-il plus ou moins de chance d'obtenir un succès qu'un élève du groupe LMy) ou « le changement dans les odds entraîné par la variation d'une unité » d'une variable quantitative (par exemple, le nombre de lettres d'un mot : dans quelle mesure un mot de $n+1$ lettres affecte-t-il les chance d'obtenir un succès par rapport à un mot de n lettres) (Bressoux, 2010, p. 236).

La procédure de modélisation est la suivante : nous mettons d'abord en place un modèle sans variable indépendante (modèle 0, voir Tableau 4-11). Ensuite, nous ajoutons progressivement les variables indépendantes dans de nouveaux modèles. En plus de vérifier l'effet des facteurs lexicaux sur la probabilité de succès, nous ajoutons éventuellement dans le modèle les interactions entre les facteurs lexicaux et les variables Niveau de scolarité et Langue maternelle pour voir si les profils des élèves diffèrent à la lumière de ces deux dernières variables. Nous sommes particulièrement intéressé de savoir si les élèves allophones ont les mêmes chances de succès en fonction du degré d'abstraction des mots que les élèves francophones. Éventuellement, nous composons des modèles prenant en compte les interactions entre les variables indépendantes. Pour chaque modèle calculé, nous obtenons une valeur de quasi-vraisemblance (QIC) qui nous informe de la vraisemblance du modèle par rapport aux données en fonction d'une structure de corrélation (nous avons conservé la

structure « indépendante », par défaut dans SPSS). Plus petite est la valeur d'un modèle comparativement à celle d'un autre modèle, plus vraisemblable est le modèle. Ensuite, pour chaque modèle, le khi-deux de Wald vérifie l'effet des variables indépendantes (voir les annexes 9 et 10 pour le détail de ces analyses). Finalement, nous obtenons les coefficients estimés pour chaque paramètre (pour l'ordonnée à l'origine puis pour chaque variable indépendante). Dans la conclusion de cette section, nous présentons les odds ratio (OR) calculés à partir des coefficients les plus pertinents considérant le cadre de notre étude.

4.4.2 Modélisation des effets des variables lexicales sur les résultats pour la variable SENS

Nous avons calculé plusieurs modèles, d'abord pour la variable SENS puis pour la variable CONTEXTE. Les annexes 9 (SENS) et 10 (CONTEXTE) les présentent les uns à la suite des autres. Dans le Tableau 4-11 qui suit, en plus du modèle 0, nous montrons les modèles pour la variable SENS. Nous décrivons ensuite le modèle dont le QIC indique la plus grande vraisemblance (c'est-à-dire la valeur de QIC la plus basse).

Tableau 4-11 Équations d'estimation généralisées pour les facteurs lexicaux (degré d'abstraction, fréquence, longueur) et les variables relatives aux élèves (niveau de scolarité, langue maternelle) sur la variable SEN\$.

	M0	M1	M2	M3	M4
QIC	16649,040	16362,833	15840,937	15808,497	15736,363
Paramètres	B	Et.s.	B	Et.s.	B
Ord.O	1,821***	,0385	1,862***	,0407	2,049***
Abs			-,131***	,0123	-,130***
Fréq			,020***	,0022	,019***
Long			-,033**	,0100	,075***
Abs*Fréq					,023***
Abs*Long					-,039***
Fréq*Long					,003***
Abs*Fréq*Long					
LM=2					
LM=1					0
Niv=1					-,347***
Niv=2					-,090
Niv=3					0
Niv=1 * Abs					
Niv=2 * Abs					
Niv=3 * Abs					
Niv=1 * Freq					
Niv=2 * Freq					
Niv=3 * Freq					
Niv=1 * Long					
Niv=2 * Long					
Niv=3 * Long					
LM=2 * Abs					
LM=1 * Abs					
LM=2 * Freq					
LM=1 * Freq					
LM=2 * Long					
LM=1 * Long					

QIC=quasi-vraisemblance sous un critère de modèle d'indépendance; Ord.O.=ordonnée à l'origine; B=coefficent estimé; Et.s.= erreur standard; *p<0,05;

p<0,01; *p<0,001

Dans tous les modèles où nous prenons en compte les effets des variables lexicales, ces variables montrent un effet significatif (voir l'annexe H pour le détail des Khi deux de Wald). Parmi les modèles, le modèle 4 montre le plus de vraisemblance (QIC=15736,363). Nous interprétons les résultats à la lumière de ce modèle. Nous soulignons d'abord les effets des facteurs lexicaux. Ensuite, nous décrivons les effets d'interaction entre ces facteurs. Nous terminons par la description des effets éventuels des variables relatives aux élèves (langue maternelle et niveau de scolarité).

Selon ce modèle, les variables lexicales ont toutes un effet statistique sur la probabilité d'obtenir 1 pour la variable SENS. Ainsi, une variation positive dans le degré d'abstraction réduit les chances d'avoir une bonne réponse, ($B_{abs}(1) = -,164$, $p<0,001$). Dans le cas de la fréquence des mots, une variation positive augmente les chances ($B_{Freq}(1) = ,030$, $p<0,001$). C'est le cas également pour la longueur des mots ($B_{Long}(1) = ,061$, $p<0,01$), ce qui est plutôt paradoxal.²⁸ L'explication réside peut-être dans un effet d'interaction avec les autres variables lexicales.

D'ailleurs, le modèle 4 montre des interactions entre les variables lexicales. Ces interactions et le changement de l'effet négatif au positif pourraient indiquer une non-linéarité dans la relation ou souligner la corrélation entre les variables lexicales (voir annexe D). Les coefficients positifs montrent que les chances d'une bonne réponse augmentent en fonction des variations entre le degré d'abstraction et la fréquence des mots ($B_{Abs*Freq}(1) = ,022$, $p<0,001$) et en fonction des variations entre la fréquence des mots et la longueur ($B_{Freq*Long}(1) = ,005$, $p<0,001$). Avec un coefficient négatif, les chances d'une bonne réponse diminuent en fonction des variations du degré d'abstraction et de la longueur des mots ($B_{Abs*Long}(1) = -,047$, $p<0,001$). Un coefficient statistique négatif est également calculé en fonction de l'interaction entre les trois facteurs lexicaux ($B_{Abs*Freq*Long}(1) = -,009$, $p<0,001$).

²⁸ Dans le modèle 1, l'effet, statistique, de la longueur est négatif ($B_{Long,Modèle1}(1) = -,033$, $p<0,001$).

Des effets apparaissent pour les variables relatives aux élèves. Les chances sont réduites pour les élèves de secondaire 1 et sont plus grandes pour les élèves francophones, comme l'ont montré les analyses dans la section précédente. Ce qui nous intéresse dans cette section ce sont les interactions entre les variables lexicales et la langue maternelle. Il n'y a pas d'interaction entre ces variables et la variable Langue maternelle ($B_{LM2*Abs}(1) = ,012$, ns; $B_{LM2*Fréq}(1) = -,008$, ns; $B_{LM2*Long}(1) = -,015$, ns), ce qui indique que les profils des deux populations sont semblables, même si les élèves allophones ont moins de chance d'obtenir un succès que les élèves francophones ($B_{LM=2}(1) = -,169$, $p<0,05$), ce qui a également été montré dans les analyses de la section précédente. Par ailleurs, il y a interaction entre les résultats pour les élèves de secondaire 2 et la variable Degré d'abstraction ($B_{Niv2*Abs}(1) = -,079$, $p<0,05$), ce qui montre une pente ou une ordonnée à l'origine différente pour les résultats des élèves de secondaire 2 en ce qui concerne le degré d'abstraction des mots. Finalement, il y a interaction entre les résultats pour les élèves de secondaire 1 et la variable Longueur ($B_{Niv1*Long}(1) = -,063$, $p<0,05$), ce qui montre également une pente ou une ordonnée à l'origine différente pour les résultats de ces élèves en ce qui concerne la longueur des mots. Nous ne nous attarderons pas sur le sujet de ce dernier phénomène.

4.4.3 Modélisation des effets des variables lexicales sur les résultats pour la variable CONTEXTE

Le Tableau 4-12 résume les modèles dont les QIC indiquent la plus grande vraisemblance (c'est-à-dire les valeurs des QIC les plus basses) pour la variable CONTEXTE (voir annexe J pour le détail de l'analyse).

Tableau 4-12 Équations d'estimation généralisées pour les facteurs lexicaux (degré d'abstraction, fréquence, longueur) et les variables relatives aux élèves (niveau de scolarité, langue maternelle) sur la variable CONTEXTE.

QIC	M0			M1			M2			M3			M4		
	23642,203	24114,065	23981,120	23843,168	23843,168	23843,168	23843,168	23843,168	23843,168	23843,168	23843,168	23843,168	23791,560	23791,560	
Paramètres	B	Er.s.	B												
Ord.O	.784	,0243***	,859***	,0274	,997***	,0319	,1.196***	,0485	,1.185***	,0485	,1.185***	,0498	,0498	,0498	,0498
Abs			-,380***	,0119	-,38***	,0120	-,384***	,0122	-,320***	,0122	-,320***	,0201	,0201	,0201	,0201
Fréq			-,026***	,0119	-,039***	,0027	-,039***	,0027	-,052***	,0027	-,052***	,0040	,0040	,0040	,0040
Long			-,016*	,0072	-,022**	,0085	-,022**	,0085	-,032*	,0085	-,032*	,0135	,0135	,0135	,0135
Abs*Fréq					-,021***	,0017	-,022***	,0017	-,022***	,0017	-,022***	,0017	,0017	,0017	,0017
Abs*Long					-,045***	,0065	-,046***	,0065	-,047***	,0065	-,047***	,0065	,0065	,0065	,0065
Fréq*Long					-,006***	,0008	-,006***	,0008	-,006***	,0008	-,006***	,0008	,0011	,0011	,0011
Abs*Fréq*Long													,005***	,005***	,0007
LM=2															
LM=1															
Niv=1															
Niv=2															
Niv=3															
Niv=1 * Abs															
Niv=2 * Abs															
Niv=3 * Abs															
Niv=1 * Freq															
Niv=2 * Freq															
Niv=3 * Freq															
Niv=1 * Long															
Niv=2 * Long															
Niv=3 * Long															
LM=2 * Abs															
LM=1 * Abs															
LM=2 * Freq															
LM=1 * Freq															
LM=2 * Long															
LM=1 * Long															

QIC=quasi-vraisemblance sous un critère de modèle d'indépendance; Ord.O.=ordonnée à l'origine; B=coefficent estimé; Er.s.=erreur standard; *p<0,05;

p<0,01; *p<0,001

Pour la variable CONTEXTE, peu importe le modèle où nous prenons en compte les effets des variables lexicales, ces variables montrent un effet significatif (voir l'annexe J pour le détail des Khi deux de Wald). Parmi les modèles, le modèle 4 montre le plus de vraisemblance (QIC=23791,560). Nous interprétons les résultats à la lumière de ce modèle. Selon ce modèle, les variables lexicales ont toutes un effet statistique sur la probabilité d'obtenir 1 pour la variable CONTEXTE. Ainsi, une variation positive dans le degré d'abstraction réduit les chances d'occurrence d'une bonne réponse, ($B_{abs}(1) = -.320$, $p<0,001$). C'est le cas également pour la longueur des mots ($B_{Long}(1) = -.032$, $p<0,05$). Dans le cas de la fréquence des mots, une variation positive réduit les chances ($B_{Freq}(1) = -.052$, $p<0,001$), ce qui est paradoxal, comme nous l'avons souligné plus tôt.

Les interactions entre les variables lexicales sont statistiques. Les coefficients négatifs montrent une réduction des chances d'avoir une bonne réponse en fonction des mots les plus abstraits et les plus fréquents ($B_{Abs*Fréq}(1) = -.022$, $p<0,001$). Cette interaction explique peut-être le paradoxe soulevé dans le paragraphe précédent. Des mots abstraits plus fréquents feraient en sorte que les mots fréquents sont moins réussis. Les variations entre les mots les plus abstraits et les plus longs affectent négativement les chances de réussite ($B_{Abs*Long}(1) = -.047$, $p<0,001$). Les variations entre les trois facteurs lexicaux donnent un coefficient positif ($B_{Abs*Fréq*Long}(1) = .005$, $p<0,001$).

En ce qui concerne les variables relatives aux élèves, les chances augmentent pour les élèves de secondaire 1 avec les mots plus fréquents ($B_{Niv=1}(1) = .011$, $p<0,05$). Les élèves montrent des profils semblables en ce qui concerne la langue maternelle et le degré d'abstraction ($B_{LM2*Abs}(1) = .011$, ns) ou la fréquence des mots ($B_{LM2*Freq}(1) = .003$, ns). Il y a une seule interaction, entre la variable Longueur des mots et la variable Langue maternelle ($B_{LM2*Long}(1) = .032$, $p<0,05$), ce qui montre une pente ou une ordonnée à l'origine différente pour ce groupe d'élèves.

4.4.4 Conclusion

Pour résumer cette analyse des effets des variables lexicales, nous reprenons les coefficients pour chacun des effets statistiques du degré d'abstraction présentés dans les deux sous-

sections qui précèdent. Nous mettons l'accent sur le degré d'abstraction parce c'est ce facteur que nous avons jugé le plus pertinent de prendre en considération dans le cadre de ce travail. Pour chaque coefficient B, nous calculons l'odds ratio à partir de l'exponentielle de ce coefficient. Ces calculs apparaissent dans le Tableau 4-13.

Tableau 4-13 Résumé des relations et interactions statistiques entre le degré d'abstraction et les autres variables lexicales, le niveau de scolarité et la langue maternelle pour les connaissances paradigmatisques (SENS) et les connaissances syntagmatiques (CONTEXTE).

Variable dépendante	Paramètre	B	OR
SENS	Abs	-0,164	0,849
	Abs*Fréq	0,022	1,022
	Abs*Long	-0,047	0,954
	Abs*Fréq*Long	-0,009	0,991
	[Niv=2] * Abs	-0,079	0,924
CONTEXTE	Abs	-0,320	0,726
	Abs*Fréq	-0,022	0,978
	Abs*Long	-0,047	0,954
	Abs*Fréq*Long	0,005	1,005
	[Niv=1] * Abs	-0,135	0,874

OR (pour Odds ratio) = $\text{Exp}(B)$

Le tableau montre les effets en termes d'odds ratio (OR) sur les chances d'obtenir 1 plutôt que 0 en fonction du degré d'abstraction des mots. Nous rappelons d'abord l'effet principal de ce facteur. Chaque degré d'abstraction de plus réduit les chances d'obtenir un 1, que ce soit pour la variable SENS ou la variable CONTEXTE : les chances de réussite plutôt que d'un échec pour un degré d'abstraction +1 donné sont de 0,849 pour la variable SENS et de 0,726 pour la variable CONTEXTE. En d'autres mots, plus le degré d'abstraction augmente, moins il y a de chance de réussite; le phénomène est plus important pour la variable CONTEXTE.

Les modèles montrent des effets d'interaction entre le degré d'abstraction et les autres facteurs lexicaux pour la dimension Sens comme pour la dimension CONTEXTE. Ces résultats ne sont pas surprenants, sachant que les facteurs lexicaux interagissent entre eux, comme le montre l'analyse dans l'annexe D.

Des interactions entre le niveau de scolarité et le degré d'abstraction apparaissent pour les deux dimensions. Les élèves de secondaire 2 ont une ordonnée à l'origine ou une pente différente des élèves des deux autres niveaux pour la dimension SENS ($OR_{[Niv2]*Abs}=0,924$). La Figure 4-26 donne une idée de la distinction entre les pentes des trois groupes, lorsque nous calculons les moyennes des résultats en fonction du degré d'abstraction et du niveau de scolarité.

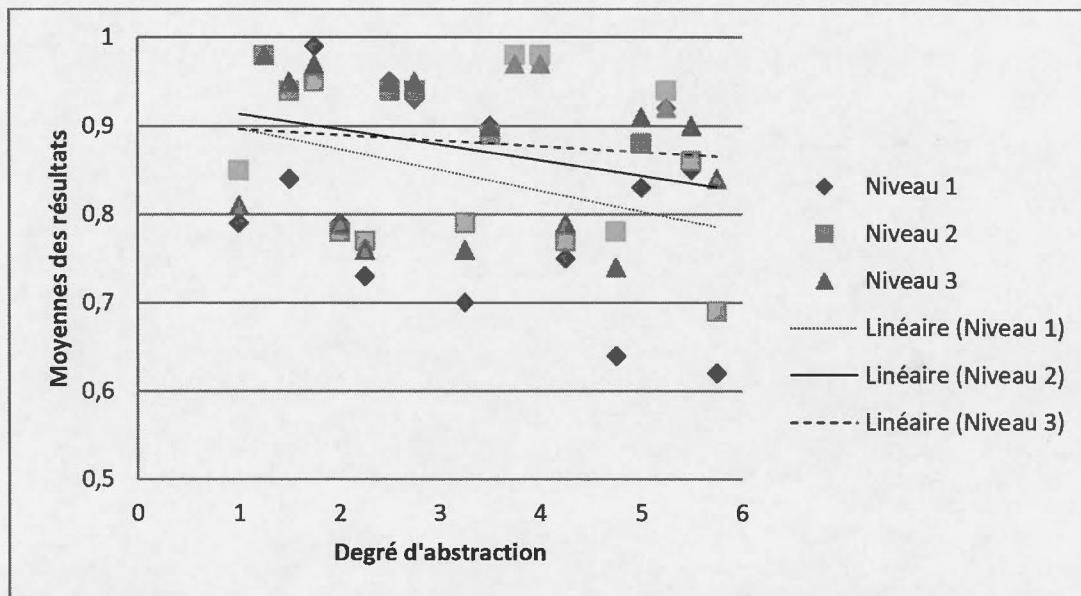


Figure 4-26 Moyennes des résultats en fonction du degré d'abstraction et du niveau de scolarité pour la dimension SENS

Les ordonnées à l'origine sont les mêmes pour les élèves de secondaire 1 et 3 (vis-à-vis le degré d'abstraction 1), mais pas pour les élèves de secondaire 2, ce qui explique peut-être l'effet d'interaction. Par contre, nous observons que les trois pentes sont tout de même négatives, ce qui illustre l'effet du degré d'abstraction pour les trois niveaux de secondaire.

Il y a également un effet d'interaction entre le niveau de secondaire 1 et le degré d'abstraction pour la dimension CONTEXTE ($OR_{[Niv1]*Abs}=0,874$). La Figure 4-27 illustre les pentes des résultats moyens des trois groupes en fonction du degré d'abstraction.

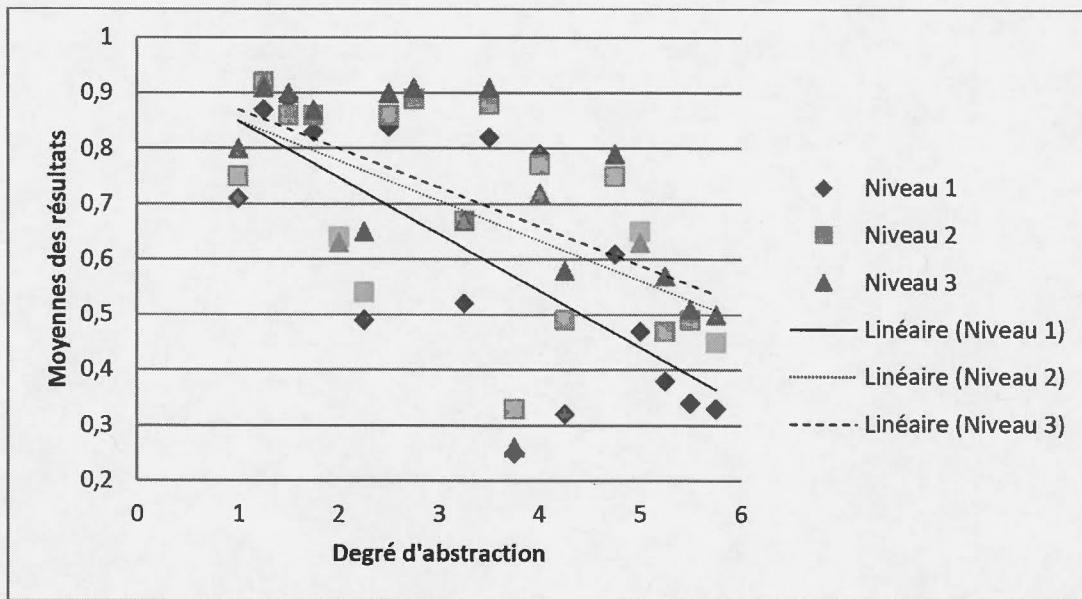


Figure 4-27 Moyennes des résultats en fonction du degré d'abstraction et du niveau de scolarité pour la dimension CONTEXTE

La figure montre une pente négative plus prononcée pour les élèves de secondaire 1, ce qui suggère un plus grand effet négatif du degré d'abstraction sur les chances de réussite pour ces élèves.

Par ailleurs, et plus intéressant dans le cadre de ce travail, il n'y a pas eu d'interaction entre la variable Langue maternelle et le facteur Degré d'abstraction, ni pour la dimension SENS ni pour la dimension CONTEXTE. Cela suggère que le degré d'abstraction pose le même degré de difficulté pour les deux populations. Nous en discutons plus amplement dans la prochaine partie.

4.5 Conclusion

Dans cette partie, nous avons présenté les résultats des mesures de l'étendue et de la profondeur de la connaissance du vocabulaire d'élèves allophones et francophones de secondaire 1 à 3. D'abord, nous avons montré que les élèves allophones comme les élèves francophones reconnaissaient autour de 90 % des noms du vocabulaire scolaire et des noms les plus fréquents du français, peu importe le niveau de scolarité ou le sexe des élèves. En ce

qui concerne la profondeur de la connaissance du vocabulaire, nous avons mesuré les connaissances paradigmatisques et syntagmatiques des élèves. Plusieurs effets statistiques ont été mesurés. Tous domaines d'apprentissages confondus, les élèves allophones ont montré des résultats plus faibles pour les connaissances paradigmatisques comme pour les connaissances syntagmatiques, particulièrement en secondaire 1 et 2. Des différences par niveau de scolarité ont été également observées, les élèves de secondaire 1 obtenant des résultats particulièrement plus faibles que les élèves de secondaire 2 et 3. Les résultats montrent un effet de la variable Sexe pour la dimension CONTEXTE chez les élèves francophones de secondaire 2 et 3, mais pas chez les allophones : les résultats des garçons francophones sont plus faibles que ceux des filles francophones. Les résultats par domaine montrent de nombreuses manifestations de plafonnement pour toutes les connaissances paradigmatisques et les mots les plus fréquents. Il y a également plafonnement dans les résultats pour les domaines Géographie et Mathématique dans le cas des connaissances syntagmatiques. Les résultats pour les connaissances syntagmatiques en français et en histoire montrent des effets de la variable Niveau scolaire : il y a amélioration statistique des résultats d'un niveau à l'autre, particulièrement pour les élèves allophones. Un effet statistique est calculé en français pour le sexe : les résultats des garçons sont plus faibles que ceux des filles, sauf pour les élèves francophones de secondaire 3, où les résultats sont semblables. Un petit effet est montré pour la variable Langue maternelle pour les mots les plus fréquents : les élèves allophones ont des résultats plus faibles que ceux des élèves francophones, particulièrement en secondaire 1 et 2. Des interactions entre la variable Niveau scolaire et Langue maternelle sont observées pour les domaines Histoire et Science et technologie, de même que pour la langue maternelle et le sexe en histoire. Par ailleurs, une grande variance dans les résultats et une faible ampleur des effets suggèrent que d'autres variables interviennent dans la connaissance en profondeur du vocabulaire.

Dans cette partie, nous avons également procédé à une analyse des relations entre des facteurs lexicaux (le degré d'abstraction des mots, leur fréquence et leur longueur en nombre de lettres) et les résultats des élèves. Pour ce faire, nous avons employé une modélisation linéaire appelée « Équations d'estimation généralisées » que nous avons appliquée à des

variables dépendantes dichotomiques (le résultat 1 ou 0 pour chacun des mots). Les Khi 2 de Wald ont montré des effets statistiques entre toutes les variables lexicales et les variables dépendantes SENS et CONTEXTE (c'est-à-dire respectivement les connaissances paradigmatisques et syntagmatiques). C'est le degré d'abstraction qui affecte le plus les chances de bien répondre : plus le degré d'abstraction d'un mot est élevé, moins grandes sont les chances de bien répondre. C'est le cas en particulier pour les élèves de secondaire 1 pour la dimension CONTEXTE. La fréquence d'un mot affecte positivement les chances de bien répondre pour la dimension SENS, mais pas, paradoxalement, pour la dimension CONTEXTE. La longueur d'un mot affecte positivement les chances de bien répondre pour la dimension SENS, ce qui est également paradoxalement. Elle affecte négativement les chances de bien répondre pour les connaissances paradigmatisques. Des interactions statistiques entre les variables lexicales ont également été mesurées. Ces interactions expliquent peut-être les résultats paradoxaux soulignés plus tôt.

Nous discutons de tous ces résultats à la lumière de la problématique et du cadre théorique dans la partie Discussion qui suit.

CHAPITRE V

DISCUSSION

5.1 Introduction

À la fin de la première partie de ce travail, nous nous sommes demandé quel était le degré d'acquisition du vocabulaire scolaire chez les élèves allophones du système scolaire francophone québécois après plusieurs années dans ce système. Ces élèves n'ont pas le français comme langue maternelle alors qu'ils doivent évoluer dans un système scolaire francophone. Ils doivent à la fois apprendre une langue, le code écrit et les domaines d'apprentissages, alors que les élèves locuteurs natifs du français ont l'avantage de déjà connaître la langue. Nous avons rappelé que, même pour des élèves dont le français est la langue première, le langage scolaire est difficile à acquérir, particulièrement son vocabulaire. Les mots à apprendre sont nombreux, souvent abstraits et principalement rencontrés à l'écrit. Cela implique que l'habileté de lecture doit être développée de même que des habitudes de lecture établies pour que l'acquisition se produise. Comme les élèves allophones connaissent peu le français quand ils commencent l'école, il leur faut plusieurs années avant d'atteindre le niveau atteint par les élèves francophones, s'ils y parviennent. Tout au long de la scolarité, la méconnaissance du vocabulaire du français peut nuire à la compréhension. Nous avons abordé les relations entre la connaissance du vocabulaire et la lecture. La connaissance de la plupart des mots d'un texte est essentielle pour lire couramment. Cette connaissance des mots passe par son étendue – un grand nombre de mots doivent être connus – et sa profondeur. Au fur et à mesure que le nombre de mots connus augmente, des liens entre les mots se tissent, approfondissant la connaissance du vocabulaire. Ces liens organisés semblent favoriser une récupération des mots pour les comprendre ou les produire efficacement et rapidement. La profondeur de la connaissance comporte plusieurs dimensions. Dans notre étude, nous avons

mis l'accent sur ses dimensions syntagmatique et paradigmatische, la première indiquant une connaissance de l'usage des mots en contexte, la seconde indiquant une connaissance du sens du mot au sein d'un système organisé, tel qu'il est présenté dans le monde scolaire. De plus, comme les mots du vocabulaire scolaire sont plus abstraits, ce facteur risque d'être un obstacle au développement du vocabulaire, puisqu'il implique, entre autres, des connaissances syntaxiques que les élèves allophones n'ont peut-être pas.

Nous avons procédé aux mesures de l'étendue et de la profondeur de la connaissance du vocabulaire et avons analysé les résultats obtenus dans la partie précédente. Dans cette partie, nous interprétons ces résultats à la lumière de la situation des élèves allophones dans le système scolaire et des recherches passées sur l'état du vocabulaire de ces élèves. Nous établissons également des liens entre les résultats obtenus et les différents objectifs poursuivis de même qu'entre les différentes variables prises en compte : le niveau de scolarité, le sexe et le degré d'abstraction des mots sont des facteurs qui ont affecté les résultats, en plus de la langue maternelle.

5.2 État de la connaissance du vocabulaire scolaire des élèves allophones

5.2.1 L'étendue

Nous avons expliqué dans le cadre théorique qu'il fallait connaître préalablement un fort pourcentage des mots d'un texte pour pouvoir le comprendre. La taille du vocabulaire étant un facteur important à considérer en lecture (Nagy et Anderson, 1984; Nation, 2001; Schmitt *et al.*, 2011; Vermeer, 2001) et celle-ci constituant une habileté fondamentale à développer pour apprendre au cours de la scolarité (Baumann et Graves, 2010), nous avons voulu mesurer l'étendue du vocabulaire des élèves allophones. Pour ce faire, nous avons développé un test en nous inspirant de ceux développés par Meara (Meara et Buxton, 1987) et Alderson (2005). Il a consisté à demander aux élèves de cocher s'ils connaissaient ou non le sens d'une séquence de lettres. Les séquences pouvaient être un nom du vocabulaire scolaire, un nom parmi les mots les plus fréquents du français ou un pseudomot. L'analyse des résultats montre que les élèves allophones connaissent autant de noms du vocabulaire scolaire et des

mots les plus fréquents que les élèves francophones. De plus, les scores élevés (autour de 90 % pour les deux populations) indiquent que ce sont la plupart des noms qui sont connus. Par contre, l'une des limites du test employé est de ne pas indiquer jusqu'à quel point un mot est connu (Read, 2000). Ce que montrent les résultats des élèves, c'est qu'ils reconnaissent les noms du vocabulaire scolaire et les mots les plus fréquents au moins sur le plan de la forme écrite. Si nous comparons les résultats obtenus au test sur l'étendue à ceux obtenus au test sur la profondeur de la connaissance, nous observons une différence importante entre les résultats. Considérant les moyennes plus grandes obtenues au test sur l'étendue ($LM1=91.75\%$, $LM2=92.02\%$) comparativement au test sur la profondeur (Sens, $LM1=87.32\%$, $LM2=84.04\%$; Contexte, $LM1=70.06\%$, $LM2=65.86\%$), le premier semble surestimer la connaissance du vocabulaire. Deux explications peuvent être proposées : soit que les élèves connaissent bien la forme des mots et les règles de formation des mots en français, soit que les élèves ont reconnu des mots apparentés dans leur langue d'origine ou dans une autre langue seconde apprise par ailleurs (l'anglais ou l'espagnol, par exemple). Cette dernière explication met en relief une autre des limites du test soulevée par Meara (1992), qui souligne la difficulté de mesurer exclusivement les mots moins fréquents de l'anglais alors qu'ils ont été pour une grande part empruntés à d'autres langues latines. Cette explication d'un effet de la connaissance des mots apparentés s'applique peut-être également aux mots du français. Par ailleurs, si les élèves n'éprouvent pas de difficulté à reconnaître les mots du français, la forme n'est qu'un aspect de la connaissance complexe que l'on peut avoir d'un mot. Son sens et son usage sont tout aussi essentiels à sa compréhension (Nation, 2001). Les résultats obtenus par les élèves au test sur la profondeur nous permettent de mieux comprendre sur lesquels de ces deux aspects les deux populations se distinguent.

5.2.2 La profondeur de la connaissance du vocabulaire

Pour évaluer la profondeur de la connaissance, nous avons choisi de mesurer les connaissances paradigmiques et les connaissances syntagmatiques. En général, les résultats obtenus à notre test sur la profondeur de la connaissance du vocabulaire appuient les résultats des autres recherches auprès des élèves allophones (Bialystok *et al.*, 2009; Droop et Verhoeven, 2003; Hacquebord, 1994; Jean et Geva, 2008; Schoonen et Verhellen, 2008;

Verhallen et Schoonen, 1998) : ceux-ci n'ont pas développé une connaissance du vocabulaire comparable à celle des élèves locuteurs natifs. Contrairement aux études précédentes sur la profondeur de la connaissance du vocabulaire (Hellman, 2011; Qian, 1999; Qian et Schedl, 2004; Read, 2000; Schoonen et Verhallen, 2008), nous avons choisi d'analyser les deux dimensions séparément. Nous avons jugé que des analyses distinctes permettraient de mieux cerner la profondeur de la connaissance des élèves. Nous discutons successivement des résultats pour les connaissances paradigmatisques, de ceux pour les connaissances syntagmatiques puis de l'écart important dans les résultats entre les deux mesures.

C'est particulièrement pendant les années scolaires que se développent les relations paradigmatisques entre les mots, l'école favorisant un tel traitement (Schoonen et Verhallen, 2008). Rappelons que les relations paradigmatisques sont des relations hiérarchiques verticales entre les mots, comme l'hyperonymie et l'hyponymie (Vermeer, 2001). À l'école, les élèves développent leur lexique en étant confronté à de nouveaux sens et de nouvelles relations entre les mots. Ces nouveaux sens et relations sont développés par généralisation, catégorisation et abstraction (Schoonen et Verhallen, 2008). Dans le cadre de notre étude, pour un nom cible dans un domaine d'apprentissages donné, les élèves devaient identifier deux mots avec lequel le mot cible partage en partie son sens. Les mots à choisir étaient des synonymes, hyperonymes ou hyponymes du mot cible.

L'analyse montre une différence statistique entre les résultats des élèves allophones et ceux des élèves francophones. Ces derniers obtiennent en moyenne un score plus élevé (87.32 %) que celui des élèves allophone (84.04 %). Cette différence de trois points de pourcentage entre les deux populations cache des distinctions et des ressemblances plus importantes quand nous prenons en compte l'analyse par domaine et le niveau de scolarité (voir les sections 5.3 et 5.4). Cependant, en général, les élèves allophones n'ont pas développé leurs connaissances paradigmatisques des noms du vocabulaire scolaire au même niveau que les élèves francophones. Nous discutons de l'impact éventuel de ces résultats sur la compréhension écrite dans la section 5.9.

Bien que les résultats obtenus montrent une différence entre les deux populations, nous observons qu'ils sont, dans l'ensemble, plutôt bien réussis. Nous avons déjà rappelé à quel point le monde scolaire favorisait l'organisation des noms dans des relations d'antonymie, de synonymie, de gradation, de coordination, d'hyperonymie et d'hyponymie (Schoonen, Verhellen, 2008 ; Qian, 2002 ; Vermeer, 2001). Nous avons également rappelé qu'après la petite enfance, cette façon d'organiser les mots prenait de l'ampleur et se poursuivait à l'âge adulte (Meara, 2009; Singleton, 1999). A la lumière de nos résultats, il semble que le développement des relations paradigmatisques entre les mots du vocabulaire scolaire ne pose pas de problème particulier chez les élèves du secondaire, peu importe la langue première des élèves. Ce n'est pas le cas avec les relations syntagmatiques.

Les connaissances syntagmatiques sont les connaissances syntaxiques et sémantiques mises en œuvre pour se représenter le sens d'un mot en contexte avec d'autres mots (Gleitman *et al.*, 2005 ; Wray, 2002). Elles peuvent également être des blocs de mots mémorisés apparaissant fréquemment dans une même séquence, c'est-à-dire les collocations (Metin et Karaoglan, 2011; Nation, 2001; Ward, 2007 ; Wray, 2002). Les mots apparaissent rarement de façon isolée et leur sens est précisé par le contexte dans lequel ils apparaissent (Nation, 2001). Le locuteur dont les connaissances syntagmatiques sont développées traite plus efficacement et rapidement les mots dans leur contexte, ce qui, en retour, contribue favorablement à la compréhension (Nation, 2001 ; Ward, 2007).

Pour les connaissances syntagmatiques, l'analyse des résultats révèle une différence statistique entre les élèves allophones et les élèves francophones. Les élèves allophones obtiennent en moyenne un score inférieur à celui des élèves francophones. Comme pour les connaissances paradigmatisques, nous constatons que les élèves allophones n'ont pas développé des connaissances syntagmatiques au même niveau que celles des élèves francophones. Considérant que ce sont des unités plus grandes qui sont traitées avec moins de succès, nous nous interrogeons sur l'impact que cet écart peut avoir sur la compréhension des textes des domaines d'apprentissages par les élèves allophones. Nous en discutons dans la section 5.9. Par ailleurs, nous observons que les scores moyens des connaissances syntagmatiques sont nettement inférieurs à ceux obtenus pour les connaissances

paradigmatiques. Ces dernières semblent plus faciles à développer pour les élèves, qui n'ont besoin d'établir qu'une relation linéaire et hiérarchique entre les mots d'une même classe (Schoonen et Verhallen, 2008 ; Vermeer, 2001 ; Miller et Fellbaum, 1991, dans Read, 2004). Pour les connaissances syntagmatiques, des relations entre mots de différentes classes doivent être établies, ce qui implique d'appliquer des connaissances syntaxiques (Gleitman *et al.*, 2005). De plus, comme nous avons sélectionné les réponses à notre test en fonction de la fréquence des collocations, les résultats des élèves illustrent peut-être leur expérience moindre en lecture.

La limite que nous rencontrons dans notre pouvoir d'interprétation des résultats des élèves pour les connaissances syntagmatiques est de savoir s'ils sont le reflet de leur degré de connaissances syntaxiques et sémantiques, de leur expérience de lecture ou des deux. Étudier plus en profondeur les connaissances syntagmatiques mérite dès lors d'être envisagé. De plus, comme les mots des disciplines apparaissent fréquemment dans des syntagmes complexes (Ward, 2007), étudier ces constructions sur les plans de la forme comme du sens permettrait de mieux comprendre comment les élèves traitent les unités plus grandes que le mot.

5.3 La connaissance du vocabulaire scolaire et le niveau de scolarité

Comme l'a montré l'analyse des résultats, la différence entre les résultats des élèves allophones et celle des élèves francophones doit être pondérée en fonction du niveau de scolarité. De façon générale, nous nous attendons à ce que les élèves, au cours de leur scolarité, développent leur vocabulaire scolaire, peu importe leur langue maternelle (Droop et Verhoeven, 2003; Jean et Geva, 2008; Schoonen et Verhallen, 2008). En fait, l'ampleur de l'effet calculée à l'aide de SPSS montre que c'est le niveau de scolarité qui est le facteur le plus important dans les différences entre les moyennes obtenues. Ce résultat corrobore celui obtenu par d'autres études, où le temps ou l'expérience scolaire sont un facteur plus important du développement du vocabulaire que la langue maternelle (Droop et Verhoeven, 2003; Jean et Geva, 2008; Schoonen et Verhallen, 2008). Dans nos résultats, l'effet du niveau de scolarité est le double de l'effet de la langue maternelle pour la dimension paradigmique

(6 % comparativement à 3 %) et il est plus de cinq fois plus important que celui de la langue maternelle pour la dimension syntagmatique (respectivement 17 % comparativement à 3 %).

Bien que notre travail ne soit pas une étude longitudinale, les résultats d'une année à l'autre suggèrent une progression dans les connaissances, particulièrement pour les élèves allophones. Les élèves francophones plafonnent dans les connaissances paradigmatisques, mais les profils des deux populations suggèrent une progression d'année en année pour les connaissances syntagmatiques. Le saut entre la première et la deuxième secondaire est le plus important. Le temps, l'expérience avec la langue ou la scolarité ont donc un effet positif sur le développement du vocabulaire scolaire des élèves. C'est le cas plus particulièrement des élèves allophones qui semblent rattraper les élèves francophones, sans les rejoindre toutefois en ce qui concerne les connaissances paradigmatisques. Par contre, un effet d'interaction entre le niveau de scolarité et la langue maternelle révélé par l'analyse des résultats suggère que les élèves allophones ont rejoint les élèves francophones en secondaire 3 en ce qui concerne les connaissances syntagmatiques. Il faudrait vérifier s'il s'agit là de l'effet de connaissances lexicales comparables ou d'une expérience de lecture semblable, comme nous l'avons abordé dans la section précédente.

Pour les deux dimensions de la profondeur de la connaissance, nous observons une différence marquée entre la première et la deuxième secondaire. Il s'agit peut-être de l'effet de certaines matières enseignées en deuxième secondaire, qui permettent aux élèves d'améliorer leurs connaissances du vocabulaire.

5.4 La connaissance du vocabulaire et le sexe de l'élève

Le sexe des élèves est souvent cité comme facteur de réussite scolaire. Selon le MELS (MELS, 2010), les garçons réussissent moins bien en lecture. Par contre, ils réussissent mieux en science et en mathématiques que les filles. Dans la problématique, nous avons résumé l'analyse du MELS (2006) en fonction de certains paramètres comme le sexe, la langue maternelle et le lieu de naissance. En ce qui concerne le sexe, le retard scolaire est plus important pour les garçons que pour les filles, qu'ils soient ou non issus de l'immigration.

Nous avons donc vérifié si nous obtenions le même portrait des garçons et des filles dans notre étude. Dans la mesure des connaissances paradigmatisques, il n'y a pas eu de différences statistiques entre les sexes, peu importe la langue maternelle. Les garçons et les filles ont développé des connaissances paradigmatisques comparables.

En ce qui concerne les connaissances syntagmatiques, l'analyse a révélé des différences statistiques. L'ampleur de l'effet sexe est très petite. Pour l'ensemble des domaines, c'est en deuxième et troisième secondaire que l'effet se manifeste : les filles francophones obtiennent de meilleurs résultats que les garçons alors que les deux sexes chez les élèves allophones sont à égalité. Dans les résultats par domaine, en français, les garçons francophones sont plus faibles que les filles, en secondaire 1 et 2, mais pas en secondaire 3. En histoire, une faible interaction entre le niveau et le sexe montre que les garçons allophones sont meilleurs que les filles allophones, alors que les filles francophones sont meilleurs que les garçons francophones en secondaire 2 et 3.

En général, nous constatons que lorsqu'il y a un effet sexe, il est de très faible ampleur et ne concerne que peu de domaines d'apprentissages. Ces résultats sont différents des résultats du MELS (2006) qui montrent des résultats plus faibles pour les garçons peu importe la langue maternelle. Il nous est impossible d'expliquer cette différence avec les données que nous avons. Éventuellement, si nos résultats ne sont pas dus au hasard ou à une particularité de notre échantillon, il faudrait vérifier si un facteur culturel ou comportemental n'entre pas en jeu. Une plus grande motivation pour la réussite scolaire chez les élèves issus des minorités ou leurs parents pourraient expliquer nos résultats (Downie, Gaudreau, Koestner et Magneau, 2008).

5.5 La connaissance du vocabulaire et les domaines d'apprentissages

Dans la problématique, nous nous sommes interrogé sur la réussite scolaire des élèves allophones. Un rapport du MELS (2006) fait état du retard scolaire chez les élèves allophones. En général, les élèves issus de l'immigration sont proportionnellement plus nombreux à présenter un retard scolaire. Selon une autre référence, plus ancienne, les élèves

allophones montrent un rendement scolaire comparable à celui des élèves dont la langue maternelle est le français ou l'anglais : les résultats aux épreuves uniques du secondaire et les taux de réussite sont semblables (MEQ, 1998). Dans certains domaines, les résultats des élèves allophones sont même plus élevés que ceux des autres élèves. Par contre, certains groupes linguistiques éprouvent plus de difficultés « et affichent des résultats nettement inférieurs à la moyenne des élèves » (MEQ, 1998, p. 10). Pour vérifier si cette description se reflétait dans les résultats de notre étude, nous avons comparé ceux-ci en fonction des domaines d'apprentissages et des mots les plus fréquents du français.

Les résultats à la mesure des connaissances paradigmatisques vont dans le même sens que ce que le rapport du MEQ (1998) suggère. Les profils des élèves sont semblables d'un domaine à l'autre. Dans tous les domaines sauf en science et technologie, il y a plafonnement des résultats, pour les deux populations, dès la deuxième secondaire. Par ailleurs, nous observons que ce sont les mots fréquents qui sont le mieux réussis, peu importe le niveau de scolarité ou la langue maternelle. Nous rappelons que ces observations ne sont pas appuyées par une analyse statistique à cause des manifestations de plafonnement. Il vaudrait la peine d'inclure dans nos mesures des mots des domaines des cycles ultérieurs pour circonvenir l'effet de plafonnement et mieux comprendre le développement des relations paradigmatisques par domaine chez les élèves.

En ce qui concerne les résultats des élèves pour les connaissances syntagmatiques, nous soulignons une différence statistique entre les élèves allophones et les élèves francophones en histoire et pour les mots les plus fréquents. En histoire, les élèves allophones réussissent moins bien. Cela pourrait s'expliquer par le fait que ces élèves se sentent peu concernés par les sujets discutés. Selon le MEQ :

les élèves d'autres cultures se reconnaissent peu dans certains contenus d'enseignement qui leur sont transmis à cause du manque de perspectives inclusives de ces derniers, ce qui peut nuire au développement de leur sentiment d'appartenance. D'autre part, on ne facilite pas suffisamment l'accès de ces élèves à une information sur le patrimoine du Québec, son histoire, ses valeurs et ses coutumes. Ils et elles ont peu de repères sur l'évolution de la société

québécoise ainsi que sur leur propre participation à la construction et à l'enrichissement de ce patrimoine (MEQ, 1998, p. 13).

Pour les mots les plus fréquents, les résultats plus faibles des élèves allophones suggèrent que ceux-ci ont davantage développé leurs connaissances syntagmatiques pour le vocabulaire scolaire que celles pour le vocabulaire plus domestique du français. Ces résultats vont dans le même sens que ceux de (Bialystok *et al.*, 2009), qui montrent que les élèves bilingues du primaire ont une meilleure connaissance des mots du vocabulaire scolaire que des mots de la conversation courante. Ces connaissances plus faibles pourraient avoir un impact négatif sur la compréhension chez les élèves allophones, considérant l'importance des mots fréquents dans les textes.

5.6 Le vocabulaire scolaire et le lexique du locuteur bilingue

Dans le cadre théorique, nous avons présenté quelques modèles du lexique de la personne bilingue (Dijkstra et van Heuven, 2002; Jiang, 2000; Kroll et Sunderman, 2003). Nous avons décrit comment les mots de langues différentes étaient traités et représentés en mémoire et comment leurs représentations, et lesquelles, étaient reliées entre elles. Cette présentation des modèles nous a permis de comprendre comment les apprenants d'une langue seconde développaient leur lexique dans une autre langue en plus de leur langue maternelle. Il semble que les lexiques de la langue première et d'une langue seconde d'un même locuteur sont connectés entre eux, mais ils ne forment pas un seul et même lexique (Kroll et Sunderman, 2003; Singleton, 1999). De plus, il semble que les relations entre un mot donné en L2 et un autre en L1 vont varier d'un individu à l'autre, tout dépendant du contexte dans lequel le mot est traité, comment il a été appris, jusqu'à quel point il est connu et dans quelles mesures les liens sémantiques sont communs d'une langue à l'autre (Kroll et Sunderman, 2003; Singleton, 1999). Ces modèles sont fondés sur des études auprès d'apprenants d'une langue seconde scolarisés dans leur langue première. Une grande importance est accordée dans ces modèles aux ponts qui doivent d'abord être établis entre les mots de la L1 et ceux de la L2. Or, nous nous sommes demandé si des apprenants allophones pouvaient développer leurs représentations des mots scolaires, en L2 donc, sans avoir une représentation de ces mots dans leur langue première.

Les résultats obtenus dans le cadre de notre travail donnent quelques indices du lexique développé par les élèves allophones. Dans le modèle de Jiang (2000), le développement de la connaissance d'un mot dans une langue seconde se déroule en trois étapes. Au cours de la première étape, l'apprenant consacre surtout ses efforts à une représentation phonologique et orthographique du mot rencontré dans la L2. Les résultats de notre étude au test sur l'étendue du vocabulaire suggèrent que les élèves allophones ont franchi cette étape. Au cours d'une deuxième étape, le nombre d'associations entre le mot de la L2 et sa traduction dans la L1 augmente, au point d'être représenté dans les dimensions sémantiques et syntaxiques du mot de la L1. Ces ponts entre les représentations d'un mot dans la L2 et la L1 sont également illustrés dans le modèle hiérarchique modifié de Kroll et Stewart (1994). Comme nous n'avons pas pris en compte les connaissances lexicales des élèves allophones dans leur langue première, il nous est impossible de vérifier si nos résultats concordent avec ces modélisations. Dans une troisième étape, le mot est entièrement représenté par ses caractéristiques dans la L2. Il s'agit de l'étape d'intégration, où les représentations et le traitement du mot en L2 ressemblent à celle d'un mot en L1. Encore une fois, nous ne pouvons savoir si les résultats des élèves allophones de notre étude reflètent un lexique de mots intégrés et plus autonomes en L2. Il manque à notre étude une mesure de la connaissance du vocabulaire scolaire dans la L1 pour mieux comprendre la construction de ce vocabulaire dans la L2. Dans notre étude, de nombreux élèves ont déclaré lire dans une langue autre que le français et l'anglais. Cette pratique de la lecture dans une autre langue les a peut-être aidés à développer la profondeur de leur connaissance du vocabulaire en français. Pour le vérifier, nous pourrions comparer les résultats des élèves allophones qui lisent dans leur langue première à ceux qui ne lisent pas dans cette langue. Un autre indice qui pourrait éclairer la situation est le temps de réponse des élèves en fonction de la langue maternelle et du degré d'abstraction des mots. Une interaction entre la langue maternelle et le degré d'abstraction des mots sur le temps de réponse pourrait révéler un traitement différent entre les élèves allophones et les élèves francophones. Nous prévoyons étudier cette variable dans un travail ultérieur. Éventuellement, nous voudrions pouvoir répondre à la question : jusqu'à quel point le vocabulaire scolaire d'un apprenant allophone peut-il se développer quand il n'est pas développé dans sa langue première? Rappelons que l'étude de Bialystok *et al.*

(2009) a montré que les élèves bilingues avaient développé un lexique pour la communication à la maison dans la langue première et un lexique dans la langue seconde pour la communication et l'apprentissage scolaires. Nous prévoyons poursuivre nos recherches dans cette direction.

5.7 La connaissance du vocabulaire scolaire et le degré d'abstraction des mots

Dans le cadre de notre étude, nous avons établi que le degré d'abstraction était un facteur qu'il fallait prendre en considération. D'abord, un grand nombre de mots du vocabulaire scolaire sont abstraits (Ward, 2007). Ensuite, les mots abstraits ne sont pas appris ni traités aussi facilement et rapidement que les mots plus concrets (Altarriba *et al.*, 1999; De Groot et Keijzer, 2000). Les mots concrets sont plus facilement appris et récupérés parce qu'ils peuvent être entreposés dans deux systèmes de représentation, un système de représentations sensoriels, à l'accès plus rapide, et le système verbal ; les mots abstraits, eux, sont entreposés dans le seul système verbal (Altarriba *et al.*, 1999; Holcomb *et al.*, 1999). Pour apprendre puis traiter les mots abstraits, des connaissances préalables sont requises, dont des connaissances syntaxiques (Gleitman *et al.*, 2005). Comme les élèves allophones sont moins susceptibles d'avoir développé les connaissances préalables, particulièrement les connaissances syntaxiques de la langue d'enseignement (August et Shanahan, 2006 ; Morris et Labelle, 2008), nous avons proposé qu'ils risquaient d'éprouver des problèmes dans l'apprentissage des mots abstraits. D'autant plus que les mots abstraits d'une langue seconde n'ont pas nécessairement les mêmes représentations que le mot traduit en langue première, alors que c'est plus souvent le cas pour les mots concrets (De Groot et Keijzer, 2000; Meuter, 2009).

Dans un premier temps, nous nous sommes assuré que les mots des domaines d'apprentissages de notre propre sélection pour le test sur la profondeur de la connaissance du vocabulaire étaient plus abstraits que les mots les plus fréquents (voir section 3.6.2). Par ailleurs, nous observons que ce sont les mots des domaines Géographie et Mathématiques qui sont les plus concrets parmi les mots des domaines d'apprentissages. Les résultats montrent des scores plus élevés dans ces domaines (en plus de ceux des mots les plus fréquents, qui

sont en moyenne les plus concrets) pour les deux dimensions de la profondeur de la connaissance.

L'analyse présentée dans la partie précédente a cependant porté sur les résultats des élèves pour chacun des mots en fonction du degré d'abstraction de ceux-ci, sans prendre en compte le domaine d'apprentissages. Les résultats ont montré que chaque degré d'abstraction de plus réduisait les chances de réussir l'item, autant pour les connaissances paradigmatisques que pour les connaissances syntagmatiques. Il n'y a pas eu d'interaction entre la variable langue maternelle et le degré d'abstraction. Cela suggère que les élèves allophones n'éprouvent ni plus, ni moins de difficulté avec les mots abstraits que les élèves francophones. Le degré d'abstraction est donc un facteur de difficulté pour tous les élèves, peu importe leur langue maternelle. En ce qui concerne les élèves allophones, considérant la faible correspondance entre des mots abstraits d'une L1 et ceux d'une L2 (Pavlenko, 2009), ces élèves ont peut-être construit leur lexique des mots abstraits du langage scolaire en français directement dans la mémoire conceptuelle, sans passer par les représentations dans la langue maternelle (De Groot et Keijzer, 2000). L'absence d'interaction entre les résultats selon le degré d'abstraction et la langue maternelle justifie cette interprétation. Selon nous, si les élèves allophones avaient construit des représentations des mots abstraits du français fondées sur des mots correspondants dans leur L1, ils auraient obtenus encore moins de réussite en fonction du degré d'abstraction, considérant que ces mots ne couvrent pas la même réalité dans une langue ou dans l'autre (De Groot et Keijzer, 2000; Pavlenko, 2009). De plus, les résultats de Verhallen et Shoonen (1998) montrent que des élèves allophones peuvent développer une plus grande profondeur du vocabulaire dans la L2 que dans la L1. Cela suggère que la connaissance des mots dans la L1 n'est peut-être même pas nécessaire au développement d'une connaissance en profondeur du vocabulaire dans la L2. Cependant, nous avons besoin d'étudier le phénomène plus précisément, en prenant en compte la connaissance du vocabulaire dans la L1, les connaissances syntaxiques dans la L2 ou en mesurant les temps de réponse aux items du test sur la profondeur de la connaissance en fonction des variables de la langue maternelle et du degré d'abstraction des mots.

Par ailleurs, l'analyse a montré un effet d'interaction assez élevé entre le degré d'abstraction des mots et le niveau de secondaire 1. L'observation des pentes en fonction du degré d'abstraction et du niveau de scolarité (voir la partie Analyse des résultats) suggère que le traitement des mots plus abstraits avaient encore moins de chance de succès pour les élèves de secondaire 1 que pour les autres élèves. Cette observation n'est pas surprenante considérant que les élèves de secondaire 2 et 3 auront eu plus d'expérience langagière avec les mots scolaires dans le cadre de leurs cours que les élèves de secondaire 1.

5.8 Le vocabulaire scolaire et le contexte sociolinguistique

Dans la problématique, nous avons présenté différents contextes de bilinguisme pour dégager des facteurs susceptibles d'influencer l'acquisition du vocabulaire scolaire. Nous avons expliqué que des enfants exposés à deux langues pendant les années de développement du langage, apprises simultanément et se développant en remplissant les mêmes fonctions, dont les fonctions liées à la littératie, développent alors un bilinguisme équilibré et additif (Hamers, 2004). Ceux qui apprennent d'abord leur langue maternelle puis une langue seconde, soit dans le voisinage, soit à l'école peuvent développer également un bilinguisme équilibré si les deux langues sont valorisées dans le réseau social de l'enfant (Hamers, 2004). Par contre, un enfant peut apprendre deux langues consécutivement, mais sans apprendre les fonctions de la langue première relatives à la littératie (Hamers, 2004). De plus, la langue première peut être dévalorisée socialement dans la famille, alors que la langue seconde est la langue privilégiée dans la société (Hamers, 2004). Les enfants peuvent alors développer un bilinguisme soustractif (Hamers, 2004). La langue première cesse d'être apprise et peut s'atrophier avec le temps et la langue seconde ne se développe pas au-delà d'un seuil critique qui permet le développement cognitif, scolaire et linguistique (Cummins, 2000). Dans notre travail, nous aurions voulu pouvoir mesurer les effets d'un tel bilinguisme soustractif sur l'acquisition du vocabulaire. Mais les résultats de la population étudiée et le milieu scolaire dans lequel elle évolue suggèrent que nous avons des élèves qui semblent développer un bilinguisme additif. C'est le cas du moins pour une grande part d'entre eux, puisque la majorité lit dans sa langue première (voir la partie Méthodologie). Le milieu scolaire (et probablement parental) dans lequel évolue la population étudiée privilégie et favorise une

haute fréquence d'étude et de lecture. Et les résultats des élèves aux tests sur l'étendue et sur la profondeur de la connaissance du vocabulaire montrent des résultats élevés ou en progression d'un niveau scolaire à l'autre. Pour mesurer l'effet éventuel d'un bilinguisme soustractif sur l'acquisition du vocabulaire, il faudrait étudier les populations allophones d'autres milieux scolaires.

Par contre, l'un des grands apports de notre étude est de montrer que, même lorsque des élèves allophones proviennent d'un milieu qui privilégie un bilinguisme additif et qui offre des conditions socioéconomiques moyennes à élevées, ils parviennent à peine à développer un vocabulaire scolaire équivalent à celui des élèves locuteurs natifs au bout de la troisième secondaire. Certaines recherches auprès des élèves allophones se sont produites dans des milieux scolaire défavorisés (Droop et Verhoeven, 2003; Hacquebord, 1994; Schoonen et Verhellen, 2008). D'abord, nous nous interrogeons sur les résultats éventuels d'élèves allophones québécois provenant de milieux plus défavorisés, considérant l'impact négatif de ces conditions sur le développement langagier (Chall *et al.*, 1990; Cummins, 2000; Droop et Verhoeven, 2003; Hacquebord, 1994). Ensuite, nos résultats nous amènent à constater que le fait de poursuivre sa scolarité quand la langue d'enseignement n'est pas la langue maternelle est un facteur qui retarde l'apprentissage du vocabulaire scolaire, peu importe les conditions sociolinguistiques ou socioéconomiques.

5.9 L'impact éventuel de la connaissance du vocabulaire sur la lecture

Nous avons déjà expliqué que l'un des principaux facteurs de la compréhension est la connaissance préalable des mots du texte à lire (Schmitt *et al.*, 2011). En lisant, un lecteur peut inévitablement rencontrer des mots qu'il ne connaît pas. Plus il y a de mots inconnus dans un texte lu, moins la compréhension a des chances de se produire, soit parce que le contenu du texte est oublié pendant que le lecteur cherche la signification des mots inconnus, soit parce que le nombre de mots inconnus ne permet simplement pas le traitement du texte (Schmitt *et al.*, 2011). Le pourcentage de mots qui doivent être préalablement connus pour lire un texte sans nuire à la compréhension varie de 95 à 98 % (Schmitt *et al.*, 2011). Les mots d'un texte sont principalement constitués de mots très fréquents (les 2000 mots les plus

fréquents constituent autour de 80 % des mots d'un texte, tout dépendant de la langue), d'un petit pourcentage de mots plus rares (Cobb et Horst, 2004) et, dans le cas des textes à lire dans le cadre scolaire, d'un autre petit pourcentage de mots des domaines d'apprentissages (Coxhead, 2000 ; Ward, 2007).

L'une des raisons pour laquelle nous nous sommes intéressé à la mesure de la connaissance du vocabulaire chez les élèves allophones est l'impact négatif que pourrait avoir des résultats plus faibles sur la compréhension et la production écrite des élèves. (Nous discutons au sujet de la production écrite dans la prochaine section.) Dans le cadre théorique, nous avons expliqué qu'à partir de la quatrième année du primaire, la tâche des élèves en lecture n'était plus d'apprendre à lire, mais à apprendre à partir de ce qui était lu (Chall *et al.*, 1990). Le contenu des textes lus va au-delà de ce qui est déjà su, linguistiquement et cognitivement, par le lecteur, avec des nouveaux mots et de nouvelles idées. Cependant, nous avons rappelé que de nombreux apprenants de certains milieux socioculturels n'ont pas accès à ce vocabulaire en dehors de l'école (Corson, 1997). En conséquence, en plus de constituer une part des activités en classe, la lecture est alors la première source d'apprentissage pour plusieurs élèves au-delà de la salle de classe (Schmitt *et al.*, 2011).

Le succès de la compréhension en lecture dépend d'une connaissance du vocabulaire en quantité (il faut connaître beaucoup de mots) et en profondeur (il faut savoir ce que le mot veut dire dans le contexte et il faut pouvoir construire des unités de sens plus grandes que le mot en traitant les séquences de mots dans lesquelles le mot apparaît) (Read, 2000; Schmitt *et al.*, 2011). Les résultats au test sur l'étendue ont montré que les élèves connaissaient la grande majorité des mots les plus fréquents et le vocabulaire scolaire. Cependant, nous avons expliqué que le test surestimait probablement la connaissance de ces mots, considérant les résultats obtenus à la mesure de la profondeur de la connaissance. Les résultats obtenus par les élèves pour cette dernière donnent une meilleure idée de l'état des connaissances du vocabulaire des mots fréquents et des mots des domaines d'apprentissages des élèves allophones. Rappelons brièvement ces résultats. En général, les élèves allophones ont des connaissances paradigmiques et syntagmatiques plus faibles que les élèves francophones.

De plus, ils ont des connaissances syntagmatiques plus faibles pour les mots les plus fréquents.

Quels sont les conséquences que peut avoir une plus faible connaissance du vocabulaire ? Qian et Schedl (2004) et Qian (2002) ont montré que les résultats obtenus avec une mesure de la profondeur de la connaissance du vocabulaire sont fortement corrélés positivement avec une mesure de la compréhension. Une connaissance en profondeur plus faible chez élèves allophones annonce des résultats plus faibles en lecture, particulièrement dans le traitement de la microstructure du texte (Qian et Schedl, 2004). Les élèves allophones comprennent sans doute le sens global du texte, mais éprouvent probablement des difficultés à saisir toutes les nuances au niveau du syntagme ou de la phrase. Même si la différence entre la connaissance des élèves allophones et celle des élèves peut sembler faible (environ 3 % en moyenne pour les connaissances paradigmatiques et 5 % pour les connaissances syntagmatiques), il suffit de ne pas connaître un ou deux mots dans un texte, particulièrement s'ils sont cruciaux pour sa compréhension, pour hypothéquer celle-ci (Schmitt et al, 2011). Pour les élèves allophones, un ou deux mots connus en moins peut donc faire la différence entre la compréhension ou pas d'une partie du texte.

De plus, rappelons que les mots des disciplines apparaissent fréquemment dans des syntagmes complexes avec d'autres noms ou adjectifs spécialisés également (Ward, 2007). Ces collocations peuvent être plus spécialisées et spécifiques à un domaine d'apprentissages que le mot seul (Ward, 2007). Les élèves doivent connaître autant les mots pris individuellement que le sens global des collocations s'ils veulent comprendre un texte. Le fait que les résultats des élèves allophones à la mesure des connaissances syntagmatiques rejoignent ceux des élèves francophones en secondaire 3 fait peut-être en sorte qu'ils comprennent dès lors mieux la microstructure des textes. Mais l'impact des connaissances syntagmatiques comparativement à celui des connaissances paradigmatiques sur la lecture reste à démontrer. Par ailleurs, nous ne savons pas quelles conséquences ont les difficultés dans le traitement de la microstructure des textes sur le développement des connaissances scolaires, connaissant le rôle de la lecture à ce sujet. C'est une question sur laquelle il faudrait se pencher.

5.10 L'impact éventuel de la connaissance du vocabulaire sur l'écriture

Nous n'avons pas traité spécifiquement de la production écrite dans notre travail. Nous tenons tout de même à souligner les conséquences possibles de la profondeur de la connaissance plus faible des élèves allophones sur leur production écrite. Dans la définition de ce que c'est que de connaître un mot, nous avons rappelé la distinction établie par Nation entre la compréhension et la production d'un mot (Nation, 2001). Savoir produire un mot, c'est entre autres pouvoir produire des synonymes et des antonymes de ce mot, pouvoir utiliser le mot correctement dans une phrase originale et pouvoir utiliser les mots qui accompagnent habituellement le mot. Dans ce dernier cas, nous avons déjà énoncé que l'emploi approprié des collocations par un locuteur est souvent un critère pour l'identifier comme locuteur natif (Nation, 2001). Nous avons également rappelé que les collocations sont une des relations parmi d'autres associées à une compréhension et à une production appropriée du vocabulaire (Nation, 2001). Considérant que la production est généralement considérée une tâche plus difficile que la compréhension, nous pressentons que les élèves allophones puissent être limités dans la production de mots dans leur contexte approprié ou dans l'usage d'une plus grande variété de synonymes, faute d'une connaissance en profondeur du vocabulaire suffisante. Ce ne sont que des hypothèses, mais elles méritent d'être explorées.

CONCLUSION

Dans cette étude, nous avons voulu mesurer le degré d'acquisition du vocabulaire scolaire chez les élèves allophones évoluant au sein du système scolaire francophone québécois. Pour ce faire, nous avons établi trois objectifs de recherche : mesurer l'étendue du vocabulaire scolaire des élèves allophones ; mesurer la profondeur de la connaissance de ce vocabulaire, plus particulièrement les connaissances paradigmatisques et les connaissances syntagmatiques ; finalement, mesurer la relation entre différents facteurs lexicaux et la profondeur de la connaissance, plus particulièrement l'effet du degré d'abstraction des mots. En ce qui concerne le premier objectif, les résultats montrent que les élèves allophones du secondaire connaissent autant de mots que leurs pairs francophones, du moins, sur le plan de la forme écrite des mots. Pour ce qui est du deuxième objectif, les résultats dépendent principalement de la langue maternelle, du niveau de scolarité et de la dimension de la profondeur mesurée. Les connaissances paradigmatisques des élèves allophones sont relativement élevées, mais systématiquement plus faibles que celles de élèves francophones. Les connaissances syntagmatiques des deux populations sont plus faibles que les connaissances paradigmatisques et celles des élèves allophones sont plus faibles que celles des élèves francophones, sauf en troisième secondaire. A la lumière de ces résultats, nous constatons que les élèves allophones mettent de nombreuses années avant d'obtenir, quand ils l'obtiennent, une profondeur de la connaissance du vocabulaire comparable à celle des élèves francophones. En ce qui concerne le troisième objectif, les résultats montrent l'effet statistique du degré d'abstraction des mots dans le traitement des relations paradigmatisques et syntagmatiques. Par contre, cet effet n'est pas plus important chez les élèves allophones. Nous avons proposé que ces derniers aient construit un lexique du français langue seconde comparable à celui des élèves francophones, soit à l'aide de leur lexique en langue première, soit indépendamment de celui-ci. Nous devons procéder à de nouvelles recherches intégrant les connaissances en langue première pour mieux comprendre comment le vocabulaire des

mots abstraits est construit chez les élèves allophones. Les résultats obtenus dans le cadre de cette étude montrent que la connaissance du vocabulaire scolaire d'élèves allophones d'un milieu socioéconomique plutôt favorisé accuse un retard sur les élèves francophones au moins jusqu'en secondaire trois. Nous nous interrogeons sur la connaissance du vocabulaire scolaire développée chez des élèves allophones de milieu socioéconomique moins favorisé. Jusqu'à quel point le facteur socioéconomique interagit-il avec le facteur de la langue maternelle ? Par ailleurs, les liens établis entre la connaissance du vocabulaire et la compréhension à l'écrit et la production écrite nous amène à nous interroger également sur l'impact de ce retard sur la performance dans les deux habiletés.

Dans cette étude, des résultats importants ont été obtenus, mais certaines lacunes et observation doivent être soulignées. Elles concernent des aspects de la problématique, du cadre théorique, de la méthodologie et des résultats. Nous en discutons tour à tour dans les paragraphes qui suivent. De plus, les résultats et les observations dont il est question ici entraînent des perspectives de recherche que nous énumérons également.

Une grande partie de notre problématique et de notre cadre théorique met l'accent sur la situation des élèves allophones dans une perspective de bilinguisme soustractif. Nous avons réalisé notre étude auprès d'élèves allophones de milieu socio-économique moyen à élevé évoluant dans d'une école privée. Leurs réponses au questionnaire sociodémographique et leur milieu d'origine n'en font pas une population qui se développe dans une telle perspective de bilinguisme. Nous n'avons donc pas pu mesurer l'impact d'un bilinguisme soustractif sur la connaissance du vocabulaire scolaire chez des élèves allophones poursuivant leur scolarité au secondaire. Par contre, les résultats obtenus dans notre étude auprès d'élèves évoluant probablement dans un contexte de bilinguisme équilibré laissent prévoir des résultats pires pour des élèves allophones qui n'ont pas développé de littératie dans leur langue première ou chez qui les langues, première ou seconde, sont dévalorisées.

Les résultats aux diverses mesures ont montré de nombreux signes de plafonnement. Nous avons déjà discuté du test sur l'étendue du vocabulaire qui surestime la connaissance des mots. Les indices de plafonnement pour la mesure des connaissances paradigmatiques nous

amènent à réfléchir sur une façon d'augmenter son pouvoir discriminant. Une partie de la solution réside peut-être dans le choix des items : dans notre sélection des noms cibles, nous n'avons pas vérifié si ces mots pouvaient être connus déjà au primaire. Il faudrait constituer un corpus du vocabulaire scolaire du primaire et sélectionner les mots selon l'ordre scolaire, en plus des domaines d'apprentissages. L'autre partie de la solution est d'ailleurs de constituer un corpus du vocabulaire du deuxième cycle du secondaire et du collégial. Nous pourrions alors développer un test de la profondeur de la connaissance du vocabulaire couvrant toute la scolarité québécoise.

D'autres résultats nous obligent à réfléchir sur la sélection des items : rappelons que l'analyse des résultats par item à la mesure des connaissances syntagmatiques a montré que les mots plus fréquents étaient moins bien réussis. La seule explication que nous pouvons proposer, c'est que les réponses pour les mots les plus fréquents offraient plus de choix plausibles, considérant la plus grande fréquence de ces mots, d'autant plus qu'une plus grande fréquence implique plus de polysémie (Hyland et Tse, 2007). Par ailleurs, l'effet mesuré est plutôt petit et ne nous amène pas à remettre en question l'effet statistique, beaucoup plus important, du degré d'abstraction.

Toujours en ce qui concerne les résultats, nous devons souligner à quel point les variances étaient grandes et l'ampleur des effets, petite. Nous avons rappelé dans le cadre théorique les nombreux facteurs intervenant dans la connaissance du vocabulaire. Ces facteurs sont autant d'ordre lexical (longueur du mot, degré d'abstraction, fréquence) que cognitif (expérience du monde, connaissances antérieures, automatité du traitement) ou socioculturel (attitudes de l'apprenant ou de ses parents vis-à-vis de la langue, habitudes de lecture). La variance inexpliquée dans les analyses s'explique par ses nombreux facteurs que nous avons mis de côté. Pour mieux comprendre la problématique des élèves allophones, nous souhaiterions prendre en considération, en plus des facteurs lexicaux, les facteurs socioculturels.

Pour terminer ce travail, nous proposons quelques pistes de recherche à la lumière de certaines de nos observations. Nous avons déjà discuté de la possibilité d'analyser les temps de réponse aux items du test sur la profondeur de la connaissance. Le temps de réponse peut

être analysé en fonction des variables lexicales (degré d'abstraction, etc.) et des variables relatives aux élèves (langue maternelle, niveau de scolarité). Une telle analyse mettra peut-être en lumière des effets des variables que les réponses des élèves n'ont pas mis en évidence. Éventuellement, nous en avons déjà discuté, nous voudrions prendre en considération les connaissances en L1 du langage scolaire ou domestique des élèves allophones pour vérifier si elles sont un facteur de la profondeur de la connaissance du vocabulaire scolaire en L2. Il faudrait finalement poursuivre la recherche sur les connaissances syntagmatiques et le vocabulaire scolaire. Plusieurs des unités syntagmatiques constituent également des éléments fréquents et importants à connaître pour comprendre un texte scolaire. Leur contenu sémantique et syntaxique pourrait apporter un éclairage sur la profondeur de la connaissance du vocabulaire des élèves.

Pour terminer, soulignons que le vocabulaire dont il a été question dans ce travail a été limité aux noms des différents domaines d'apprentissages. Nous avons besoin de mesurer également la connaissance des verbes et des adjectifs dont le vocabulaire scolaire est abondamment constitué. De même, il faudrait mesurer la connaissance des mots partagés entre les différents domaines, mais dont le sens peut varier d'un domaine à l'autre. Finalement, considérant l'importance de la littérature dans les années scolaires subséquentes (au cégep, particulièrement), une étude sur la connaissance du vocabulaire de la littérature mérite d'être prise en considération.

ANNEXE A

FEUILLE DE CONSENTEMENT DES ENSEIGNANTS

Madame Michèle Bourget,

J'accepte que mes élèves participent à l'étude sur le vocabulaire menée par Michel Bastien de l'UQÀM. Je m'engage à faire parvenir le formulaire de consentement à leurs parents et à les recueillir par la suite.

Liste des foyers concernés :

Nom du professeur : _____

Signature : _____

Date : _____

Feuille de consentement des parents

Université du Québec à Montréal

Le 23 mars 2012**Cher(s) parent(s),**

Dans le cadre de mon projet de thèse en linguistique à l'Université du Québec à Montréal, je mène une étude sur l'acquisition du vocabulaire scolaire auprès des élèves du secondaire. Les objectifs de mon projet sont de mesurer l'étendue et la profondeur de la connaissance du vocabulaire scolaire des élèves. Le but est de voir comment se développe le vocabulaire scolaire des élèves après plusieurs années à l'école.

Pour réaliser cet objectif, je dois recruter des élèves dans des écoles secondaires francophones de la grande région de Montréal. Ces élèves doivent être en première, deuxième ou troisième secondaire. La participation à l'étude se produit au cours d'une séance dans une salle d'ordinateurs de l'école. Une séance dure environ 45 minutes. Mes outils de mesure consistent en deux questionnaires suivis de deux tests de vocabulaire. Les questionnaires permettent de recueillir des informations relativement aux langues parlées et écrites des élèves, à leurs habitudes de lecture, à leurs loisirs et à leurs attitudes vis-à-vis l'école, la lecture et le vocabulaire. Le premier test de vocabulaire permet d'estimer le nombre de mots du vocabulaire scolaire connus par les élèves. Le second mesure la profondeur de la connaissance du vocabulaire, c'est-à-dire la connaissance du sens des mots et celle des autres mots souvent rencontrés dans le même contexte.

Toutes les réponses des élèves sont confidentielles. Les données seront gardées sur un serveur sécurisé et sur mon ordinateur, dont l'accès est également sécurisé. Les élèves seront identifiés par des codes et en aucun temps l'école ou les élèves ne seront nommés lors de

présentations ou de publications. Tous les groupes qui participeront recevront une analyse de leur performance. De plus, pour remercier les élèves de leur participation, je ferai tirer des certificats d'achat d'une librairie connue. Les risques encourus par les élèves qui participeront à cette étude sont comparables à ceux présents lors d'une journée scolaire typique. Il me fera plaisir de répondre à vos questions en tout temps. Vous pouvez me joindre à l'adresse suivante :

M. Michel Bastien
Étudiant au doctorat en linguistique
Département de linguistique
Université du Québec à Montréal
CP 8888, succursale Centre-Ville,
Montréal, QC, H3C 3P8
Bastien.m@uqam.ca
(514) 987-3000 #1794

Il n'y a aucune obligation de participer à ce projet. De plus, un participant peut choisir de se retirer du projet en tout temps sans pénalité ou pression. Veuillez adresser toute plainte à mon adresse ci-dessus. Si j'étais dans l'impossibilité de résoudre un problème, je vous invite à vous adresser à l'agence suivante : *Comité institutionnel d'éthique de la recherche avec des êtres humains de l'Université du Québec à Montréal* (secrétariat: 987-3000 poste 7753).

DECLARATION DE CONSENTEMENT ÉCLAIRÉ

J'ai LU et COMPRIS TOUTE l'information ainsi que les garanties données par M. Bastien ci-dessus.

J'accepte que mon/mes enfant(s) participe(nt) au projet OUI NON (Encerclez la mention pertinente)

Signature:..... DATE:.....

Nom du parent ou du tuteur..... (lettres moulées)

SI VOUS ACCEPTEZ QUE VOTRE ENFANT PARTICIPE AU PROJET, Veuillez COMPLÉTER LA SECTION SUIVANTE.

Nom de mon enfant/mes enfants :

Date de naissance de mon enfant/mes enfants :

Sexe de l'enfant :

M	F	M	F

Numéro de téléphone à la maison :

Veuillez signer cette feuille et demander à votre enfant de la retourner à la personne responsable du projet à l'école. Je suggère de conserver la première page pour vos dossiers.

ANNEXE B

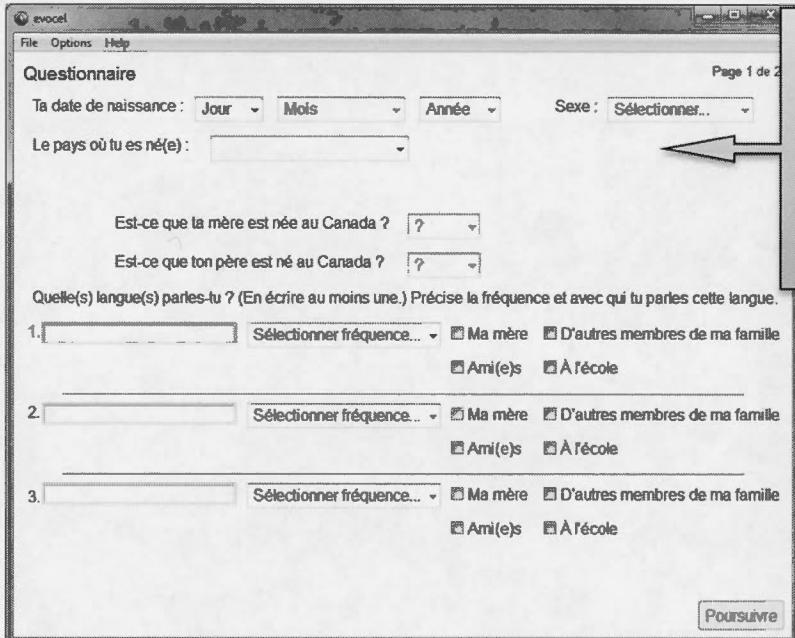
LES FENÊTRES DES DIFFÉRENTS OUTILS DE MESURE SUR LE SITE ÉVOCEL (WWW.MICROBE.CA/EVOCEL)

Questionnaire sociodémographique

« Consigne

Dans les deux prochaines pages, réponds sincèrement aux différentes questions en entrant du texte à partir du clavier ou en sélectionnant une réponse avec la souris. »

Page 1



Si la personne répond
« Canada », une nouvelle liste
apparait pour choisir la province
de naissance. Si un autre pays est
choisi, il faut ensuite répondre à
quel âge on est arrivé au Canada

Page 2

The screenshot shows a window titled "Questionnaire" with the sub-tittle "Page 2 de 2". The main text asks: "Quelle(s) langue(s) lis-tu ? (En écrire au moins une.) Précise la fréquence à laquelle tu lis dans cette langue." Below this are three input fields labeled 1., 2., and 3., each with a dropdown menu "Sélectionner fréquence...".

Below the input fields, a question asks: "Si ta langue maternelle n'est pas le français, as-tu déjà eu des cours parmi les suivants :". There are four checkboxes: Classe d'accueil en français, Francisation, Langue d'origine (PELO), and Autre langue.

Next, it asks: "Coche toutes les années où tu as étudié en français :". For "Au primaire", there are checkboxes for Préscolaire, Première, Deuxième, Troisième, Quatrième, Cinquième, and Sixième. For "Au secondaire", there are checkboxes for Première and Deuxième.

At the bottom right of the window is a "Poursuivre" button.

Questionnaire sur les habitudes de lecture et les attitudes

« Consigne

Dans les trois prochaines pages, réponds sincèrement aux différentes questions en sélectionnant une réponse avec la souris. »

Page 1

evocel

File Options Help

Loisirs et occupations Page 1 de 3

Quelles activités pratiques-tu parmi les suivantes ? Précise la fréquence.

1. Regarder la télévision (émissions ou films) en français Fréquence?

2. Jouer à des jeux électroniques, vidéos ou autres en français Fréquence?

3. Utiliser un ordinateur pour des raisons personnelles (Internet, courriel) Fréquence?

4. Activités culturelles (par exemple, théâtre, concert, musée) Fréquence?

5. Activités communautaires (sport, plein air, religion, bénévolat) Fréquence?

6. Suivre des leçons supplémentaires ou cours de soutien Fréquence?

7. Lire en français (livres, revues, journaux) Fréquence?

Quel type de document lis-tu en français? Précise la fréquence. Journaux Fréquence?

Romans, nouvelles, poésie Fréquence? Bandes dessinées Fréquence?

Manuels scolaires Fréquence? Textes sur Internet Fréquence?

Poursuivre

Page 2

evocel

File Options Help

La langue Page 2 de 3

Dis Jusqu'à quel point tu es d'accord avec les phrases suivantes.

1. J'aime lire. Ton avis?

2. Je lis seulement si je dois le faire. Ton avis?

3. J'aime recevoir des livres en cadeau. Ton avis?

4. Je suis curieux de savoir ce que mes proches lisent. Ton avis?

5. Je ne suis pas capable de lire très longtemps. Ton avis?

6. La plupart des textes que je lis à l'école sont ennuyeux. Ton avis?

7. J'aime aller à la librairie ou à la bibliothèque. Ton avis?

8. J'aime apprendre de nouveaux mots. Ton avis?

9. Quand je rencontre un mot que je ne connais pas, je cherche à savoir ce qu'il veut dire. Ton avis?

10. J'aime m'exprimer avec des mots justes. Ton avis?

11. Je n'ai pas besoin de connaître tous ces mots rencontrés lors des activités à l'école. Ton avis?

12. J'aime quand on me montre des mots que je ne connais pas. Ton avis?

Poursuivre

Page 3

La vie familiale

Page 3 de 3

Précise la fréquence de ces situations dans ta famille jusqu'à présent.

1. Nous avons voyagé à l'extérieur du Canada. Fréquence?
2. Nous avons voyagé au Canada ou au Québec. Fréquence?
3. Nous avons participé à des activités culturelles (cinéma, concert, musée). Fréquence?
4. Nous avons participé à des activités sportives ou de plein air. Fréquence?
5. Nous avons participé à des activités religieuses. Fréquence?
6. Pendant les vacances, nous sommes partis en dehors de la ville. Fréquence?
7. J'ai trouvé de l'aide pour mes devoirs. Fréquence?
8. On s'est intéressé à mes succès et à mes problèmes scolaires. Fréquence?
9. On m'a écouté et on m'a aidé quand j'ai éprouvé des problèmes. Fréquence?
10. On m'a encouragé à bien travailler dans mes projets scolaires ou autres. Fréquence?
11. On m'a félicité quand j'ai réussi quelque chose. Fréquence?
12. On s'est intéressé aux activités que j'ai entreprises. Fréquence?

Poursuivre

Mesure de l'étendue - Consigne

Premier exercice de vocabulaire

Consigne

Dans l'exercice qui vient, tu vas voir des suites de lettres suivies de boîtes à cocher à l'écran. Par exemple:

fraise Oui Non

Pour chaque suite, tu dois dire si c'est un mot que tu connais ou pas. Par exemple, est-ce que 'fraise' est un mot que tu connais? Si tu connais le sens, tu cliques sur Oui. Si tu ne le connais pas, tu cliques sur Non. Il faut donner une réponse à chacune des suites de lettres.

Il y a 80 suites de lettres en tout.

Poursuivre

Exemple de fenêtre du test

Page 1 de 5

Connais-tu le sens des suites de lettres suivantes? Clique sur Oui si tu le connais, clique sur Non si tu ne le connais pas.

base	<input checked="" type="radio"/> Oui <input type="radio"/> Non	religion	<input checked="" type="radio"/> Oui <input type="radio"/> Non
comparaison	<input checked="" type="radio"/> Oui <input type="radio"/> Non	segment	<input checked="" type="radio"/> Oui <input type="radio"/> Non
variable	<input checked="" type="radio"/> Oui <input type="radio"/> Non	point	<input checked="" type="radio"/> Oui <input type="radio"/> Non
PIB	<input checked="" type="radio"/> Oui <input type="radio"/> Non	prudiction	<input checked="" type="radio"/> Oui <input type="radio"/> Non
mase	<input checked="" type="radio"/> Oui <input type="radio"/> Non	procédé	<input checked="" type="radio"/> Oui <input type="radio"/> Non
doubhisme	<input checked="" type="radio"/> Oui <input type="radio"/> Non	sable	<input checked="" type="radio"/> Oui <input type="radio"/> Non
délanche	<input checked="" type="radio"/> Oui <input type="radio"/> Non	terre	<input checked="" type="radio"/> Oui <input type="radio"/> Non
phénomène	<input checked="" type="radio"/> Oui <input type="radio"/> Non	objet	<input checked="" type="radio"/> Oui <input type="radio"/> Non

Poursuivre

Mesure de la profondeur de la connaissance du vocabulaire - Consigne

Deuxième exercice de vocabulaire

Consigne

(Clique sur le bouton Continuer.)

Revenir

Continuer

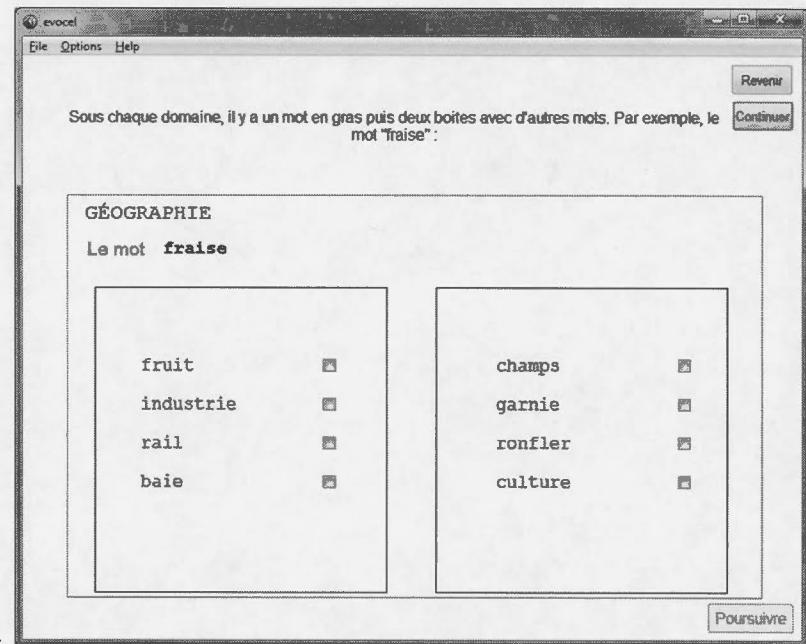
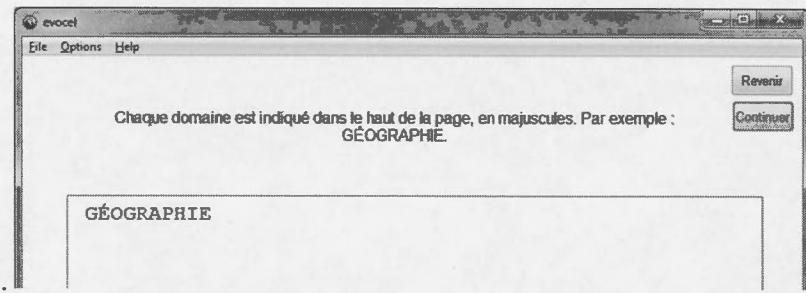
1.

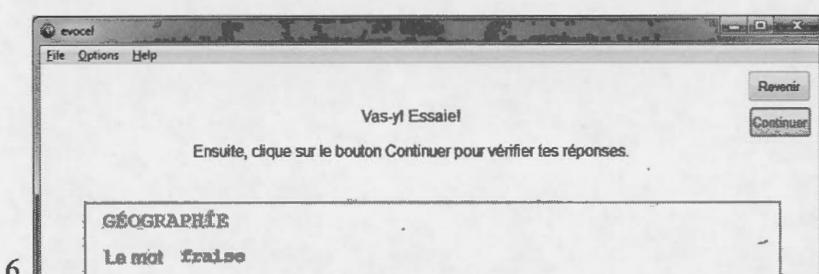
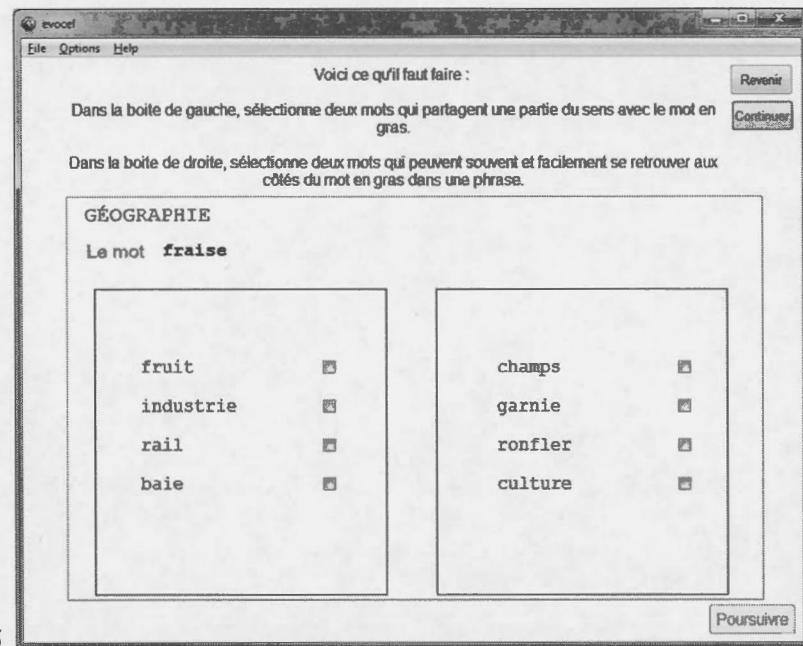
Dans l'exercice suivant, tu vas associer des mots. Ces mots proviennent des domaines d'apprentissage scolaires : français, géographie, histoire, sciences et technologie et mathématiques. En plus, il y a des mots qu'on voit souvent en français (appelés MOTS FRÉQUENTS).

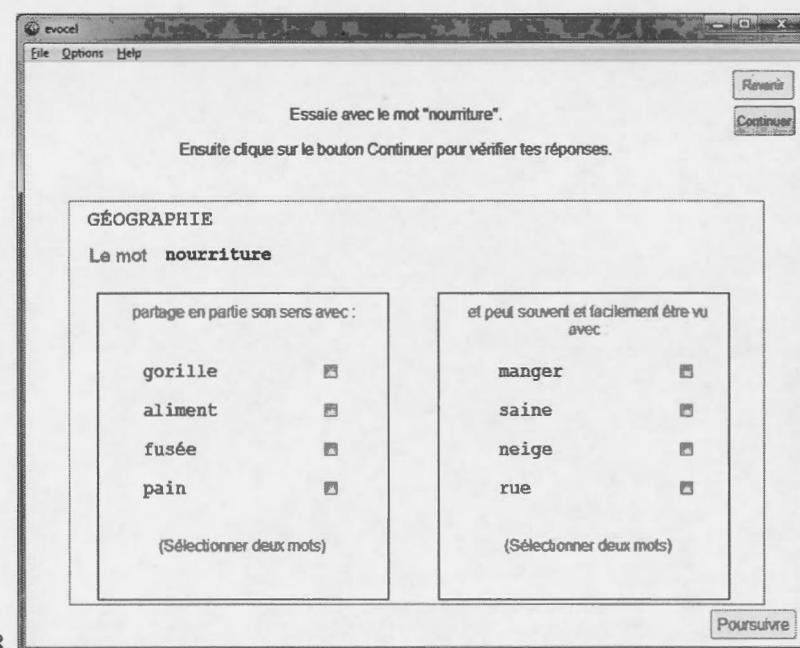
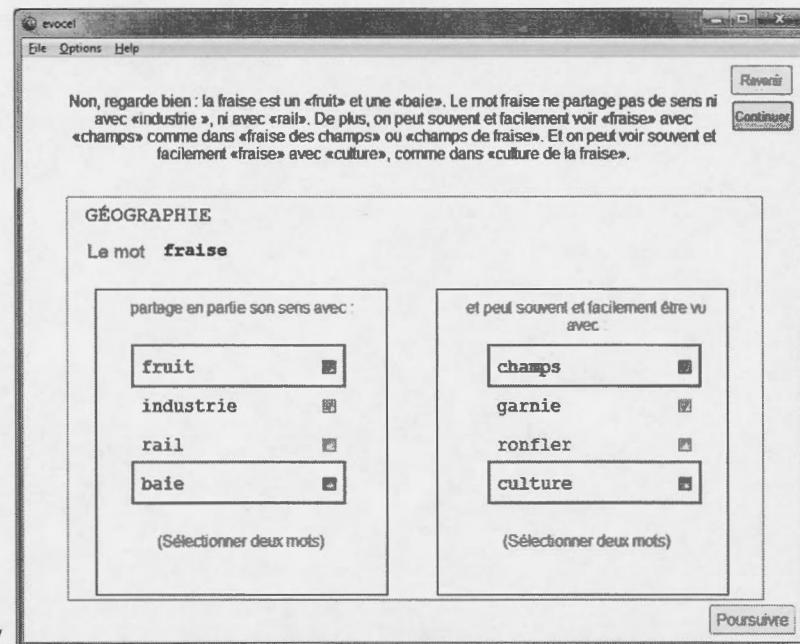
Revenir

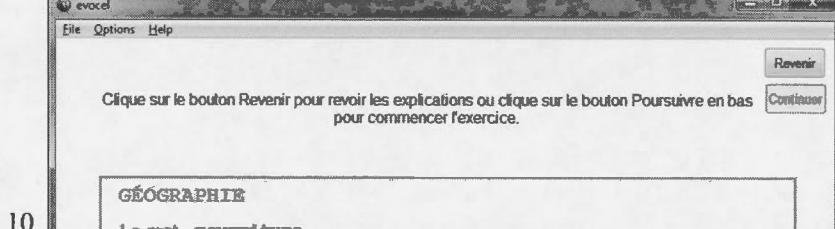
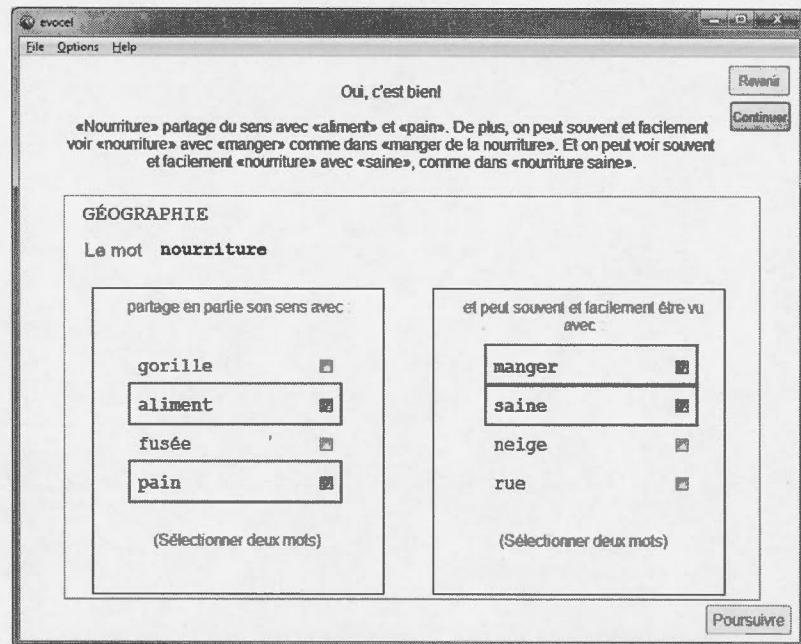
Continuer

2.

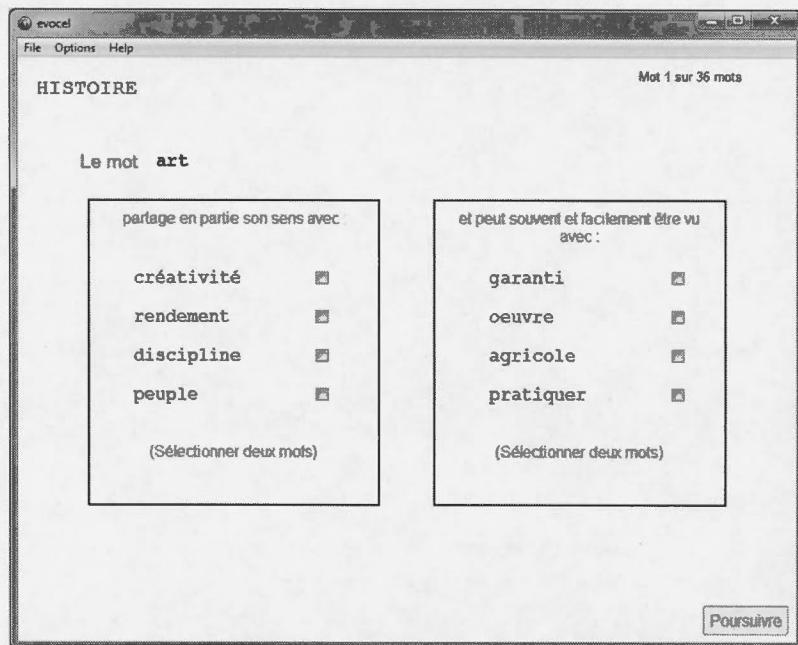








Exemple de fenêtre du test



ANNEXE C

LISTE DES MOTS ET RÉFÉRENCES DES MANUELS SCOLAIRES

Test sur l'étendue du vocabulaire - Liste des mots

Mot	Domaine	Mot	Domaine	Mot	Domaine
suspension	fr	commune	geo	art	hist
relation	fr	fossiles	geo	autorité	hist
comparaison	fr	nourriture	geo	capital	hist
déterminant	fr	phénomène	geo	époque	hist
discours	fr	PIB	geo	lumières	hist
groupe	fr	procédé	geo	nationalisme	hist
point	fr	relief	geo	production	hist
pronome	fr	sable	geo	religion	hist
relation	fr	usage	geo	société	hist
sens	fr	végétaux	geo	terre	hist
cercle	Maths	base	Techno	choses	lexorg
échantillon	Maths	cellule	Techno	coup	lexorg
losange	Maths	changement	Techno	fille	lexorg
nombre	Maths	contenant	Techno	fils	lexorg
objet	Maths	fonction	Techno	heure	lexorg
probabilité	Maths	forme	Techno	matin	lexorg
rectangle	Maths	mélange	Techno	mort	lexorg
segment	Maths	période	Techno	peur	lexorg
entier	Maths	état	Techno	question	lexorg
variable	Maths	photosynthèse	Techno	train	lexorg

Liste des pseudomots

Les pseudomots ont été créés de façon à les rendre également plausible dans l'une ou l'autre discipline : un mot dans une discipline a inspiré la création du pseudomot correspondant.

Mot	Domaine	Pseudomot	Mot	Domaine	Pseudomot
relative	fr	relaison	urbain	geo	urvain
radical	fr	radival	usage	geo	usège
comparaison	fr	comparuction	forment	geo	morfent
groupe	fr	droupe	sable	geo	sagle
production	his	prudiction	géométrique	maths	léométrique
moyen	his	moyage	variable	maths	valiable
bouddhisme	his	doubhisme	échantillon	maths	sémantillon
âge	his	ouge	possibles	maths	porribles
mélange	techno	délanche	peur	lexorg	neur
chaîne	techno	baire	mort	lexorg	cort
photosynthèse	techno	lotosynthèse	fils	lexorg	tils
base	techno	mase	veux	lexorg	teux

Mesure de la profondeur de la connaissance

Liste des mots

Mot	Domaine	Mot	Domaine	Mot	Domaine
comparaison	français	fossiles	géo	art	histoire
discours	français	phénomène	géo	autorité	histoire
point	français	PIB	géo	époque	histoire
pronome	français	procédé	géo	lumières	histoire
relation	français	sable	géo	nationalisme	histoire
sens	français	végétaux	géo	terre	histoire
cercle	maths	cellule	techno	choses	lexorg
échantillon	maths	changeement	techno	coup	lexorg
losange	maths	fonction	techno	fille	lexorg
nombre	maths	forme	techno	matin	lexorg
probabilité	maths	mélange	techno	mort	lexorg
variable	maths	photosynthèse	techno	question	lexorg

Relations paradigmatiques

Mot cible	Domaine	paradigme 1	paradigme 2	leurre 1 - paradigm	leurre 2 - paradigm
comparaison	français	jugement	rapprochement	racine	cause
discours	français	énoncé	message	ensemble	association
point	français	signe	caractère	lien	signification
pronome	français	mot	catégorie	affirmation	signe
relation	français	lien	association	catégorie	caractère
sens	français	définition	signification	agent	subordonnée
fossiles	géo	restes	empreintes	denrées	revenus
phénomène	géo	événement	fait	relation	gravier
PIB	géo	revenu	rendement	asperité	sol
procédé	géo	moyen	méthode	reste	rendement
sable	géo	sol	gravier	voyageur	rapport
végétaux	géo	plantes	flore	méthodes	empreintes
art	histoire	discipline	créativité	rendement	peuple
autorité	histoire	contrôle	pouvoir	croyance	développement
époque	histoire	ère	période	terrain	propriétaire
lumières	histoire	connaissances	savoir	investissement	fascisme
nationalisme	histoire	patriotisme	appartenance	colonialisme	expansionnisme
terre	histoire	terrain	monde	discipline	créativité
choses	lexorg	objets	trucs	craines	chocs
coup	lexorg	choc	bruit	enfant	demande
fille	lexorg	femme	gamine	décès	choc
matin	lexorg	aube	avant-midi	assuré	bruit
mort	lexorg	décès	fin	gamine	gens
question	lexorg	demande	interrogation	moment	femme
cercle	maths	figure	rond	haut	chance
échantillon	maths	portion	partie	quadrilatère	figure
losange	maths	quadrilatère	polygone	portion	unité
nombre	maths	chiffre	grandeur	polygone	solide
probabilité	maths	chance	prévisibilité	droite	rond
variable	maths	symbole	paramètre	surface	quadrilatère
cellule	techno	organisme	vivant	étape	composé
changement	techno	transformation	modification	fondement	assimilation
fonction	techno	rôle	utilité	stade	planète
forme	techno	aspect	constitution	créature	engrenage

mélange	techno	combinaison	composé	forme	comparaison
photosynthèse	techno	assimilation	transformation	série	entraînement

Relations syntagmatiques

Mot	Domaine	synt. 1	Cat.	type de relation	synt. 2	Cat.	type de relation	leurre 1	leurre 2
comparaison	fr	faire	VER	complément direct	point	NOM	complément du nom	partie	finale
discours	fr	partie	NOM	complément du nom	analyse	NOM	complément du nom	point	établir
point	fr	se terminer	VER	complément	exclamation	NOM	avec complément nominal	indéfini	direct
pronom	fr	personnel	ADJ	avec épithète	indéfini	ADJ	avec épithète	voix	phrase
relation	fr	établir	VER	complément direct	directe	ADJ	avec épithète	nominale	indéfinie
sens	fr	chercher	VER	complément du nom	figuré	ADJ	avec épithète	partie	direct
fossiles	géo	dinosauré	NOM	avec complément nominal	trouver	VER	complément direct	étalement	croissance
phénomène	géo	naturel	ADJ	avec épithète	expliquer	VER	complément direct	habitant	fabrication
PIB	géo	habitant	NOM	avec épithète	croissance	NOM	complément du nom	nouer	trouver
procédé	géo	industriel	ADJ	avec épithète	fabrisation	NOM	avec complément nominal	gouvernement	cultivé
sable	géo	désert	NOM	avec complément nominal	banc	NOM	complément du nom	manger	croissance
végétaux	géo	croissance	NOM	complément du nom	cultiver	VER	complément direct	entente	industriel
art	his	pratiquer	VER	complément direct	œuvre	NOM	complément du nom	agricole	garanti
autorité	his	exercer	VER	complément direct	soumis	ADJ	complément autre	siècle	vivre
époque	his	contexte	NOM	complément du nom	moderne	ADJ	avec épithète	culturel	électricité
lumières	his	philosophes	NOM	complément du nom	siècle	NOM	complément du nom	peupler	risque
nationalisme	his	montée	NOM	complément du nom	exacerbé	ADJ	avec épithète	gain	soumis
terre	his	peupler	VER	complément direct	cultiver	VER	complément direct	dénoncer	guerre
chooses	lexorg	plein	ADJ	complément d'adjectif	vision	NOM	complément du nom	derniers	matins
coup	lexorg	retentir	VER	sujet	donner	VER	complément direct	pas	élever
fille	lexorg	petite	ADJ	avec épithète	épouser	VER	complément direct	partir	causer
matin	lexorg	beau	ADJ	avec épithète	partir	VER	complément autre	poser	répondre

mort	lexorg	subite	ADJ	avec épithète	causer	VER	complément direct	pleine	main
question	lexorg	poser	VER	complément direct	répondre	VER	complément autre	vision	épouser
cercle	maths	rayon	ADJ	complément du nom	tracer	VER	complément direct	divisé	pyramide
échantillon	maths	analyse	NOM	complément du nom	faible	ADJ	avec épithète	rayon	perpendiculaire
losange	maths	régulier	ADJ	avec épithète	extrémités	NOM	complément du nom	augmenter	positif
nombre	maths	diminuer	VER	complément direct	diviser	VER	complément autre	bord	tracer
probabilité	maths	augmenter	VER	sujet	calculer	VER	complément direct	angle	diviser
variable	maths	valeur	NOM	complément du nom	mesurée	ADJ	avec épithète	régulier	bord
cellule	techno	reproductrice	ADJ	avec épithète	végétale	ADJ	avec épithète	ébullition	passer
changement	techno	chimique	ADJ	avec épithète	couleur	NOM	avec complément nominal	unité	verser
fonction	techno	remplir	VER	complément direct	mécanique	ADJ	avec épithète	atteindre	produit
forme	techno	prendre	VER	complément direct	changer	VER	complément autre	courroie	éléments
mélange	techno	verser	VER	complément direct	separation	NOM	complément du nom	adolescence	humains
photosynthèse	techno	processus	NOM	complément du nom	plantes	NOM	avec complément nominal	déchet	substance

Manuels scolaires consultés

Les mots cibles sont tirés des manuels scolaires suivants :

Français

Desfossés, Nicole, Marcel Lamarre et Michel Therrien. *Au fil des mots et des idées*. Mont-Royal, Québec: Thomson Groupe Modulo, 2006.

Bourbeau, Geneviève et Marie-Hélène Gosselin. *Laissez-passer. français, 1er cycle du secondaire*. Laval, Québec: Éditions Grand Duc HRW, 2005.

Fortin, Jacqueline, Carole Tremblay et Sophie Trudeau. *Rendez-vous. français, 1er cycle du secondaire*. Montréal: Graficor Chenelière éducation, 2005.

Géographie

Dalongeville, Alain, Antoine-Michel LeDoux, Romain Dalongeville, Éric Mottet, Yann Roche et Philippe Vigneault. *Cap sur les territoires: géographie, 1er cycle du secondaire*. Anjou, Québec: Les Éditions CEC, 2005.

Boudrias, Nathalie, Danielle Marcheterre et Mélanie Langlais. *Enjeux et territoires. géographie, 1er cycle du secondaire*. Montréal: Graficor Chenelière Éducation, 2005.

Laurin, Suzanne et Julie Boivin. *Territoires: géographie, 1er cycle du secondaire*. Saint-Laurent, Québec: Erpi, 2005.

Histoire

Laville, Christian, Maude Ladouceur et Geneviève Leblanc. *D'hier à demain: histoire et éducation à la citoyenneté, 1er cycle du secondaire*. Montréal: Graficor : Chenelière Éducation, 2005.

Lord, France et Jean Léger. *Histoire en action: histoire et éducation à la citoyenneté : 1er cycle du secondaire*. Mont Royal: Groupe Modulo, 2005.

Blouin, Claude et Jean Roby. *L'Occident en 12 événements: histoire et éducation à la citoyenneté, 1er cycle du secondaire*. Laval, Québec: Éditions Grand Duc HRW, 2005.

Gagnon, Hervé, Line Lamarre et Michel Vervais. *Réalités. histoire et éducation à la citoyenneté: 1er cycle du secondaire*. Saint-Laurent, Québec: Éditions du Renouveau pédagogique, 2005.

Bachand, Charles-Antoine. *Regards sur les sociétés, 1er cycle du secondaire*. Anjou: Éditions CEC, 2005.

Mathématiques

Coupal, Michel et Lynn Marotte. *À vos maths!. mathématique, 1er cycle du secondaire*. Montréal: Graficor, Chenelière éducation, 2005.

Cadieux, Richard, Isabelle Gendron, Antoine Ledoux, André Deschênes et Dominique Boivin. *Panoram@th: mathématique, 1er cycle du secondaire*. Anjou, Québec: CEC, 2005.

Guay, Sylvio, Jean-Claude Hamel et Steeve Lemay. *Perspective mathématique: 1er cycle du secondaire*. Laval, Québec: Éditions Grand Duc - HRW, 2005.

Science et technologie

Gagné, Claudette, Luc Bachand, Robert Durocher, Ghyslain Samson et Régent Bouchard. *Action*. Montréal, Québec: LIDEC, 2005.

Escrivá, Inés et Khanh-Thanh Tran. *Explorations: science et technologie : 1er cycle du secondaire*. Montréal: Graficor, 2006.

Chenouda, Atef et Mathieu Dubreuil. *Galileo. science et technologie, 1er cycle du secondaire*. Anjou, Québec: Éditions CEC, 2005.

Lalonde, J. Robert. *Univers, science et technologie. 1er cycle du secondaire*. Saint-Laurent, Québec: Éditions du Renouveau pédagogique, 2005.

ANNEXE D

LE DEGRÉ D'ABSTRACTION

Nous retrouvons dans cette annexe la description des procédures pour établir le degré d'abstraction, la longueur et la fréquence des mots utilisés dans le test sur la profondeur de la connaissance. À la fin, un tableau synthétise les informations lexicales pour chaque mot.

Degré d'abstraction

Pour cette tâche de jugement sur le degré d'abstraction, nous nous sommes inspiré des tâches résumées dans Ferrand (Ferrand, 2001) et de celle développée par Altarriba, Bauer et Benvenuto (Altarriba *et al.*, 1999). Nous avons demandé à quatre personnes ayant étudié en linguistique de juger les mots des tests sur la profondeur de la connaissance. Nous croyons que des linguistes sont les mieux placés pour distinguer le degré d'abstraction d'un mot d'autres considérations comme la fréquence, la familiarité ou la différence entre le mot lui-même et son référent.

À chacun des juges, nous avons remis un fichier Excel avec la consigne et les mots sur lesquels ils devaient porter leur jugement. Voici la consigne :

Classer chaque mot selon qu'il est plus concret (1) ou plus abstrait (6). Un mot est concret si son référent peut être perçu par un ou plusieurs sens. Un mot est abstrait si ce à quoi il réfère ne peut être défini qu'à l'aide du langage ou de connaissances. Analyser les mots en tenant compte de la matière scolaire. Par exemple, le mot 'angle' en mathématiques est plus concret que le même mot en histoire, de même que le mot 'champ' en français sera plus abstrait que le même mot en géographie. Répondre le plus spontanément possible; ne pas passer plus de dix secondes par mot.

Sous la consigne, nous présentions une liste des mots et l'échelle du degré d'abstraction. En voici un exemple pour le domaine du français :

	1	2	3	4	5	6
comparaison						
déterminant						
discours						
emphatique						
<i>Etc.</i>						

Ensuite, nous avons calculé la moyenne des degrés d'abstraction des juges pour chaque mot.

Voici le résultat de ce calcul :

Nom_cible	Domaine	J1	J2	J3	J4	Moyenne
comparaison	français	5	3	3	6	4,25
discours	français	3	3	2	5	3,25
point	français	2	1	1	4	2
pronome	français	4	2	3	5	3,5
relation	français	6	6	2	6	5
sens	français	4	6	6	6	5,5
fossiles	géo	2	1	2	1	1,5
phénomène	géo	5	5	6	5	5,25
PIB	géo	6	3	6	6	5,25
procédé	géo	6	4	6	6	5,5
sable	géo	1	1	1	1	1
végétaux	géo	2	1	2	1	1,5
art	histoire	3	2	3	5	3,25
autorité	histoire	4	3	6	6	4,75
époque	histoire	4	2	4	5	3,75
lumières	histoire	3	1	4	5	3,25
nationalisme	histoire	5	6	6	6	5,75
terre	histoire	1	1	1	4	1,75
choses	lexorg	4	6	3	2	3,75
coup	lexorg	3	1	1	3	2
fille	lexorg	1	1	2	1	1,25
matin	lexorg	3	1	2	4	2,5
mort	lexorg	3	1	5	2	2,75
question	lexorg	4	1	1	4	2,5
cercle	maths	1	1	2	1	1,25
échantillon	maths	5	2	5	5	4,25
losange	maths	1	1	2	1	1,25

nombre	maths	3	1	3	3	2,5
probabilité	maths	5	6	6	6	5,75
variable	maths	5	3	6	5	4,75
cellule	techno	3	1	6	4	3,5
changement	techno	3	3	4	6	4
fonction	techno	5	5	6	6	5,5
forme	techno	3	2	1	3	2,25
mélange	techno	3	2	1	3	2,25
photosynthèse	techno	5	3	6	5	4,75

Nous avons vérifié l'accord interjuge à l'aide de coefficients de corrélation de Spearman. Cette analyse est présentée au début de la section Analyse des résultats.

Longueur des mots

Il s'agit de la longueur des mots en termes de lettres. Nous avons utilisé la procédure automatisée sur Lexique.org.

Fréquence des mots

Nous avons utilisé l'indice de fréquence des mots dans le dictionnaire de l'application Antidote.

Synthèse des indices lexicaux pour tous les mots

Mot	Domaine	Degré abstr.	Fréquence	Longueur
comparaison	fra	4,25	56	11
discours	fra	3,25	62	8
point	fra	2	69	5
pronome	fra	3,5	41	6
relation	fra	5	65	8
sens	fra	5,5	66	4
fossiles	geo	1,5	42	8
phénomène	geo	5,25	61	9
PIB	geo	5,25	51	3
procédé	geo	5,5	55	7
sable	geo	1	54	5
végétaux	geo	1,5	51	8
art	hist	3,25	65	3
autorité	hist	4,75	63	8
époque	hist	5,25	65	6
lumières	hist	3,25	62	8
nationalisme	hist	5,75	49	12
terre	hist	1,75	65	5
cercle	maths	1,25	56	6
échantillon	maths	4,25	53	11
losange	maths	1,25	37	7
nombre	maths	2,5	68	6
probabilité	maths	5,75	50	11
variable	maths	4,75	52	8
cellule	sc_tech	3,5	58	7
changement	sc_tech	4	64	10
fonction	sc_tech	5,5	64	8
forme	sc_tech	2,25	66	5
mélange	sc_tech	2,25	55	7
photosynthèse	sc_tech	4,75	34	13
choses	mot_freq	3,75	69	6
coup	mot_freq	2	66	4
fille	mot_freq	1,25	66	5
matin	mot_freq	2,5	63	5
mort	mot_freq	2,75	65	4
question	mot_freq	2,5	72	8

Corrélations entre les variables lexicales (sans PIB)

Corrélations

		Abs	Freq	Long
Abs	Corrélation de Pearson		1	-.025
	Sig. (bilatérale)			.887
	N	35	35	35
Freq	Corrélation de Pearson	-.025	1	-.502**
	Sig. (bilatérale)	.887		.002
	N	35	35	35
Long	Corrélation de Pearson	.499**	-.502**	1
	Sig. (bilatérale)	.002	.002	
	N	35	35	35

**. La corrélation est significative au niveau 0.01 (bilatéral).

Corrélation positive entre le degré d'abstraction et la longueur des mots $P(35)=.499$, $p<0,01$.

Corrélation négative entre la fréquence et la longueur des mots $P(35)=-.502$, $p<0,01$.

Pas de corrélation entre le degré d'abstraction et la fréquence des mots.

ANNEXE E

STATISTIQUES DESCRIPTIVES POUR LE TEST SUR L'ÉTENDUE DU VOCABULAIRE

Il s'agit d'un fichier trop volumineux pour être imprimé. Il est possible de le consulter à l'adresse suivante : microbe.ca/these/ae.pdf

ANNEXE F

STATISTIQUES DESCRIPTIVES POUR LES RÉSULTATS AU TEST SUR LA PROFONDEUR DE LA CONNAISSANCE

Il s'agit d'un fichier trop volumineux pour être imprimé. Il est possible de le consulter à l'adresse suivante : microbe.ca/these/af.pdf

ANNEXE G

ANALYSES DE VARIANCES POUR LES RÉSULTATS AU TEST SUR LA PROFONDEUR DE LA CONNAISSANCE

Il s'agit d'un fichier trop volumineux pour être imprimé. Il est possible de le consulter à l'adresse suivante : microbe.ca/these/ag.pdf

ANNEXE H

MODÈLES DE RÉGRESSION LOGISTIQUE DES RÉSULTATS PAR ITEM À L'AIDE D'ÉQUATIONS D'ESTIMATION GÉNÉRALISÉES POUR LA VARIABLE SENS

Il s'agit d'un fichier trop volumineux pour être imprimé. Il est possible de le consulter à l'adresse suivante : microbe.ca/these/ah.pdf

ANNEXE I

MODÈLES DE RÉGRESSION LOGISTIQUE DES RÉSULTATS PAR ITEM À
L'AIDE D'ÉQUATIONS D'ESTIMATION GÉNÉRALISÉES POUR LA VARIABLE
CONTEXTE

Il s'agit d'un fichier trop volumineux pour être imprimé. Il est possible de le consulter à l'adresse suivante : microbe.ca/these/ai.pdf

ANNEXE J

RÉSULTATS MOYENS EN FONCTION DES FACTEURS LEXICAUX

Moyennes des résultats en fonction du degré d'abstraction pour la dimension SENS

abs	Moyenne	N	Ecart-type
1.00	.82	589	.382
1.25	.98	1767	.147
1.50	.93	1178	.250
1.75	.96	589	.190
2.00	.79	1178	.409
2.25	.76	1178	.427
2.50	.95	1767	.228
2.75	.94	589	.237
3.25	.76	1767	.426
3.50	.90	1178	.305
3.75	.98	589	.152
4.00	.97	589	.168
4.25	.77	1178	.418
4.75	.74	1767	.437
5.00	.88	589	.320
5.25	.93	1178	.257
5.50	.88	1767	.328
5.75	.75	1178	.436
Total	.86	20615	.346

Moyennes des résultats en fonction du degré d'abstraction pour la dimension CONTEXTE

abs	Moyenne	N	Ecart-type
1.00	.77	589	.423
1.25	.91	1767	.287
1.50	.88	1178	.324
1.75	.86	589	.348
2.00	.63	1178	.482
2.25	.58	1178	.494
2.50	.87	1767	.334
2.75	.90	589	.300
3.25	.65	1767	.478
3.50	.89	1178	.318
3.75	.29	589	.453
4.00	.75	589	.434
4.25	.51	1178	.500
4.75	.75	1767	.435
5.00	.61	589	.487
5.25	.50	1178	.500
5.50	.48	1767	.500
5.75	.46	1178	.498
Total	.69	20615	.464

Moyennes des résultats en fonction de la fréquence des mots pour la dimension SENS

freq	Moyenne	N	Ecart-type
34	.58	589	.494
37	.96	589	.198
41	.90	589	.305
42	.89	589	.314
49	.54	589	.499
50	.95	589	.209
51	.98	589	.152
52	.68	589	.467
53	.92	589	.271
54	.82	589	.382
55	.85	1178	.360
56	.81	1178	.394
58	.90	589	.305
61	.88	589	.322
62	.83	1178	.375
63	.97	1178	.165
64	.94	1178	.240
65	.88	2945	.328
66	.87	2356	.339
68	.88	589	.322
69	.81	1178	.392
72	.98	589	.152
Total	.86	20615	.346

Moyennes des résultats en fonction de la fréquence des mots pour la dimension CONTEXTE

freq	Moyenne	N	Ecart-type
34	.81	589	.389
37	.88	589	.330
41	.82	589	.384
42	.87	589	.337
49	.17	589	.380
50	.74	589	.441
51	.89	589	.309
52	.58	589	.494
53	.79	589	.410
54	.77	589	.423
55	.68	1178	.468
56	.56	1178	.497
58	.95	589	.213
61	.83	589	.380
62	.51	1178	.500
63	.80	1178	.399
64	.55	1178	.498
65	.70	2945	.460
66	.61	2356	.487
68	.91	589	.286
69	.48	1178	.500
72	.95	589	.223
Total	.69	20615	.464

Moyennes des résultats en fonction de la longueur des mots pour la dimension SENS

long	Moyenne	N	Ecart-type
3	.62	589	.485
4	.94	1767	.235
5	.83	3534	.374
6	.94	2945	.231
7	.89	2356	.316
8	.88	5301	.322
9	.88	589	.322
10	.97	589	.168
11	.83	1767	.372
12	.54	589	.499
13	.58	589	.494
Total	.86	20615	.346

Moyennes des résultats en fonction de la longueur des mots pour la dimension CONTEXTE

long	Moyenne	N	Ecart-type
3	.93	589	.263
4	.63	1767	.483
5	.75	3534	.432
6	.62	2945	.486
7	.80	2356	.403
8	.68	5301	.467
9	.83	589	.380
10	.75	589	.434
11	.58	1767	.493
12	.17	589	.380
13	.81	589	.389
Total	.69	20615	.464

RÉFÉRENCES

Aarts, Rian; Verhoeven, Ludo. (1999). Literacy attainment in a second language submersion context. *Applied Psycholinguistics*. *Applied Psycholinguistics*, 20, 377-393.

Aitchison, J. (2005). *Words in the mind : an introduction to the mental lexicon*. Malden, MA: Blackwell.

Alderson, C. J. (2005). *Diagnosing Foreign Language Proficiency*. London: Continuum.

Altarriba, J. et Basnight-Brown, D. (2009). An Overview of Semantic Processing in Bilinguals: Methods and Findings. In A. Pavlenko (Ed.), *The bilingual mental lexicon: interdisciplinary approaches*. Bristol UK ; Buffalo NY: Multilingual Matters.

Altarriba, J., Bauer, L. M. et Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments & Computers*, 31(4), 578-602.
doi:10.3758/BF03200738

Armand, F. (2005). Les élèves immigrants nouvellement arrivés et l'école québécoise. *Santé, société et solidarité*, 1, 141-152.

Armand, F., Beck, I. A. et Murphy, T. (2009). *Vie pédagogique*, (152).

Armand, F. et d' Anglejan, A. (1996a). Acquisition de la littératie par des élèves sous-scolarisés intégrés dans une classe d'accueil multi-âges au primaire. *L'éducation multiculturelle, ACPLS, rapport no 3*, 73-87.

Armand, F. et d' Anglejan, A. (1996b). *Éducation multiculturelle, état de la question* (No. 3). Toronto et Laval: Association canadienne des professeurs de langue seconde.

August, D. et Shanahan, T. (Ed.). (2006). *Developing Literacy in Second-Language Learners*. Mahwah: Lawrence Erlbaum Associates.

Baumann, J. F. et Graves, M. F. (2010). What Is Academic Vocabulary? *Journal of Adolescent et Adult Literacy*, 54(1), 4–12. doi:10.1598/JAAL.54.1.1

Bernhardt, Elizabeth B., Rivera, R. J. et Kamil, M. L. (2004). The Practicality and Efficiency of Web-Based Placement Testing for College-Level Language Programs. *Foreign Language Annals*, 37(3).

Bialystok, E., Luk, G., Peets, K. F. et Yang, S. (2009). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13(04), 525-531. doi:10.1017/S1366728909990423

Boivin, M.-C. et Pinsonneault, R. (2011). *Projet FQRSC Grammaire et écriture*. FQRSC.

Bressoux, P. (2010). *Modélisation statistique appliquée aux sciences sociales*. Bruxelles: De Boeck.

Carrière, I. et Bouyer, J. (2002). Choosing marginal or random-effects models for longitudinal binary responses: application to self reported disability among older persons. *BMC Medical Research Methology*, 2(15), 1-10.

Chall, J. S., Jacobs, V. A. et Baldwin, L. E. (1990). *The Reading Crisis*. Cambridge: Harvard University Press.

Cobb, T. (s. d.). Compleat Lexical Tutor. Consulté à l'adresse lextutor.ca

Cobb, T. et Horst, M. (2001). Reading Academic English: Carrying Learners Across the Lexical Threshold. In J. Flowerdew et M. Peacock (Éd.), *Research perspectives on English for academic purposes* (p. 315-329). Cambridge: Cambridge University Press.

Cobb, T. et Horst, M. (2004). Is there room for an academic word list in French? In P. Bogaards et B. Laufer (Éd.), *Vocabulary in a Second Language* (p. 15-38). Amsterdam: John Benjamins Publishing Company.

Collier, V. P. (1989). How long? A Synthesis of Research on Academic Achievement in a Second Language. *TESOL Quarterly*, 23(3), 509-531.

Corson, D. (1997). The Learning and Use of Academic English Words. *Language Learning*, 47(4), 671-718.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
doi:10.2307/3587951

Crutch, S. J., Connell, S. et Warrington, E. K. (2009). The different representational frameworks underpinning abstract and concrete knowledge: Evidence from odd-one-out judgements. *Quarterly Journal of Experimental Psychology*, 62(7), 1377-1390.
doi:10.1080/17470210802483834

Cummins, J. (2000). *Language, power, and pedagogy: bilingual children in the crossfire.* Clevedon [England] ; Buffalo [N.Y.]: Multilingual Matters.

De Groot, A. M. B. et Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, 50(1), 1-56.
doi:10.1111/0023-8333.00110

Desrochers, A. M. et Berger, M.-J. (2011). Littératie, éducation et société. In A. M. Desrochers et M.-J. Berger (Éd.), *L'évaluation de la littératie*. Ottawa [Ont.]: Presses de l'Université d'Ottawa. Consulté à l'adresse <http://public.eblib.com/EBLPublic/PublicView.do?ptiID=653326>

Dijkstra, T. et van Heuven, W. J. B. (2002). The Architecture of the Bilingual Word Recognition System: From Identification to Decision. *Bilingualism: Language and Cognition*, 5(03), 175-197. doi:10.1017/S1366728902003012

Downie, M., Gaudreau, P., Koestner, R. et Magneau, G. A. (2008). *L'influence d'une motivation culturelle sur la satisfaction scolaire et la persistance des minorités ethniques* (Rapport commercial No. 118481) (p. 5). Ministère de l'Éducation, du Loisir et du Sport (MELS).

Droop, M. et Verhoeven, L. (2003). Language proficiency and reading ability in first- and second language learners. *Reading Research Quarterly*, 38(1), 78-103.

Fazio, L. et Lister, R. (1998). Immersion and Submersion Classrooms: A Comparaison of Instructional Practices in Language Arts. *Journal of multilingual and multicultural development*, 19(4), 303-317.

Ferrand, L. (2001). *Cognition et lecture*. Bruxelles: DeBoeck Université.

Field, A. P. (2009). *Discovering statistics using SPSS*. Los Angeles, [Calif.]; London: SAGE.

Giasson, J. (2011). *La lecture: apprentissage et difficultés*. Montréal: G. Morin.

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A. et Trueswell, J. C. (2005). Hard Words. *Language Learning et Development*, 1(1), 23-64. doi:10.1207/s15473341lld0101_4

Hacquebord, H. (1994). L2-reading in the content areas: text comprehension in secondary education in the Netherlands. *Journal of Research in Reading*, 17(2), 83-98.

Hamers, J. F. (2004). A sociocognitive model of bilingual development. *Journal of Language and Social Psychology*, 23(1), 70-98.

Hanley, J. A., Negassa, A., Edwardes, M. D. et Forrester, J. E. (2003). Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. *American Journal of Epidemiology*, 157(4), 364-375. doi:10.1093/aje/kwf215

Hellman, A. B. (2011). Vocabulary size and depth of word knowledge in adult-onset second language acquisition. *International Journal of Applied Linguistics*, 21(2), 162–182. doi:10.1111/j.1473-4192.2010.00265.x

Holcomb, P. J., Kounios, J., Anderson, J. E. et Caroline, W. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3), 721-742. doi:10.1037/0278-7393.25.3.721

Horst, M. et Cobb, T. (2006). Editorial: Second Language Vocabulary Acquisition / Éditorial: Acquisition du vocabulaire d'une langue seconde. *The Canadian Modern Language review/La revue canadienne des langues vivantes*, 63(1), 1-12.

Hyland, K. et Tse, P. (2007). Is There an « Academic Vocabulary »? *TESOL Quarterly*, 41(2), 235-253.

ICC. (2008). *Présence en 2008 des immigrants admis au Québec de 1997 à 2006* (p. 32). Québec: Immigration et communautés culturelles, Québec.

ICC. (s. d.). Immigration et Communautés culturelles. Consulté à l'adresse www.micc.gouv.qc.ca/fr/index.html

Jean, M. et Geva, E. (2008). The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics*, 30, 153. doi:10.1017/S0142716408090073

Jiang, N. (2000). Lexical Representation and Development in a Second Language. *Applied Linguistics*, 21(1), 47-77.

Keuleers, E. et Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627-633. doi:10.3758/BRM.42.3.627

Kroll, J. F. et Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149-174.

Kroll, J. F. et Sunderman, G. (2003). Cognitive Processes in Second Language Learners and Bilinguals: The Development of Lexical and Conceptual Representations. In C. Doughty (Éd.), *The handbook of second language acquisition*. Malden MA: Blackwell Pub.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York; London: Routledge.

Le Bras, Y. et Lonsdale, D. (2009). *A Frequency Dictionary of French: Core Vocabulary for Learners*. Routledge.

Lefrançois, P. et Armand, F. (2003). The role of phonological and syntactic awareness in second-language reading: the case of Spanish-speaking learners of French. *Reading and Writing: An Interdisciplinary Journal*, 16, 219-246.

Lété, B. (2004). MANULEX: Le lexique des manuels scolaires de lecture. Implications pour l'estimation du vocabulaire des enfants de 6 à 11 ans. *Didactique du lexique: Contextes, démarches, supports*, 241-257.

Lété, B., Sprenger-Charolles, L. et Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments & Computers*, 36, 156-166.

Leung, C. et Franson, C. (1989). The multilingual classroom: the case for minority language pupils. *Journal of Multilingual and Multicultural Development*, 10(6), 461-472.

Meara, P. (1992). EFL Vocabulary Tests, second edition (2010). Consulté à l'adresse <http://www.lognistics.co.uk/vlibrary/meara1992z.pdf>

Meara, P. et Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language testing*, 4, 142-151.

Meara, P. M. (2009). *Connected words: word associations and second language vocabulary acquisition*. Amsterdam ; Philadelphia: John Benjamins Pub. Co.

MELS. (1998). *Politique d'intégration scolaire et d'éducation interculturelle : une école d'avenir*. Québec: MELS.

MELS. (2006). *Portrait scolaire des élèves issus de l'immigration: de 1994-1995 à 2003-2004*. Québec: MELS.

MELS. (2010). *La performance des jeunes en lecture, en mathématiques et en sciences résultats obtenus par les élèves québécois de 15 ans*. Québec: Éducation, loisir et sport Québec. Consulté à l'adresse <http://collections.banq.qc.ca/ark:/52327/2007421>

Messier, M. (1997). *Les modèles de service réservés aux élèves nouveaux arrivants: une étude comparée entre Montréal et Toronto*. Université de Montréal, Montréal, mémoire de maitrise.

Metin, S. K. et Karaoğlan, B. (2011). Measuring Collocation Tendency of Words. *Journal of Quantitative Linguistics*, 18(2), 174-187. doi:10.1080/09296174.2011.556005

Meuter, R. (2009). Neurolinguistic Contributions to Understanding the Bilingual Mental Lexicon. In A. Pavlenko (Éd.), *The bilingual mental lexicon : interdisciplinary approaches* (p. 1-25). Bristol UK ; Buffalo NY: Multilingual Matters.

Morris, L. et Labelle, M. (2008). *Une juste mesure : Développement d'instruments et de critères d'évaluation linguistique pour des élèves allophones*. (Projet financé dans le cadre du programme des Actions concertées du FQRSC en partenariat avec le MELS). Montréal: UQAM. Consulté à l'adresse http://www.frsqc.gouv.qc.ca/upload/editeur/RF-Lori_Morris%281%29.pdf

Nagy, W. E. et Anderson, R. C. (1984). How many words are there in printed school English. *Reading Research Quarterly*, 19(3), 304-330.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. et Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-12.

New, B. et Pallier, C. (2005). Manuel de Lexique 3, version 3.1. Consulté à l'adresse <http://www.lexique.org/docLexique.php>

New, B. et Pallier, C. (s. d.). Lexique. Consulté à l'adresse <http://www.lexique.org/>

OCDE. (s. d.). Organisation de coopération et de développement économiques. Consulté à l'adresse http://www.oecd.org/home/0,3305,fr_2649_201185_1_1_1_1_1,00.html

Ordóñez, C. L., Carlo, M. S., Snow, C. E. et McLaughlin, B. (2002). Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology*, 94(4), 719-728. doi:10.1037/0022-0663.94.4.719

Painchaud, G., d' Anglejan, A., Armand, F. et Jezak, M. (1994). Diversité culturelle et littératie. *Repères*, 15, 77-94.

Pavlenko, A. (2009). Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon: interdisciplinary approaches* (p. 125-160). Bristol UK ; Buffalo NY: Multilingual Matters.

Perea, M., Rosa, E. et Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception et Psychophysics*, 67(2), 301-314.

Qian, D. D. (1999). Assessing the Roles of Depth and Breadth of Vocabulary Knowledge in Reading Comprehension. *Canadian Modern Language Review*, 56(2), 282-307.

Qian, D. D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perpective. *Language Learning*, 52(3), 513-536.

Qian, D. D. et Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21(1), 28-52.
doi:10.1191/0265532204lt273oa

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards et B. Laufer (Éd.), *Vocabulary in a Second Language* (John Benjamins Publishing Company., p. 209-227). Amsterdam.

Schmitt, N., Jiang, X. et Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 95(1), 26-43.
doi:10.1111/j.1540-4781.2011.01146.x

Schoonen, R. et Verhellen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing*, 25(2), 211-236.
doi:10.1177/0265532207086782

Simard, D. (2008). *Difficultés en lecture vécues par des enfants lusophones de l'île de Montréal : rôle explicatif des habiletés syntaxiques et métasyntaxiques en L1 et en L2*. CRSH.

Singleton, D. (1999). *Exploring the Second Language Mental Lexicon*. Cambridge: Cambridge University Press.

Thornbury, S. (2002). *How to teach vocabulary*. Harlow: Longman.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. *Heritage languages in America: Preserving a national resource*, 37–77.

Verhellen, M. et Shoonen, R. (1998). Lexical Knowledge in L1 and L2 of Third and Fifth Graders. *Applied Linguistics*, 19(4), 452 -470. doi:10.1093/applin/19.4.452

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(02), 217-234.

Ward, J. (2007). Collocation and technicality in EAP engineering. *Journal of English for Academic Purposes*, 6(1), 18-35. doi:10.1016/j.jeap.2006.10.001

Wesche, M. et Paribakht, T. S. (1996). Assessing Second Language Vocabulary Knowledge: Depth versus Breadth. *The Canadian modern language review*, 53(1), 13-40.

West, M. (1953). *A General Service List of English Words* (Longman, Green etCo.). London.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Zahar, R., Cobb, T. et Spada, N. (2001). Acquiring Vocabulary through Reading: Effects of Frequency and Contextual Richness. *Canadian Modern Language Review/ La Revue canadienne des langues vivantes*, 57(4), 541-572. doi:10.3138/cmlr.57.4.541

Zorn, C. J. W. (2001). Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *American Journal of Political Science*, 45(2), 470-490.
doi:10.2307/2669353

Zouali, O. (1997). Maintien de la langue d'origine et acquisition de la langue seconde chez des élèves maghrébins à Montréal. In M.-L. Lefebvre et M.-A. Hily (Éd.), *Les situations plurilingues et leurs enjeux*. Montréal: L'Harmattan.