

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Hassiba Benbouali de Chlef

Faculté des Langues Etrangères

Département de Français



THÈSE

Réalisée pour l'obtention du diplôme de

DOCTORAT

Filière : Français

Spécialité : Didactique du FLE

Par

Mohammed BENDAOU

Thème :

L'ENSEIGNEMENT-APPRENTISSAGE DU VOCABULAIRE AU COLLEGE : CAS DES CLASSES DE 1^{ère} ANNEE

Soutenu le 06/03/2023, devant le jury composé de :

Mohand Amokrane AIT DJIDA	Professeur	Université de Chlef	Président
El Djamhouria SLIMANI	Professeur	Université de Chlef	Rapporteur
Hakim MENGUELLET	Professeur	Université de Blida 2	Examineur
Hassane MISSOUM-BENZIANE	MCA	Université de Chlef	Examineur
Redhouane KERROUZI	MCA	Université de Chlef	Examineur
Amir MEHDI	MCA	Université de Tيارت	Examineur

DÉDICACES

A la mémoire de mes parents

Mohammed Bendacud

REMERCIEMENTS

J'aimerais tout d'abord remercier ma directrice de thèse, le Pr. El Djamhouria Slimani Aït Saâda pour avoir accepté l'encadrement de ce travail et apporté ses éclairages judicieux.

Je tiens également à exprimer ma gratitude envers le Pr. Mohand Amokrane AIT DJIDA pour sa disponibilité, son amabilité et ses conseils pertinents qui m'ont aidé à réaliser ce travail.

Je tiens également à formuler mes remerciements en signe de gratitude et de reconnaissance aux membres de Jury qui ont pris la peine de lire et relire ce travail, de le corriger et de m'honorer de leur présence et enfin d'évaluer cette recherche.

Enfin, je ne peux passer outre ma reconnaissance envers ma petite famille pour sa présence, sa confiance en moi et son enthousiasme.

RESUME

Il n'y a presque aucune information dans la littérature actuelle sur le lexique FLE appris dans les écoles algériennes aujourd'hui. Dans cette recherche qui s'inscrit dans le domaine de la didactique du lexique, nous avons examiné deux des principaux aspects de l'enseignement/apprentissage du vocabulaire : la taille du vocabulaire réceptif chez les apprenants et la structure lexicale des textes supports du manuel de la 1ère AM. Beaucoup de chercheurs estiment que l'étendue du vocabulaire est l'une des principales dimensions de la compétence lexicale. Or, pour beaucoup de collégiens algériens, le problème réside dans l'absence du vocabulaire de base indispensable à tout développement de la compétence lexicale. A ce problème de carence lexicale s'ajoute l'inadéquation des textes supports du manuel de 1ère AM étant donné que pour la grande partie d'élèves algériens, le manuel scolaire est l'unique source d'apport de vocabulaire. Notre étude comportera deux volets. Dans le premier volet, nous examinerons la taille du vocabulaire basique chez une population de 61 élèves de 1ère AM évoluant dans deux milieux distincts et ce à l'aide de X_Lex - un test de vocabulaire oui/non dans lequel les élèves sont invités à cocher les mots qu'ils connaissent sur 40 mots et 20 pseudo-mots. Le second volet sera consacré à une analyse lexicographique des textes supports contenus dans le manuel de 1ère AM. Les résultats de l'expérimentation ont révélé un grand déficit structurel dans les profils lexicaux chez la population testée. L'analyse lexicographique effectuée grâce à des outils informatiques a montré que la structure lexicale des textes supports ne favoriserait pas l'acquisition du vocabulaire. Les résultats de notre travail préconisent d'octroyer plus d'intérêt à l'évaluation de la taille du vocabulaire chez les apprenants ainsi qu'à la charge lexicale des textes supports du manuel.

Mots-clés : taille de vocabulaire – carence lexicale – texte support – manuel scolaire - charge lexicale

ملخص

لا توجد معلومات تقريباً في الأدبيات الحالية حول مفردات FLE الذي يتم تعلمها في المدارس الجزائرية اليوم. في هذه الدراسة ، التي تعد جزءاً من مجال تعليمية تعلم المفردات ، قمنا بفحص جانبيين من الجوانب الرئيسية لتدريس / تعلم المفردات: حجم المفردات المستقبلية للمتعلمين والبنية المعجمية لنصوص الداعمة في الكتاب المدرس للسنة الأولى متوسط. يعتقد العديد من الباحثين أن نطاق المفردات هو أحد الأبعاد الرئيسية للكفاءة المعجمية. ومع ذلك ، بالنسبة للعديد من طلاب المدارس الإعدادية الجزائريين ، تكمن المشكلة في غياب المفردات الأساسية الضرورية لأي تطوير للكفاءة المعجمية. يضاف إلى مشكلة النقص المعجمي عدم ملاءمة النصوص الداعمة في الكتاب المدرسي للسنة الأولى متوسط بالنظر إلى أن الكتاب المدرسي بالنسبة لغالبية التلاميذ الجزائريين هو المصدر الوحيد لمدخلات المفردات. دراستنا سوف تتكون من جزئين. في الجزء الأول ، سنفحص حجم المفردات الأساسية في مجموعة مكونة من 61 طالباً من السنة أولى متوسط يتطورون في بيئتين مختلفتين باستخدام X_Lex - اختبار مفردات نعم / لا حيث يتم دعوة الطلاب لتحديد الكلمات التي يعرفونها. 40 كلمة و 20 كلمة زائفة. سيخصص الجزء الثاني لتحليل معجمي لنصوص الدعم الواردة في الكتاب المدرسي للسنة الأولى متوسط . كشفت نتائج التجربة عن عجز بنيوي كبير في الملامح المعجمية في العينة الذي تم اختبارها. أظهر التحليل المعجمي الذي تم إجراؤه باستخدام أدوات الكمبيوتر أن البنية المعجمية للنصوص الداعمة لا تشجع على اكتساب المفردات. توصي نتائج عملنا بمنح المزيد من الاهتمام لتقييم حجم مفردات المتعلمين بالإضافة إلى العبء المعجمي للنصوص الداعمة للكتاب المدرسي.

الكلمات المفتاحية: حجم المفردات - النقص المعجمي - نص السند- الكتاب المدرسي - الحمل المعجمي

TABLE DES MATIERES

DEDICACES	1
REMERCIEMENTS	3
RESUME.....	4
LISTE DES FIGURES	9
LISTE DES TABLEAUX.....	12
INTRODUCTION GENERALE	13

CADRE THEORIQUE ET CONCEPTUEL

CHAPITRE 1 :	
CONCEPTS FONDAMENTAUX	22
1.1 Introduction	22
1.2 Lexique et/ou vocabulaire	23
1.3 Qu'est-ce qu'un mot ?	24
1.4 Qu'est-ce que connaître un mot ?	26
1.5 Qu'est-ce qu'un bon test de vocabulaire ?	30
1.6 Conclusion	34
CHAPITRE 2 :	
DIFFICULTE DES MOTS, FREQUENCE ET ACQUISITION DES MOTS : PROFILS LEXICAUX.....	35
2.1 Introduction	35
2.2 Différences dans la fréquence des mots	36
2.3 Le modèle de fréquence : dessiner un profil de fréquence.....	38
2.4 Test du modèle et dessin des profils de fréquence réelle	39
2.5 Le développement des profils au cours de l'apprentissage	42
2.6 Profils de vocabulaire chez les apprenants individuels.....	46
2.7 Difficulté des mots et comment cela influence le modèle de fréquence.....	47
2.8 Conclusion	54

CHAPITRE 3 :	
RELATION ENTRE FREQUENCE ET COUVERTURE	57
3.1 Introduction	57
3.2 La fréquence des mots et la couverture : la loi de Zipf	57
3.3 La couverture requise pour la compréhension	64
3.4 Couverture dans les corpus écrits et parlés	68
3.5 La conclusion à en tirer n'est pas seulement que la compréhension.....	70
3.6 Couverture dans les lexiques spécialisés.....	70
3.7 Couverture dans différentes langues	74
3.8 Conclusion	78
CHAPITRE 4 :	
MESURE DE L'ÉTENDUE DU VOCABULAIRE PASSIF	80
4.1 Introduction	80
4.2 Tests d'étendue du vocabulaire	80
4.3 Prédications de croissance de l'étendue du vocabulaire	84
4.4 Attentes en matière de connaissances du vocabulaire dans les examens et autres programmes d'études.....	86
4.5 Études sur la croissance de l'étendue du vocabulaire	88
4.6 Temps passé en classe et taux d'apprentissage	95
4.7 Conclusion	100
CHAPITRE 5 :	
TAILLE DU VOCABULAIRE ET COMPETENCES LINGUISTIQUES	102
5.1 Introduction	102
5.2 Les quatre compétences et la taille du vocabulaire	103
5.3 Taille du vocabulaire et performance à l'examen	106
5.4 Taille du vocabulaire et le CECR.....	109
5.5 Conclusion	112
CHAPITRE 6 :	
EXPOSITION DU VOCABULAIRE DANS LES MANUELS SCOLAIRES	113
6.1 Introduction	113

6.2	Sélection du vocabulaire dans les manuels	114
6.3	Apprentissage du vocabulaire à partir de l'input en classe	117
6.4	Recyclage et acquisition du vocabulaire dans les manuels.....	117
6.5	Conclusion	119

PARTIE PRATIQUE

INTRODUCTION	121
CHAPITRE 1 :	
VERIFICATION DE LA 1 ^{ERE} HYPOTHESE (ETUDE 1)	122
1.1 Méthodologie	122
1.1.1 Population	122
1.1.2 Le profil d'entrée « présumé » d'un élève de 1 ^{ère} AM	124
1.1.3 Le profil de sortie « présumé » d'un élève de 1 ^{ère} A.M	124
1.1.4 Le cadre opérationnel.....	125
1.1.5 Questions de recherche	125
1.1.6 Choix et élaboration des instruments de mesure.....	125
1.1.7 Administration des épreuves.....	130
1.1.8 Collectes de données	131
1.2 RESULTATS	131
1.2.1 Les proportions de connaissance de la première bande de fréquence chez un élève de 1 ^{ère} AM en début d'année.....	131
1.2.2 Le niveau d'un élève de 1 ^{ère} année par rapport au niveau CECR	136
1.2.3 Les mots les plus cochés lors des deux tests	138
1.2.4 La moyenne du nombre de mots appris par un élève de 1 ^{ère} AM au cours d'une années scolaire.....	140
1.3 DISCUSSION	142
1.4 CONCLUSION PARTIELLE.....	143
1.5 LIMITES	145
CHAPITRE 2 :	
VERIFICATION DE LA 2 ^{IE} ME HYPOTHESE (ETUDE 2).....	147

2.1	Méthodologie :	147
2.1.1	Matériaux	149
2.1.2	Modification des textes	149
2.1.3	Questions de recherche	151
2.1.4	Conception de la recherche	151
2.1.5	Analyse des données	154
2.1.6	Profilage statistique.....	157
2.2	RÉSULTATS	157
2.2.1	Proportions des listes de fréquences dans les textes supports du manuel de 1 ^{ère} année.....	157
2.2.2	Le comportement de chaque liste de mots le long des huit textes supports.	159
2.2.3	Les mots les plus fréquents dans les textes supports du manuel.....	161
2.2.4	Le recyclage du vocabulaire thématique.....	163
2.3	DISCUSSION	168
2.4	CONCLUSION PARTIELLE.....	170
2.5	LIMITES DE RECHERCHE.....	173
CHAPITRE 3 :		
	IMPLICATIONS ET RETOMBEES PEDAGOGIQUES	175
3.1	Introduction.....	175
3.2	Le vocabulaire, élément incontournable dans l'apprentissage des langues	176
3.4	La fréquence du vocabulaire comme modèle d'apprentissage	178
3.5	L'apprentissage du vocabulaire et les différences individuelles.....	181
3.6	La variation individuelle et la dimension de la profondeur du vocabulaire.....	185
3.7	Variation individuelle, âge et difficulté des mots	186
3.8	Implications pour la pratique de l'enseignement du vocabulaire.....	188
	CONCLUSION GENERALE.....	191
	BIBLIOGRAPHIE	194
	ANNEXES	

LISTE DES FIGURES

CADRE THÉORIQUE ET CONCEPTUEL

- Fig. 2-1: Fréquence moyenne des mots dans les bandes de fréquences de Baudot
- Fig. 2-2 : Profil de vocabulaire d'un apprenant type
- Fig. 2-3: Profil de fréquence pour les apprenants grecs d'EFL
- Fig. 2 4 : Profil de fréquence des apprenants britanniques du français
- Fig. 2-5: Profils de fréquence pour les apprenants grecs d'EFL divisés par classe
- Fig. 2-6: Profils individuels des 10 étudiants les mieux notés à Milton
- Fig. 3-1: Les bandes les plus fréquentes du tableau 3.2 présentées sous forme de graphique
- Fig. 3-2 : Connaissance des mots impliquée par les corpus pour 80 et 95% de couverture du texte
- Fig. 3-3 : Auto-déclarations des élèves sur la compréhension et la connaissance du vocabulaire
- Fig. 3-4 : Comparaison de la couverture entre Carroll et al. et le Corpus National Hellénique
- Fig. 3-5 : Comparaison de la couverture entre le corpus anglais de Carroll et al. (1971) et le corpus français de Baudot (1992)
- Fig. 4-1 : Exemple d'un test de liste de contrôle (version française de Meara et Milton, 2003)
- Fig. 4-2 : Exemple de test de niveau tiré de Nation (2001, p. 416)
- Fig. 4-3 : Croissance lexicale des apprenants d'EFL en Grèce (Milton, 2006a)
- Fig. 4-4: Croissance lexicale des apprenants d'EFL en Hongrie (Milton, 2009, p. 80)
- Fig. 4-5: Croissance lexicale des apprenants de français langue étrangère au Royaume-Uni
- Fig. 4-6: Comparaison des mots appris par heure de classe au cours de chaque année d'études au Royaume-Uni et dans une école privée en Grèce (Milton, 2009, p. 89)
- Fig. 5-1 : Taille du vocabulaire et note de niveau « A »
- Fig. 6-1 : Vocabulaire commun à trois manuels EFL pour débutants
- Fig. 6-2: Vocabulaire commun entre Encore Tricolore et français fondamental (Tschichold, cité dans Milton, 2009)

PARTIE PRATIQUE

Fig. 1-1 : Répartition du nombre d'élèves en fonction des scores obtenus au prétest du vocabulaire (Classe A)

Fig. 1-2 Répartition du nombre d'élèves en fonction des score obtenus au prétest du vocabulaire (Classe B)

Fig. 1-3 : Profils lexicaux des deux classes

Fig. 1-4 : Comparaison entre les consommations moyennes de vocabulaire par heure

Fig. 2-1 : Séquence de collecte des données de la méthode mixte

Fig. 2-2 : L'utilisation des corpus dans l'apprentissage et l'enseignement des langues

Fig. 2-3 : Pourcentages de représentation des listes de fréquences dans les textes support du manuel de la 1ère année moyenne

Fig. 2-4 : Proportions des listes de fréquence dans les huit textes supports du manuel de la 1ère AM

Fig. 2-5 : Courbe d'Ebbinghaus

Fig. 3-1 : Profil de fréquence illustrant un déficit de niveau 2 selon Milton

Fig. 3-2 Profil de fréquence illustrant un déficit structurel Milton (2007a)

Fig. 3-3 : Scores moyens aux tests d'aptitude pour deux groupes d'apprenants avec des profils de vocabulaire différents

LISTE DES TABLEAUX

CADRE THÉORIQUE ET CONCEPTUEL

Tableau 1-1: Qu'implique la connaissance d'un mot?

Tableau 2-1 : Les mots les plus fréquents, et les mots de la bande des 5000 mots, et leurs occurrences dans le BNC

Tableau 2-2 : Comparaison des lexiques les plus fréquents des livres pour enfants (Vassiliu, cité dans Milton, 2009), des textes automobiles (Milton & Hales, 1997) et du BNC

Tableau 2-3 Absorption du vocabulaire par classe de mots

Tableau 2-4: Scores moyens de longueur de syllabe et de parenté par niveau de fréquence

Tableau 2-5: Corrélations entre le niveau de fréquence, le nombre de syllabes et la parenté

Tableau 3-1 : Rangs et fréquences des mots en anglais et en français

Tableau 3-2 : Chiffres de couverture types pour différentes bandes de fréquences

Tableau 3-3 : Couverture des textes académiques fournis par le GSL et l'AWL

Tableau 3-4 : Couverture dans business academic et littéraires

Tableau 4-1 : Objectifs de croissance du vocabulaire actif et passif dans le programme national de base hongrois

Tableau 4-2 : Consommation de vocabulaire anticipée par heure en Hongrie, dans la frontière grecque et au Royaume-Uni

Tableau 4-3: Vocabulaire appris par heure de contact

Tableau 5-1; Corrélations de Spearman entre les scores de taille de vocabulaire et les scores de lecture, d'écoute et d'écriture (n=88)

Tableau 5-2 : Corrélations entre les scores de taille de vocabulaire et les scores IELTS

Tableau 5-3 : Détails de la taille moyenne du vocabulaire aux niveaux GCSE et « A »

Tableau 5-4: Coefficients de classement de Spearman et signification statistique des corrélations entre les scores X-Lex et le GCSE et les notes AS prévues

Tableau 5-5 Scores moyens de taille de vocabulaire EFL et CECR

Tableau 5-6 Scores de taille du vocabulaire français langue étrangère et CECR

Tableau 5-7 Scores de taille du vocabulaire grec langue étrangère et CECR

Tableau 5-8 : Synthèse des scores moyens pour chaque niveau du CECR dans trois langues étrangères

Tableau 6-1 : Contenu du vocabulaire du texte pour débutant par plage de fréquences

PARTIE PRATIQUE

Tableau 1-1 : Portrait de l'ensemble des sujets

Tableau 1-2 : Répartition du nombre d'élèves en fonction des score obtenus au prétest (Classe A)

Tableau 1-3 : Répartition du nombre d'élèves en fonction des scores obtenus au prétest (Classe B)

Tableau 1-4 : Résultats au prétest du vocabulaire

Tableau 1-5 : Taille du vocabulaire du français langue étrangère et CECR chez les apprenants en Grèce (Milton & Alexiou, 2009, p. 202)

Tableau 1-6 : Taille du vocabulaire du français langue étrangère et CECR chez les apprenants en Espagne

Tableau 1-7 : Résultats au prétest du vocabulaire

Tableau 1-8 : Quelques mots les plus cochés dans les deux versions du test

Tableau 1-9 : Résultats au prétest interprétés en nombre de mots connus

Tableau 1-10 : Moyennes de mots appris lors de l'année et par heure de contact

Tableau 2-1 : Compositions lexicographiques des textes supports du manuel de la 1ère AM

Tableau 2-2. Pourcentage des fréquences dans les textes supports

Tableau 2-3 : Les 100 mots les plus fréquents dans les textes supports du manuel de la 1ère année moyenne

Tableau 2-4 : Occurrences des mots appartenant au vocabulaire thématique dans les huit textes supports du manuel

INTRODUCTION GENERALE

Le vocabulaire a été considéré comme un « aspect négligé de l'apprentissage des langues » (Meara, 1980), une « victime de discrimination » dans l'acquisition d'une langue seconde (Levenston, cité dans Laufer, 1990) puisque, par rapport à la phonologie, la syntaxe et le discours, beaucoup moins de recherches ont été menées sur l'acquisition du vocabulaire dans une langue seconde (Laufer 1986a). O'Dell (1997, p. 258) précise que le vocabulaire et le lexique sont absents des principaux ouvrages sur la théorie de l'enseignement des langues tout au long des années 1970 et 1980. Son omission peut avoir une histoire encore plus longue. Wilkins (cité dans Milton 2009), écrivant au début des années 1970, suggère qu'elle date du développement de la linguistique structurale. Pour Milton (2009), cela est également le produit d'approches structurelles devenues très omniprésentes dans l'enseignement des langues. En effet, en dehors du domaine des études de vocabulaire spécialisées, il semble exister une idée de longue date selon laquelle les mots ne sont que des mots et que l'apprentissage des mots n'est pas systématique. Par conséquent, le vocabulaire ne peut incontestablement pas être esquivé en tant que problème pédagogique et didactique. Dans les approches structurelles de l'apprentissage, la partie de l'apprentissage des langues est la manière dont les règles et systèmes linguistiques sont acquis, et avec cette approche, il n'y a pas lieu de trop se préoccuper des mots auxquels ces règles et systèmes s'appliquent. Certains chercheurs à l'instar de Milton (2009) assignent la minimisation du rôle du vocabulaire à la croyance persistante parmi les enseignants et les apprenants, qu'il est possible de devenir très compétent ou même un utilisateur averti dans une langue étrangère avec des ressources de vocabulaire très limitées. Il s'avère pourtant que des milliers de mots sont nécessaires même pour une communication de base, sans parler de la fluidité. Mais l'idée que l'enseignement des langues vivantes étrangères ne nécessite qu'une poignée de mots persiste, probablement parce que c'est aussi le produit d'un vœu pieux (Milton, 2009). Apprendre une langue n'est pas toujours facile. Pour s'exprimer comme un locuteur natif, il faut apprendre des milliers de mots, découvrir quels mots peuvent être combinés et lesquels ne le peuvent pas, et maîtriser de nombreuses règles du langage. Cela peut prendre des années d'efforts même pour atteindre les niveaux de base d'expression et de compréhension. Les

enseignants doivent essayer d'intégrer tout cela dans un emploi du temps restreint et de maintenir la motivation des apprenants à tout moment.

Harris & Snow (2004, p. 55) précisent qu'il y a aussi la croyance largement répandue selon laquelle l'enseignement explicite du vocabulaire n'est qu'une perte de temps et que peu de mots sont retenus de ceux qui sont « appris » ou « enseignés » par l'enseignement direct et que la plupart du vocabulaire de la L2 est appris de manière fortuite, en grande partie par voie orale. Il s'ensuit alors, toujours selon cette croyance, que la meilleure façon d'aborder le vocabulaire est de ne pas l'enseigner du tout car les apprenants vont s'en imprégner comme par osmose de la langue qui les entoure à l'intérieur ou à l'extérieur de la classe (Milton, 2009). Pour Milton, tout cela s'avère fantaisiste car des données probantes suggèrent que l'assimilation du vocabulaire résultant d'une exposition au langage vraiment fortuite est généralement négligeable et que les apprenants qui réussissent acquièrent de grandes quantités de vocabulaire à partir des mots explicitement enseignés en classe et complètent leur apprentissage en ciblant le vocabulaire dans des activités, comme apprendre les mots de chansons, en dehors de la classe. Pourtant, ce qui se passe en réalité semble confirmer que les idées illusoires et les contraintes du temps l'emportent sur les preuves tangibles dans la construction des programmes d'enseignement et dans la pratique de l'enseignement.

Alors que l'étude du vocabulaire est devenue récemment beaucoup plus à la mode au niveau académique, cet intérêt ne s'est pas encore transféré au courant dominant de l'enseignement des langues étrangères. Milton (2009, p. 3) déplore la volte-face du Cadre Européen Commun de Référence sur les premiers travaux consacrés aux listes de vocabulaire et déplore que les travaux du CECR se concentrent uniquement, et ce depuis 2001, sur des descriptions des compétences et de connaissances tout en omettant la compétence lexicale.

En Algérie, et à l'instar d'autres pays, l'adhésion du système éducatif aux approches communicatives a forgé l'idée que le fait de mettre l'accent sur le vocabulaire est toujours préjudiciable aux apprenants et qu'il pourrait et devrait être évité, même lorsque la communication est l'objectif principal de l'apprentissage des langues. Dans nos écoles, cet effet se traduit à la fois par la réduction des volumes de vocabulaire présentés aux apprenants et des volumes de vocabulaire appris. Même les séances dédiées au vocabulaire traitent uniquement des notions métalexicales et omettent tout objectif explicite visant à enrichir le vocabulaire thématique de l'apprenant. Aujourd'hui, il est facile de constater que l'un des composants

importants de la compétence communicative, en l'occurrence la compétence lexicale, est mise à l'écart dans les débats entretenus au sein du corps enseignant lors des journées pédagogiques.

Pourtant, Le vocabulaire est l'un des éléments constitutifs de la langue. L'importance du vocabulaire dans l'apprentissage des langues a été soulignée par beaucoup de linguistes et de méthodologistes. Nickel (1973) soutient que la communication est basée beaucoup plus sur des éléments lexicaux que grammaticaux. Wilkins (cité dans Laufer, 1990) avance que « *While without grammar very little can be conveyed; without vocabulary nothing can be conveyed* » (Wilkins, 1972, p. 111). Pour Hatch, la compétence communicative de base concerne en grande partie les stratégies que les apprenants utilisent pour solliciter le vocabulaire dont ils ont besoin afin de faire passer le sens. (Hatch, 1983, p. 74). Cuq (2004, p. 91) déclare que « *L'entrée par le vocabulaire est un penchant naturel à tout apprenant de langue étrangère, car le vocabulaire est pour l'apprenant le canal le plus direct qui le relie à son système conceptuel* ».

La théorie récente de l'apprentissage des langues suggère que la réduction des volumes de vocabulaire acquis par les apprenants peut en fait nuire au développement d'autres aspects de la langue ; par exemple, l'apprentissage des mots peut en fait conduire au développement de connaissances structurelles (Milton, 2009). Il est alors possible de remettre en cause à un niveau théorique les approches d'apprentissage de ce vocabulaire marginal ou de le réduire à des niveaux minuscules. Il est également possible, à partir de travaux récents sur la compréhension et la couverture, d'apporter une justification très pratique à l'enseignement du vocabulaire en plus grand volume.

Les mesures disponibles des ressources de vocabulaire des apprenants remettent en question le mythe selon lequel il est possible d'être un utilisateur de langue précis et très communicatif avec un très petit vocabulaire. Dans ses deux ouvrages, Galisson (1983, 1991) s'élève contre la négligence dont fait l'objet le vocabulaire et réfute l'argument selon lequel le vocabulaire s'apprend tout seul par le contexte. Il plaide pour un enseignement intentionnel du vocabulaire. En effet, le souci de comprendre le sens global des messages détournerait l'attention de l'apprenant des mots, de la correspondance entre leur morphologie et leur sens ainsi des relations sémantiques qu'entretiennent les mots entre eux.

Dans son fameux rapport adressé au ministre français de l'éducation, le linguiste Bentolila tire la sonnette d'alarme sur l'ampleur et la gravité de la déficience lexicale observée chez les élèves français. Il écrit que « si l'abondance des mots n'est pas toujours garant d'une

communication efficace, leur rareté et leur imprécision sont souvent gages d'échec ». Ce constat, pourtant si saillant, n'a pas fait l'unanimité. Néanmoins, entant que praticien, je souscris pleinement à l'appréciation du linguiste Bentolila puisque ce manque de mots répond toujours présent dans toutes les activités linguistiques en classe de langue et plonge souvent la salle de cours dans une sorte d'aphasie. Lorsque Bentolila (2007, p. 3) déplore le manque de précision des mots « *Lorsque les mots précis manquent aux élèves, c'est le sens qu'ils tentent de donner au monde qui s'obscurcit.* », nous, nous déplorons, de notre part, le manque crucial des mots chez les apprenants de l'école algérienne.

Incessamment et urgemment, le besoin crucial d'enrichissement lexical se fait sentir chez les apprenants natifs à fortiori les apprenants non-natifs. Les chercheurs préconisent en conséquent une étude provisoire des mots décontextualisés avant et après la lecture. Il apparaît donc qu'opter pour l'option « enseignement implicite du vocabulaire » n'a pas porté ses fruits et les apprenants continuent à éprouver énormément de difficultés à s'approprier du vocabulaire. Pourtant, De nombreux scientifiques reconnaissent que l'enseignement de mots ciblés est une intervention efficace pour l'apprentissage et le développement du vocabulaire des élèves ayant des troubles d'apprentissage. Beck et McKeown (2004) ajoutent que ce type d'enseignement est plus pertinent lorsque le mot cible est utile aux élèves. Plus précisément, l'étude de Laufer et Shmueli (1997) montre que l'enseignement des mots cibles peut être mieux mémorisé car l'attention de l'apprenant est uniquement focalisée sur ces mots. Contrairement à l'enseignement du vocabulaire par la lecture, les enfants peuvent être distraits par la compréhension globale. Dès lors, il est aujourd'hui constatable que cette approche n'est toujours pas la bienvenue dans nos écoles.

Notre recherche s'inscrit donc dans le domaine de la didactique du français langue étrangère et plus précisément la didactique du lexique. En effet, le vocabulaire qualifiée de « brebis galeuse » par Bogaards continue toujours à être « redouté » du fait de son système ouvert qui rend tout apprentissage quasi infini. De notre part, nous pensons que le développement du vocabulaire chez les apprenants doit faire l'objet de la plus grande attention. Cette attention doit impliquer une évaluation de la taille du vocabulaire chez les apprenants dans chaque fin de cycle scolaire ainsi que la sélection du vocabulaire tout en tenant compte de l'occurrence des mots dans le manuel scolaire. Notre recherche comporte deux volets, l'un s'intéressera à la mesure de la taille du vocabulaire, l'autre examinera l'occurrence des mos dans les textes supports du manuel de 1ère AM. Les deux volets aboutiront à des conclusions et des retombées didactiques et pédagogiques.

Notre choix s'est rabattu sur ce sujet d'autant que les études relatives à l'appropriation du vocabulaire chez les jeunes apprenants algériens ne courent pas les rues. De plus, et en tant que praticien, les effets de la carence lexicale observée chez les apprenants sur le déroulement de toutes les séances sont palpables au quotidien et les apprenants de langue y voient souvent la plus grande source de problèmes. Stephen Krashen constate que « *Lorsque les étudiants voyagent, ils ne transportent pas de livres de grammaire, ils transportent des dictionnaires* » (cité dans Lewis, 1993)

Il est donc vital pour les chercheurs de comprendre son acquisition. Heureusement, ce domaine d'investigation s'est développé depuis que Meara (1980) a décrit la recherche sur l'acquisition du vocabulaire comme un domaine oublié de la recherche sur l'acquisition d'une langue seconde.

Par ailleurs, nous avons choisi comme public les élèves de la première année moyenne étant donné que la première année constitue le socle du cycle moyen et qu'à ce niveau, la plupart des apprenants sont à mi-chemin de la durée de la scolarité obligatoire qui s'étend en Algérie jusqu'à seize ans.

Dans cette perspective, notre problématique s'articule autour de la question directrice suivante: pourquoi les apprenants de la première année éprouvent-ils des difficultés quant à l'appropriation du vocabulaire et l'enrichissement de leur stock lexical ?

En réalité, il n'y a jamais eu d'intérêt particulier à cette problématique de carence lexicale bien qu'elle « s'invite » tout le temps dans nos classes. Même en France, cela a conduit selon Bentolila à « une grave insécurité linguistique ». Bentolila (2007, pp. 8,9) ajoute que tout manque sérieux de vocabulaire risque de rendre inutile l'apprentissage indispensable des relations grapho-phonologiques. Traduire des lettres ou des groupes de lettres en sons ne permettra pas à l'apprenant d'accéder au sens des mots écrits précisément parce qu'ils n'apparaissent pas dans son vocabulaire oral. Faute de vocabulaire suffisant, la maîtrise du code le conduira alors à ne produire que du bruit et non du sens. Nous estimons qu'il est peut-être impossible d'identifier tous les facteurs responsables de ces lacunes que l'école sera incapable de combler dans la plupart des cas. Mais, nous tenterons à travers ce travail d'aborder une problématique tant présente mais souvent « boudée » par les chercheurs en Algérie. Dans la littérature française, les travaux traitant

des problèmes de l'étendue du vocabulaire¹ et de son développement se font très rares. Nous avons constaté que les travaux de Meara, Milton et Lewis accordent un intérêt particulier à l'aspect quantitatif du vocabulaire. Imprégné par la lecture de ces travaux, nous avons émis ces deux hypothèses :

- 1- L'élève de 1^{ère} AM ne disposerait pas de vocabulaire de base susceptible de constituer un moyen déterminant pour réduire l'écart entre l'apport langagier enseigné et ses pré-acquis. Cet écart entre la charge lexicale des différents supports altère sérieusement le processus d'acquisition d'un nouveau vocabulaire.
- 2- Le choix des textes supports proposés dans le manuel scolaire de 1^{ère} AM ne tiendrait pas en compte l'occurrence des mots qui représente l'un des facteurs essentiels dans la rétention des mots.

En effet, ces collégiens ne possèderaient pas déjà le vocabulaire de base qui leur favoriserait l'acquisition du vocabulaire d'autant que le vocabulaire de base recouvre le noyau dur du lexique d'une langue. Cette situation génère inéluctablement « l'effet Mattieu » qui se résume par la célèbre formule « les riches s'enrichissent et les pauvres s'appauvrissent ». Autrement dit, les élèves qui n'ont pas un vocabulaire de base continuent à être « lexicalement » pauvres et ceux qui en ont déjà assez parviendront à s'en approprier facilement un nouveau. Pour être plus claire, la pauvreté du stock lexical chez l'apprenant implique des difficultés d'appropriation du vocabulaire. Les psychologues supposent qu'en général, le vocabulaire est difficile à apprendre s'il n'a aucune relation, association ou similitude avec un vocabulaire déjà appris (Higa, cité dans O'dell, 1990). Outre cette déficience lexicale constatée, les textes supports considérés comme la source principale d'input ne disposeraient pas de nombre d'occurrence nécessaire à la rétention des mots, et en particuliers les mots relevant du vocabulaire thématique.

En quête de validation de ces deux hypothèses, nous avons été amené à embrasser une approche à la fois empirique et descriptive. L'étude empirique procèdera à l'évaluation de la taille du vocabulaire des collégiens de 1^{ère} année moyenne. Quant à l'étude descriptive, elle portera sur l'analyse de la structure lexicale des textes supports du manuel de la 1^{ère} AM.

¹ La définition de l'étendue du vocabulaire que nous utilisons dans cette étude se rapporte principalement à la taille du vocabulaire, c'est-à-dire le nombre de mots que les apprenants connaissent.

Pour servir notre recherche, nous avons mis en place deux outils que nous avons pu « dénicher » à l'issue de nos lectures sur le sujet choisi. Le premier outil consiste en le test X_Lex conçu par Meara et Milton. Ce test évalue le vocabulaire réceptif sur la base de la fréquence des mots. Nous nous contenterons de mesurer la connaissance relative aux mille mots les plus fréquents de la langue française. En effet, plus un mot est fréquent dans une langue, plus il est susceptible d'être appris facilement et tôt. Le second outil consiste en deux logiciels informatiques qui sont utilisés comme des outils linguistiques de corpus. Nous effectuerons moyennant de ces logiciels un profilage statistique des textes supports du manuel de 1ère année moyenne.

Architecture de la recherche

La présente thèse se divise en deux parties : un cadre théorique et conceptuel dans lequel nous inscrirons notre recherche, et une partie pratique associant un volet expérimental et un volet descriptif. Cette partie pratique est consacrée exclusivement à la présentation de l'expérimentation menée auprès des apprenants, et des résultats auxquels nous avons abouti ainsi qu'à une analyse de contenu.

La partie théorique se subdivise en six chapitres :

Dans le premier chapitre, nous tenterons de donner des explications pratiques sur le vocabulaire et les différentes façons de le mesurer. Le chapitre n'abordera pas toutes les options et tous les détails sur les raisons pour lesquelles ces mesures ont évolué exactement comme elles l'ont fait, mais devrait nous fournir une compréhension des termes utilisés dans cette thèse. Il couvrira la notion du « mot », celle de la connaissance des mots et la méthode utilisée pour mesurer cette connaissance.

Dans le deuxième chapitre, nous examinerons le modèle de fréquence de l'apprentissage du vocabulaire. Nous illustrerons également les profils lexicaux que possèdent les groupes et les individus et nous montrerons comment ces profils évoluent au fil du temps et à mesure que les connaissances linguistiques globales augmentent. Nous aborderons aussi l'influence de la difficulté du mot sur son apprentissage.

Le troisième chapitre est consacré à l'examen de la relation entre le vocabulaire le plus fréquent et la couverture du texte. Il introduira la loi de Zipf, qui permet de représenter graphiquement la relation entre la fréquence des mots et la couverture. Cela suggérera :

- la quantité de vocabulaire nécessaire pour lire un texte pour une compréhension de base de l'essentiel ;
- la quantité de vocabulaire nécessaire pour écouter un texte parlé normal ;
- la quantité de vocabulaire nécessaire pour une compréhension complète à la fois en lecture et en écriture ;

Le quatrième chapitre examinera comment mesurer l'étendue du vocabulaire, le nombre de mots qu'un apprenant connaît ou reconnaît, et ce qui se passe lorsque ces tests sont utilisés pour collecter des mesures parmi des apprenants à différents niveaux de connaissances et dans différents environnements d'apprentissage. Il prendra en compte les volumes et la régularité de la croissance du vocabulaire auxquels on pourrait s'attendre, les exigences des examens et des programmes, les études de croissance du vocabulaire au cours de l'apprentissage ainsi qu'au temps passé en classe et le taux d'apprentissage du vocabulaire.

Le cinquième chapitre examinera les preuves disponibles selon lesquelles la taille du vocabulaire est étroitement lié au niveau d'examen et conséquemment aux hiérarchies de performance telles que le Cadre européen commun de référence pour les langues (CECR).

Le sixième chapitre mettra en lumière l'apport du vocabulaire que reçoit l'apprenant en classe de langue. Nous nous focaliserons sur l'apport du manuel scolaire qui, bien qu'il ne soit pas nécessairement la seule source d'apport de vocabulaire pour les apprenants, il est pour de nombreux apprenants en langues étrangères la principale source des mots qu'ils apprennent.

Quant à la partie pratique, elle sera destinée à la vérification des hypothèses émises. Elle se subdivise en deux volets. Le premier volet est expérimental et sera consacré à vérifier la première hypothèse. L'expérimentation que nous mettrons en œuvre aura comme objectif de :

- évaluer les proportions de connaissances de la première bande de fréquence, autrement dit, des 1000 mots les plus fréquents chez les apprenants de 1^{ère} AM ;
- assigner aux apprenants un niveau CECR équivalent en se référant à leurs scores ;
- connaître les mots les plus connus des élèves de 1^{ère} année moyenne ;
- estimer le taux d'apprentissage en vocabulaire chez les apprenants de 1^{ère} AM.

Pour ce qui concerne le second volet qui sera consacré à la vérification de la deuxième hypothèse, nous y mettrons en exergue la structure lexicale des textes supports du manuel de 1^{ère} année, et ceci dans le but de :

- présenter les proportions des bandes de fréquence dans les textes supports du manuel de 1^{ère} AM ;
- vérifier le comportement de chaque liste de mots le long des huit textes supports du manuel ;
- connaître les mots les plus fréquents dans ces textes supports ;
- vérifier dans quelle mesure le vocabulaire thématique a-t-il été recyclé.

CADRE THÉORIQUE ET CONCEPTUEL

CHAPITRE 1 :

CONCEPTS FONDAMENTAUX

1.1 Introduction

Nous vivons dans une société où nous mesurons tout le temps : notre taille, notre poids, notre pointure, la vitesse de notre voiture. Nous le faisons automatiquement et pensons rarement aux unités que nous utilisons pour la mesure jusqu'à ce que les unités changent pour une raison quelconque.

La mesure de la langue, et des connaissances en vocabulaire en particulier, ne fait pas exception. Si nous comprenons mal les unités ou utilisons les mauvaises unités, nous n'apprendrons probablement que très peu sur la langue que nous essayons de comprendre. Le but de ce premier chapitre est d'expliquer quelles sont ces unités de mesure pour décrire l'acquisition du vocabulaire et comment nous procédons pour mesurer la connaissance du vocabulaire.

Mesurer le langage n'est pas aussi simple que mesurer la distance ou le poids. Les techniques que nous utilisons fréquemment pour évaluer les performances des apprenants semblent loin de l'objectivité d'autant que nous avons tendance à noter les performances plutôt qu'à les mesurer, ce qui peut conduire à des interprétations erronées. Par exemple, si deux essais reçoivent une note sur 10, et que l'un reçoit 8 et l'autre 4, cela ne signifie pas que le premier apprenant a deux fois plus de connaissances ou de capacités que le second, même si la note est deux fois plus grande. L'utilisation de nombres pour le classement suggère que cela devrait être le cas, mais ce n'est pas le cas. Dans ces circonstances, il est difficile de caractériser les connaissances et les progrès en langue étrangère avec exactitude ou précision ; il est difficile de mesurer le langage.

L'un des avantages de l'examen de l'apprentissage du vocabulaire dans une langue étrangère est que, superficiellement du moins, il s'agit d'une qualité qui semble dénombrable ou mesurable dans un sens significatif. Vous pouvez compter les mots dans un passage ou estimer le nombre de mots qu'un apprenant connaît, et les nombres qui émergent ont plus de sens qu'une note sur 10 pour un essai. Un passage de 400 mots est deux fois plus long qu'un passage de 200

mots. On peut dire qu'un apprenant qui connaît 2000 mots dans une langue étrangère a deux fois plus de connaissances qu'un apprenant qui ne connaît que 1000 mots. Bien que le principe semble très prometteur, en réalité, évaluer la connaissance du vocabulaire n'est pas si facile. Il n'est pas toujours clair, par exemple, ce qu'est exactement un mot, et ce qui semble être une simple tâche de compter le nombre de mots dans un texte peut aboutir à plusieurs réponses possibles. Encore une fois, en estimant le nombre de mots qu'un apprenant connaît, il est possible de proposer plusieurs définitions de la connaissance, certaines plus exigeantes que d'autres, qui pourraient produire des estimations de taille très différente. Les sections suivantes expliqueront les termes utilisés pour mesurer la connaissance et l'apprentissage du vocabulaire, et établiront certaines règles de base pour les termes utilisés dans cette thèse.

1.2 Lexique et/ou vocabulaire

Malgré certains usages quasi synonymes et arbitraires des termes vocabulaire et lexique, il convient de préciser que le lexique est « *l'ensemble des mots qu'une langue met à la disposition des locuteurs* » (PICOCHÉ, 1992, p. 45), alors que le vocabulaire représente « *l'ensemble des mots utilisés par un locuteur donné dans des circonstances données.* » (PICOCHÉ, 1992, p. 45). Néanmoins, ces deux termes ne sont pas pour autant cloisonnés puis que le lexique qui transcende le vocabulaire n'est accessible que par ce dernier. Ainsi, le lexique peut être synonyme de vocabulaire lorsqu'il désigne « *l'ensemble des unités formant le vocabulaire, la langue d'une communauté, d'une activité humaine, d'un locuteur, etc.* » (DUBOIS Jean, 1999, p. 282). Dubois définit le lexique comme étant l'ensemble des unités lexicales de la langue et le vocabulaire comme étant l'ensemble des éléments de la parole ou d'un texte (DUBOIS Jean, p. 508). En d'autres termes, le lexique est une réalité qui relève de la langue. Cette réalité n'est accessible que par la connaissance des vocabulaires particuliers qui représentent la réalité du discours.

Comme la présente recherche s'inscrit dans le cadre de la recherche en éducation, il est important de connaître les acceptions de ces deux termes (vocabulaire et lexique) dans ce champ précis de la recherche. Ainsi, le terme « lexique » est défini selon le dictionnaire actuel de pédagogie de Légendre comme suit :

« *Nombre de termes que peut posséder et utiliser un être humain à un moment particulier de son cheminement intellectuel. Répertoire qui inventorie ces termes. Ensemble des mots que comporte une langue à un moment déterminé de son histoire.* »

Quant au vocabulaire, Legendre le définit ainsi : « *ensemble des mots que connaît un individu et dont il peut disposer pour entrer en communication avec autrui ; bagage lexical d'un individu ; partie de l'enseignement et de l'apprentissage des langues traitant des relations de sens entre les mots constitués* » (Legendre, 2005, p. 1449) .

En comparant ces deux définitions appartenant précisément au champ de la didactique des langues, nous remarquons que les termes vocabulaire et lexique s'emploient comme synonymes quand ils désignent tous les deux le nombre d'unités de la langue que connaît un individu. Par contre, lorsqu'on parle de partie l'enseignement/apprentissage des langues traitant des relations sémantiques, ce sens est assigné uniquement au vocabulaire.

Le lexique inclus dans les dictionnaires est mis à jour constamment. Chaque année, des mots apparaissent, d'autres disparaissent d'où l'impossibilité d'établir un inventaire précis. Ce perpétuel mouvement a été démontré notamment après le dépouillement du dictionnaire *Le petit Larousse* entre 1949 et 1960, ce qui a illustré la vitesse à laquelle se renouvelle le lexique d'une langue. De ce fait, il n'est pas possible d'arrêter un état de synchronie autant pour le lexique que pour le vocabulaire d'un individu particulier. Car, un individu n'utilise pas tous les mots qu'il connaît d'autant qu'il y a un vocabulaire actif et un vocabulaire passif.

1.3 Qu'est-ce qu'un mot ?

Etant donné que le lexique désigne conventionnellement l'ensemble des mots d'une langue, il y a lieu de préciser cette notion du « mot », de l' « unité lexicale » en tant qu'élément de base de l'ensemble. Néanmoins, le mot n'a jamais été facile à définir. En effet, toutes les tentatives des linguistes pour trouver une définition précise du mot étaient vaines d'autant plus que les définitions proposées ne s'appliquaient pas à toutes les langues. Ces définitions sont soit trop générales, soit trop particulières puis qu'elles mettent parfois en relief quelques côtés du mot et en omettent d'autres.

En linguistique française, la définition du mot présente un certain problème à discuter. A l'heure actuelle, les points de vue divergent en deux points fondamentaux :

1. Le premier affirme que la notion du mot ne concerne que la forme phonique et morphologique tandis que la signification est du domaine extérieur.
2. Le second, se basant sur la théorie de la signification de Ferdinand de Saussure, considère la signification comme un élément essentiel et indissociable à la notion du mot.

Dans les dernières années, l'attention des spécialistes s'est considérablement accrue sur la dimension de la langue dont l'objet est le vocabulaire. Cet intérêt est généré par le déplacement des idées qui s'est opéré du structuralisme vers le fonctionnalisme. La définition de l'unité d'apprentissage inclut désormais tous les éléments du discours en mettant davantage l'accent sur le sens des messages plutôt que sur la forme des énoncés.

Actuellement, la didactique tente de réconcilier une approche pragmatique qui vise à engager l'apprenant dans un processus spontané d'utilisation de la langue avec une approche fonctionnelle-analytique inéluctable pour déterminer les matériaux de base, soit les éléments du code et les éléments du discours nécessaires au développement d'une compétence de communication. Pour se faire, il faut essayer de trouver une notion consensuelle des unités de signification. L'unité de signification que nous allons appeler désormais l'unité lexicale ne correspond pas toujours à la définition qui présente le mot comme un groupement de lettres, séparé, à gauche et à droite, par un blanc des autres éléments du texte. Elle peut inclure également des unités ou lexèmes un peu plus complexes telles que les mots composés, des proverbes ou des expressions imagées, il s'agit alors de la lexicalisation d'ensembles.

Afin de délimiter ces unités lexicales, Niklas-Salminen parle d'abord de critères phoniques mais il se rend compte vite que « en français, le rôle démarcatif de l'accent dans l'identification des mots est très faible par rapport aux accents de langues à accent lexical. ». Ensuite, sur le plan syntactico-sémantique, il parle des procédés d'*inséparabilité* et de *commutation*. En effet, en effectuant des manipulations dans le but de segmenter les unités de la chaîne parlée, on constate que le critère d'inséparabilité fonctionne pour les unités morphologiquement et graphiquement simples comme *père, famille, ministre* ainsi que pour les unités morphologiquement composées, mais graphiquement simple telles que *in-contrôl-able, em-poisonn-eur, ir-réproch-able*. Ce critère est essentiel dans la détermination de la lexicalité des unités graphiquement complexes comme *pomme de terre* ou *assistante sociale*. La cohésion de ces unités lexicales est tellement forte qu'on ne peut y insérer un autre élément. Il est impossible de dire par exemple *assistante très sociale* ou *pomme jaune de terre*.

Le critère de commutation renforce ce que démontre le critère d'inséparabilité. Lorsqu'une unité complexe est lexicalisée, (pomme de terre, assistance sociale, chemin de fer), c'est-à-dire sentie comme un mot, elle joue exactement le même rôle qu'une unité simple. »

L'assistante sociale reçoit les demandes.

Le fonctionnaire reçoit les demandes.

Niklas-Salminen finit par constater que le mot est « *avant tout une structure phonique et graphique stable, que l'on apprend à reconnaître et à reproduire.* »

Il est donc nécessaire que l'étude du vocabulaire en classe de langue ne se limite pas au mot en tant qu'unité simple.

[...] connaître un mot, ce n'est pas seulement être capable d'en donner une définition, ni même de pouvoir le situer dans son microsystème de relations paradigmatiques (antonymes, synonymes). C'est aussi connaître les propriétés distributionnelles des mots du point de vue de leur combinatoire sémantique, syntaxique, discursive, et savoir qu'il existe un nombre considérable de mots qui mobilisent toujours, ou presque toujours, d'autres mots qui les complètent, ou au niveau de l'ensemble du discours, qui leur font écho en les relayant ou en les caractérisant (Treville, M.-C, Duquette, L. 1996 : 15)

1.4 Qu'est-ce que connaître un mot ?

Il existe de nombreux types de connaissances nécessaires pour pouvoir utiliser un mot correctement et efficacement dans une langue étrangère.

Une position courante consiste à diviser la connaissance des mots en connaissances réceptives (passives) et en connaissances productives (actives). Généralement, on pense que les connaissances réceptives d'un apprenant, les mots qui sont reconnus lorsqu'ils sont entendus ou lus, sont supérieures aux connaissances productives d'un apprenant qui peuvent être rappelées à l'esprit et utilisées à l'oral ou à l'écrit. Adhérents à cette idée, certains ministères de l'éducation et concepteurs de matériels divisent leurs listes de mots en mots qu'ils s'attendent à ce que les apprenants connaissent passivement et ceux qu'ils s'attendent à ce que les apprenants connaissent activement. Krizsán (cité dans Milton, 2009) mentionne que le programme national hongrois de base, par exemple, suggère que les apprenants devraient apprendre quelque 1600 mots d'ici la 8^e année, dont 1200 devraient être connus activement et 400 autres passivement. Milton (2009, p.

13) précise que la distinction passive et active peut ne pas être aussi nette qu'elle n'y paraît au premier abord, car de bonnes compétences passives exigent souvent du lecteur ou de l'auditeur qu'il anticipe activement les mots qui se produiront.

Anderson et Freebody (1981) suggèrent de faire la distinction entre l'étendue de la connaissance des mots et la profondeur de la connaissance des mots. Pourtant, la distinction est d'une simplicité trompeuse. L'étendue des connaissances fait référence au nombre de mots qu'un apprenant connaît et la profondeur des connaissances fait référence à ce que l'apprenant sait de ces mots. Cela permet de faire une distinction entre les apprenants qui auraient appris beaucoup de mots par l'apprentissage par cœur de listes de traduction, mais ne sauraient pas vraiment comment les utiliser, et ceux qui auraient également appris comment les mots qu'ils connaissent s'associent à d'autres mots. Milton (2009, p. 14) voit que le concept de profondeur du vocabulaire est encore plus difficile à cerner, car il peut impliquer la connaissance des mots associés, de la collocation ou de la fonction des mots. De son côté, Meara (dans Milton, 2009) implique que cet aspect pourrait être réuni dans un cadre de liens entre les mots, quel que soit exactement ce lien.

Nation (2013) divise la connaissance des mots en trois domaines : la connaissance de la forme, connaissance du sens et connaissance de l'usage. Chacun de ces domaines est ensuite subdivisé.

La connaissance de la forme des mots peut impliquer la forme écrite d'un mot et la forme phonologique. Nation ajoute à ce domaine la connaissance des parties de mots par lesquelles il entend la connaissance des affixes que nous utilisons pour ajouter ou changer le sens d'un mot. Par exemple, cela inclut de comprendre que vous pouvez faire le contraire de plusieurs mots en ajoutant le préfixe *in* au début, comme dans connu et inconnu.

La connaissance du sens des mots est également divisée en trois parties. La première sous-division, forme et sens, est la partie à laquelle la plupart d'entre nous penseront en termes de connaissance d'un mot. Il s'agit de pouvoir lier la forme, quelle qu'elle soit, à un sens. Dans une langue étrangère, il s'agit souvent de faire le lien entre un mot en langue étrangère et sa traduction dans la langue maternelle. Cependant, les langues ne sont pas exactement parallèles les unes aux autres dans la façon dont elles utilisent leur vocabulaire. Les autres sous-divisions, concepts, référents et associations, indiquent donc qu'un mot dans une langue peut nécessiter plusieurs traductions ou porter des significations et des associations subtilement différentes dans une autre langue. Un mot tel que *fat* a des connotations très négatives en anglais lorsqu'il décrit une personne, et par conséquent, les locuteurs natifs doivent utiliser ce mot avec prudence. Mais,

dans certaines parties du sud du Nigeria, le concept a historiquement eu des connotations très positives ; il y avait des maisons d'engraissement dans le vieux Calabar, par exemple, où les femmes allaient prendre du poids pour être belles. Les utilisateurs d'anglais de ce milieu pourraient s'attendre à ce que le terme soit presque élogieux. Balhouq (cité dans Laufer, 1990) souligne que, dans un contexte de langue arabe musulmane, un « bistrot » n'est pas un endroit où les gens se réunissent pour prendre un verre entre amis, mais un endroit où les gens pèchent contre Dieu et la morale. Des mots comme 'holocauste', 'concentration' (même sans 'camp'), 'diaspora' ont une valeur émotive particulièrement forte lorsqu'ils sont traduits en hébreu.

Les apprenants en langues ont souvent besoin de connaître ce type d'informations s'ils ne veulent pas semer la confusion ou être offensés par un mauvais choix de mots.

La connaissance de l'usage des mots est également divisée en trois parties. Les fonctions grammaticales consistent à savoir quelle partie du discours est un mot et comment il se liera avec d'autres mots en conséquence. Si vous connaissez le mot noir, par exemple, et savez qu'il fonctionne comme un adjectif, alors en français il faut le placer après le nom qu'il qualifie et non avant : *une veste noire* et non *une noire veste*. Les apprenants anglophones de langues romanes telles que le français par exemple doivent s'adapter à la règle car les adjectifs de couleurs viennent toujours après le nom qu'ils qualifient. Certains mots apparaissent très fréquemment à côté d'autres et on dit que ces mots cohabitent les uns avec les autres, ce sont des collocations.

La distinction réceptive et productive s'adapte facilement à ce modèle. Nation divise chacune des subdivisions de son tableau en connaissances réceptives (indiquées par un R) et productives (indiquées par un P) et retient la convention qu'il existe une distinction mesurable entre ces deux types de connaissances. La distinction entre largeur et profondeur est moins nette. L'étendue du vocabulaire inclurait la catégorie Forme, mais pourrait également impliquer la sous-catégorie forme et signification de la section suivante. La profondeur du vocabulaire serait, par implication, tout le reste du tableau.

Dans le Tableau 1-1, Nation tente un résumé plus complet et systématique de ce que sont les différents types de connaissance des mots.

Tableau 1-1: Qu'implique la connaissance d'un mot?

Forme	parlée	R	À quoi ressemble le mot ?
		P	Comment se prononce le mot ?
	Écrite	R	A quoi ressemble le mot ?
		P	Comment le mot est-il écrit et orthographié ?
	Parties de mots	R	Quelles parties sont reconnaissables dans ce mot ?
		P	Quelles parties de mots sont nécessaires pour exprimer le sens ?
Sens	Forme et signification	R	Quel sens cette forme de mot signale-t-elle ?
		P	Quelle forme de mot peut être utilisée pour exprimer ce sens ?
	Concepts et référents	R	Qu'est-ce qui est inclus dans le concept ?
		P	À quels éléments le concept peut-il se référer ?
	Associations	R	A quels autres mots ce mot nous fait-il penser ?
		P	Quels autres mots pourrions-nous utiliser à la place de celui-ci ?
Utilisation	Fonctions grammaticales	R	Dans quels modèles le mot apparaît-il ?
		P	Dans quels modèles devons-nous utiliser ce mot ?
	Collocations	R	Quels mots ou types de mots apparaissent avec celui-ci ?
		P	Quels mots ou types de mots devons-nous utiliser avec celui-ci ?
	Contraintes d'utilisation	R	Où, quand et à quelle fréquence rencontrerions-nous ce mot ?
		P	Où, quand et à quelle fréquence peut-on utiliser ce mot ?

Source: Nation (2013, p. 49)

Daller et al. (2007) ont tenté de résumer ces aspects de la connaissance dans un espace théorique tridimensionnel qui oppose largeur et profondeur à une qualité de fluidité. La fluidité distingue la facilité et la vitesse avec lesquelles un apprenant peut accéder et utiliser les mots qu'il connaît, de la simple reconnaissance des mots et de la

connaissance de la façon de les utiliser. L'idée est que certains apprenants ont une grande maîtrise et peuvent utiliser n'importe quelle connaissance de la langue qu'ils ont facilement et sans hésitation et peuvent être très communicatifs, tandis que d'autres apprenants ont des difficultés à accéder aux mots qu'ils connaissent et les tentatives de communication sont caractérisées par de fréquentes pauses et hésitations tandis que le l'apprenant puise dans sa mémoire le mot ou l'expression qu'il essaie d'utiliser. Les détails manquent dans ce modèle, mais une façon de l'opérationnaliser consiste à supposer que l'étendue et la profondeur sont des aspects de la connaissance passive des mots, tandis que la fluidité fait référence à la connaissance productive des mots qu'un utilisateur possède.

Une chose qui semble évidente dans cette discussion est qu'un seul test ne pourrait pas espérer mesurer tous les aspects de la connaissance des mots. Il est difficile d'imaginer comment susciter la langue pour l'évaluation du vocabulaire productif tout en préparant simultanément les apprenants avec le lexique des langues étrangères pour évaluer les connaissances réceptives, par exemple. On pense généralement qu'il est nécessaire d'utiliser plusieurs mesures, différents tests et mesures, afin de recueillir les informations qui peuvent caractériser de manière exhaustive la connaissance du vocabulaire d'un apprenant. Par conséquent, lorsqu'ils tentent de mesurer la connaissance des mots, les chercheurs ont tendance à choisir une méthode viable qui peut caractériser un ou plusieurs éléments du mélange complexe décrit ci-dessus. Pour Milton (2009, p. 16), il s'avère souvent difficile de trouver une méthode qui puisse mesurer sans ambiguïté un seul élément de la connaissance du vocabulaire. Étant donné que les connaissances linguistiques doivent être évaluées indirectement, divers éléments de la connaissance du vocabulaire sont souvent évalués en même temps, et d'autres types de connaissances et de compétences peuvent être requis par l'apprenant, ce qui peut colorer le vocabulaire produit.

1.5 Qu'est-ce qu'un bon test de vocabulaire ?

Les estimations de la connaissance de vocabulaire sont toujours une question très relative. On doit être alors très prudent quant à leur traitement, en particulier lorsque les articles et les données sont plutôt anciens. Pendant longtemps, il n'y avait pas de tests standardisés dans le domaine des tests de vocabulaire, et les tests devaient être créés ad hoc. Il était souvent

impossible de comparer les résultats d'une expérience avec une autre de manière significative. Les conventions et le système de description de la connaissance des mots considérés ci-dessus sont un phénomène assez récent. Au cours des 20 dernières années environ, et même maintenant, les chercheurs peuvent varier dans les éléments de vocabulaire qu'ils excluent des dénombrements ou dans la façon dont ils définissent un mot si les circonstances semblent exiger un changement. Néanmoins, et selon Milton (2009, p. 17), il n'existe pas encore un ensemble complet de tests qui nous permettent de tester facilement et de manière fiable tous les aspects de la connaissance du vocabulaire d'un apprenant, mais nous avons un petit nombre de tests bien établis dans certains domaines, qui permettent des études à grande échelle, des comparaisons dans le temps et même des comparaisons inter-langues à faire. Comment ces tests sont-ils créés ? Habituellement, il y a deux questions principales à considérer dans la création de tout test. L'un est la fiabilité et l'autre est la validité. La fiabilité, au sens large, est la capacité d'un test à mesurer quelque chose de manière cohérente et précise. Cela inclut si le test est stable dans le temps, donc si, par exemple, un test de vocabulaire est administré à un apprenant deux fois le même après-midi alors que le vocabulaire de l'apprenant ne peut pas avoir changé de manière significative, alors il devrait donner le même score à chaque fois. Si c'est le cas, on dit qu'il est fiable et si ce n'est pas le cas, on pense qu'il n'est pas fiable. Tréville et Duquette ajoutent une troisième qualité. Il s'agit en fait de la « puissance discriminative ». Cette dernière se définit comme la « distinction effectuée par un test, entre les sujets capables de réussir certaines tâches et ceux qui ne le sont pas. La capacité d'un test d'effectuer cette distinction s'appelle son pouvoir discriminatif (ou puissance discriminative) » J-C Mothe (cité dans Tréville & Duquette, 1996).

L'évaluation de la fiabilité d'un test peut également inclure des estimations d'équivalence ; si différentes formes d'un test se comparent bien et produisent des résultats équivalents. Si l'apprenant qui passe deux tests de vocabulaire dans l'après-midi devait passer deux formes différentes de test, il devrait produire des scores qui montrent le même niveau de connaissances ou d'aptitudes et non des estimations différentes. Les estimations d'équivalence deviennent une préoccupation majeure dans les tests de langue, en particulier au Royaume-Uni mais aussi ailleurs, avec des accusations selon lesquelles les examens standard sont de plus en plus faciles à réussir. En règle générale, les tests dits objectifs de vocabulaire, les tests à choix multiples et les tests à réponse forcée ont une bonne fiabilité telle que mesurée par les méthodes de test et de retest, et certaines méthodes peuvent également produire de bons scores d'équivalence. Les tests subjectifs sont beaucoup moins satisfaisants et les tests de langue en font un usage intensif

à travers des questions de rédaction et des questions de compréhension ouvertes. Rédiger un essai dans une langue étrangère, par exemple, demande à l'apprenant non seulement de démontrer sa connaissance de la langue étrangère, mais aussi de choisir un style et un registre, et de faire un choix de contenu et de lectorat visé. Si ceux-ci n'étaient pas suffisants pour déstabiliser un test, alors pendant le processus de notation, l'évaluateur doit porter un jugement subjectif sur des éléments tels que le vocabulaire utilisé ; est-il étendu ou limité ou approprié à la tâche ? Beaucoup d'efforts sont actuellement déployés pour mesurer objectivement la langue produite par les apprenants dans des essais écrits ou des examens oraux, en utilisant par exemple des scores de richesse du vocabulaire, mais même ces méthodes ne peuvent pas compenser entièrement le fait que les essais et les entretiens oraux sont des moyens très indirects de mesurer connaissances linguistiques et le vocabulaire, et ne sont pas toujours de très bons moyens.

La validité répond à la question de savoir si un test mesure ce qu'il est censé mesurer et pas autre chose. Cela peut être un domaine complexe avec diverses questions à examiner.

La validité du contenu examine si un test a le contenu nécessaire et approprié pour mesurer ce qu'il est censé faire. Fréquemment, les tests d'étendue du vocabulaire utilisent des informations de fréquence comme base de sélection de mots pour les tests. Le test de niveau de vocabulaire de Nation (1990), par exemple, tire des mots des deuxième, troisième, cinquième et dixième bandes de fréquences de 1000 mots en anglais et de la liste de mots de l'université (Nation, 1990). L'hypothèse est que l'apprentissage sera concentré dans ces domaines et qu'un test de ce matériel fournira une bonne mesure globale de la connaissance du vocabulaire et de la capacité générale. Il utilise également des listes de mots lemmatisés comme base de sa sélection, ce qui reflète la croyance que les lemmes² ont une certaine réalité tant qu'unité de stockage dans l'esprit des apprenants (Levelt, cité dans Milton, 2009). Le test est généralement considéré comme ayant une bonne validité de contenu en raison de ces choix. Décider ce qui constitue une bonne validité de contenu pour d'autres aspects de la connaissance du vocabulaire peut être plus difficile. La profondeur du vocabulaire est un terme qui couvre un certain nombre de constructions distinctes et il n'est pas clair comment celles-ci sont liées les unes aux autres ou à l'étendue du vocabulaire. Un auteur, Vermeer (2001), suggère que la profondeur n'existe

² Dans le domaine linguistique, la notion de lemme désigne « la forme type de mot, adoptée par convention, permettant le regroupement, sous une même adresse lexicale (ou entrée), des diverses réalisations flexionnelles que ce mot peut connaître en discours. Par exemple, les formes *petit*, *petits*, *petites*, *petites*, ont pour lemme la forme *petit*. » (Neveu, 2004, p. 176)

peut-être pas vraiment en tant que construction distincte mais c'est une extension de l'étendue des connaissances, car les tests des deux sont si bien corrélés. En conséquence, les aspects individuels de la profondeur du vocabulaire, tels que la connaissance des collocations ou des idiomes, ont tendance à être mesurés séparément.

Afin de tester si un test de qualité nouvellement créé ou problématique, tel que le vocabulaire productif, fonctionne vraiment bien, deux tests différents de la même qualité sont souvent essayés sur les mêmes apprenants. Si le test fonctionne correctement, les résultats d'un test doivent bien se comparer aux résultats du second. C'est comme la mesure d'équivalence décrite ci-dessus, mais dans ce contexte, elle devient ce qu'on appelle un test de validité concurrente. C'est une méthode fréquemment utilisée pour aider à valider la construction et le contenu d'un test. Mais lorsque les deux tests utilisent des méthodes assez indirectes pour accéder aux connaissances linguistiques et que les performances sont influencées par d'autres domaines de connaissances, les corrélations entre les deux peuvent être modestes et il peut être difficile de conclure à quel point l'un ou l'autre des tests fonctionne vraiment (Fitzpatrick, 2007).

Enfin, il y a la question de la validité apparente, c'est-à-dire si le test est crédible pour les utilisateurs en tant que test de ce qu'il est censé mesurer. Les tests de vocabulaire peuvent exciter des passions surprenantes chez les utilisateurs, et même les tests avec une bonne validité de construction et de contenu peuvent être contestés par les apprenants. Les apprenants peuvent avoir des idées très précises sur ce à quoi devrait ressembler un test de langue, et ces tests n'impliquent pas toujours une mesure explicite du vocabulaire. Lorsque les rédacteurs de tests utilisent des données de fréquence et produisent des tests de connaissance du vocabulaire soigneusement ciblés, l'échelle comparativement petite et la simplicité des tests peuvent souvent susciter des doutes dans l'esprit des utilisateurs. Les avantages potentiels d'un test court et simple évitent la perte de concentration et d'autres complications peuvent être perdus pour les utilisateurs. Même les universitaires de domaines autres que les études de vocabulaire signalent un problème de crédibilité avec ces tests. La mesure du langage ne devrait cependant pas être une question de croyance, mais de collecte et d'évaluation de preuves empiriques.

1.6 Conclusion

Dans ce chapitre, nous avons tenté de résumer ce que l'on entend par un mot et ce qu'est la connaissance des mots. Il est essentiel pour la construction de tout bon test mesurant la connaissance du vocabulaire d'être sûr de ce qui est mesuré. Des questions concernant la façon dont la connaissance du vocabulaire peut être mesurée et les qualités d'un bon test ont été également examinées. Nous examinerons plus en détail dans les deux prochains chapitres les mots sélectionnés pour mesurer la connaissance du vocabulaire ; essentiellement en examinant la validité de contenu de la plupart des techniques de mesure modernes dans ce domaine. Sur cette base, les chapitres suivants examineront les détails de la construction de tests pour des qualités particulières de connaissance du vocabulaire, leur validité de construction, et appliqueront ces tests aux apprenants de langues étrangères.

CHAPITRE 2 :

DIFFICULTE DES MOTS, FREQUENCE ET ACQUISITION DES MOTS : PROFILS LEXICAUX

2.1 Introduction

Étant donné que la fréquence des mots et l'apprentissage sont si étroitement liés, il est généralement jugé nécessaire de cibler la connaissance des mots les plus fréquents dans une langue pour l'évaluation. Cela permet de construire une bonne mesure de la connaissance du vocabulaire qui fonctionne avec précision et d'une manière efficace.

Les mots peuvent varier de toutes sortes de façons. Ils peuvent varier dans les sons et les lettres qui les composent. Ils diffèrent également par leur longueur, la façon dont les sons et les lettres se combinent et à quel point ils sont similaires à la langue maternelle d'un apprenant. Ils peuvent différer dans la façon dont ils changent et créent des formes dérivées ou infléchies, telles que les pluriels et les temps passés. Et ils peuvent varier dans la gamme de nuances et de sens qu'ils véhiculent et, par conséquent, dans les situations où ils sont utilisés. Incontestablement, ceux-ci peuvent tous influencer si, et dans quelle mesure un mot est appris. Ces types de différences entre les mots ont fait l'objet d'études approfondies, généralement sous l'idée générale de la charge d'apprentissage ; ce qui rend un mot difficile ou facile à apprendre.

Cependant, la différence la plus importante entre les mots est peut-être la fréquence à laquelle ils apparaissent. C'est important parce que la fréquence détermine quels mots un apprenant est susceptible de rencontrer et à quelle fréquence ils sont rencontrés. Milton (2009) pense que ce facteur crée des différences générales quant au moment où un mot est susceptible d'être appris : certains mots ont tendance à être appris au début de l'apprentissage de la langue et d'autres sont plus susceptibles de l'être plus tard. Il est essentiel de comprendre le processus d'apprentissage du vocabulaire et de déterminer dans quelle mesure ce processus peut être étudié et mesuré. Si nous savons quels mots sont susceptibles d'être appris et quels mots ne le sont pas, alors il est possible de construire des tests de connaissance du vocabulaire bien meilleurs que cela ne serait possible autrement. Nous tenterons d'examiner dans ce chapitre la relation entre la fréquence d'occurrence d'un mot et l'apprentissage.

2.2 Différences dans la fréquence des mots

Dans le langage normal, nous utilisons certains mots beaucoup plus souvent que d'autres, et un bref balayage de n'importe quelle page de texte, servira généralement à illustrer cela. Des mots tels que les articles *le/la/l'/les* et *un/une/des*, des prépositions telles que *dans* et *de*, des conjonctions telles que *et* et des pronoms tels que *ce*, se produisent très fréquemment. Ils se produisent des millions de fois dans la plupart des grands corpus. D'autres mots tels que *curiosité* et *gravier* sont beaucoup moins fréquents. Le tableau 2.1 répertorie, par ordre de fréquence, les 20 mots les plus fréquents dans le *British National Corpus* (BNC) et 20 autres mots du début de la bande de fréquence de 5000 mots, pour aider à illustrer la différence.

Tableau 2-1 : Les mots les plus fréquents, et les mots de la bande des 5000 mots, et leurs occurrences dans le BNC

Ordre	Occurrences	Mot	Ordre	Occurrences	Mot
1	6,187,2	the	5001	1188	regulatory
2	4,239,6	be	5002	1188	cylinder
3	3,093,4	of	5003	1187	curiosity
4	2,687,8	and	5004	1185	resident
5	2,186,3	a	5005	1185	narrative
6	1,924,3	in	5006	1185	cognitive
7	1,620,8	to	5007	1184	lengthy
8	1,375,6	have	5008	1184	gothic
9	1,090,1	it	5009	1184	dip
10	1,039,3	to	5010	1184	adverse
11	887,87	for	5011	1184	accountability
12	884,59	I	5012	1183	hydrogen
13	760,39	that	5013	1183	gravel
14	695,49	you	5014	1182	willingness
15	681,25	he	5015	1182	inhibit
16	680,73	on	5016	1182	attain
17	675,02	with	5017	1181	specialise
18	559,59	do	5018	1180	steer
19	534,16	at	5019	1180	selected
20	517,17	by	5020	1180	like

Source: Kilgariff (2006)

Le tableau 2-1 permet également d'illustrer que les mots les plus fréquents sont presque toujours des mots de fonction ou de structure, qui semblent avoir eux-mêmes peu de poids de sens, mais sont cruciaux pour créer un langage grammatical et significatif. Les mots moins fréquents ont tendance à être des mots de contenu ou lexicaux, des noms, des verbes principaux et des adjectifs qui, pourrait-on penser, semblent porter une plus grande charge de sens dans n'importe quelle phrase. Les deux sont importants, bien sûr, et les deux sont essentiels à la maîtrise d'une langue étrangère. Les mots les plus fréquents ont également tendance à s'interconnecter beaucoup plus fréquemment.

Le tableau 2-1 montre également à quel point les mots peuvent être différents dans leur fréquence d'apparition. Les mots fréquents, ceux du haut de la liste du tableau 2-1, sont beaucoup plus fréquents que ceux du bas ou même du milieu de la liste. Les mots les plus fréquents apparaissent des millions de fois dans le BNC, tandis que la sélection de mots de la bande de 5000 mots se produit environ 1200 fois. Milton (2009) a essayé d'illustrer l'ampleur de la différence sur la figure 2-1 : il a puisé dans le corpus de Baudot (1992) du français moderne (environ 1,1 million de mots) et calculé le nombre moyen de fois que des mots dans les cinq premières bandes de fréquence de 1000 mots apparaissent.

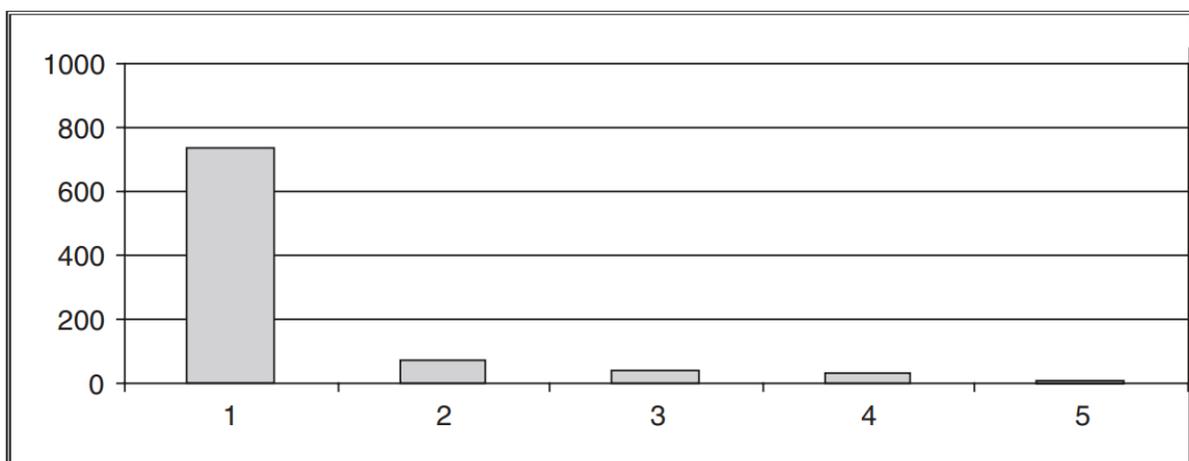


Fig. 2-1: Fréquence moyenne des mots dans les bandes de fréquences de Baudot (1992), Milton (2009)

Milton constate alors que les mots de la première bande de 1000 mots sont environ 11 fois plus fréquents, en moyenne, que les mots de la deuxième bande, et 40 fois plus fréquents que les mots de la cinquième bande de fréquences. Les mots de la première bande apparaissent en moyenne environ 800 fois dans le corpus, les mots de la deuxième bande 74 fois et les mots de la cinquième bande seulement 20 fois en moyenne. Le mot le plus fréquent du corpus, *de*, apparaît plus de 40 000 fois et le second, l'article défini *le*, plus de 25 000 fois. La majorité des mots de ce corpus n'apparaissent qu'une ou deux fois. Avec des différences de cet ordre, il n'est pas surprenant que les apprenants se familiarisent très rapidement, au moins dans une certaine mesure, avec certains des mots les plus fréquents d'une langue.

2.3 Le modèle de fréquence : dessiner un profil de fréquence

On suppose généralement qu'il existe une forte relation entre la fréquence d'un mot et la probabilité qu'un apprenant le rencontre et l'apprenne. C'est une idée très importante dans la mesure de l'acquisition du vocabulaire en langue seconde (L2) car c'est un principe sur lequel sont basés les tests d'étendue de vocabulaire les plus couramment utilisés. L'idée n'est pas nouvelle et remonte à environ 100 ans à la méthode scientifique pré-structuraliste dans l'enseignement des langues. Palmer (1917, p. 123) par exemple, a écrit que « ... *the more frequently used words will be the more easily learnt...* ». Les auteurs ultérieurs acceptent cela comme une évidence ; par exemple, Mackey (1965) et McCarthy (1990) répètent sans réserve l'idée de Palmer. Cependant, malgré l'usage répandu de l'idée au cours des dernières années, ce modèle de fréquence n'est pour Wesche et Paribakht (1996, p. 14), qu'une hypothèse et il semble n'avoir été démontré que très récemment.

Heureusement, l'un des avantages de cette idée est qu'elle peut être transformée en un modèle qui peut ensuite être testé empiriquement. Meara (1992) le fait en traçant un graphique de la relation et en produisant un profil de fréquence qui, selon lui, devrait ressembler à la figure 2-2.

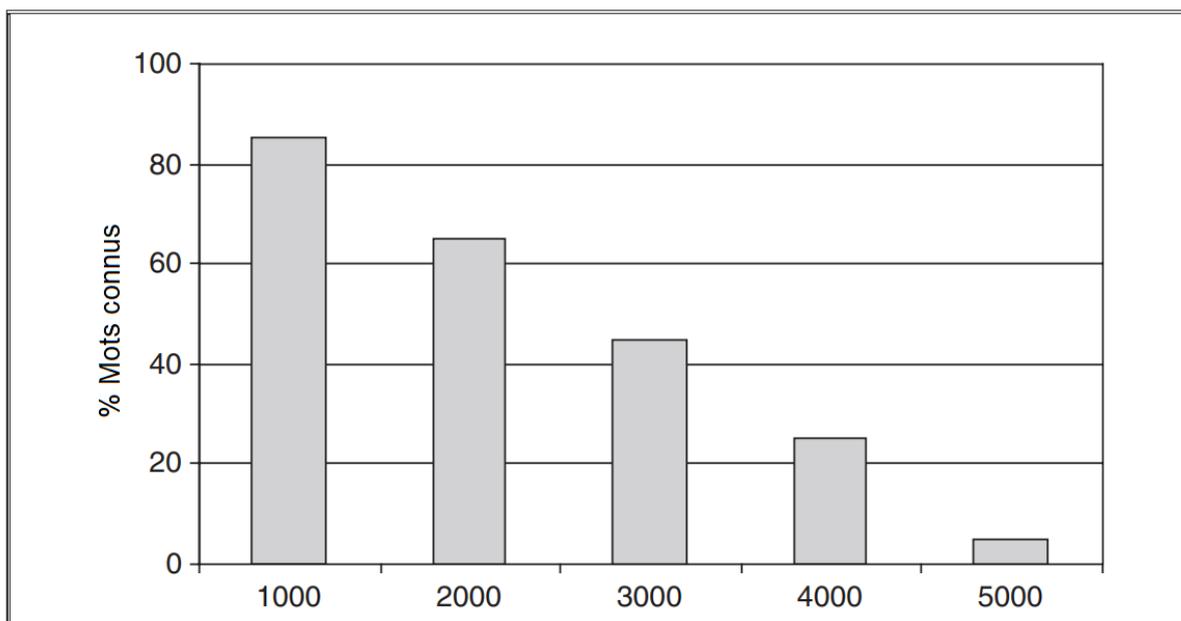


Fig. 2-2 : Profil de vocabulaire d'un apprenant type (Meara, 1992, p. 14)

La colonne 1 représente la connaissance des mille premiers mots les plus fréquents dans une langue, la colonne 2 les 1000 mots les plus fréquents suivants, et ainsi de suite. Selon Meara, les connaissances d'un apprenant typique sont élevées dans les colonnes fréquentes et inférieures dans les colonnes moins fréquentes, ce qui donne une pente descendante distinctive de gauche à droite. Les apprenants auront tendance à en savoir plus sur les mots à haute fréquence que sur les mots à basse fréquence. Au fur et à mesure que les connaissances de l'apprenant augmentent, le profil se déplace vers le haut jusqu'à atteindre un plafond à 100 % (tous les mots de cette bande de fréquence sont connus) lorsque le profil s'aplatit aux niveaux les plus fréquents et la pente descendante, de gauche à droite, se déplace vers la droite dans des bandes de vocabulaire moins fréquentes.

2.4 Test du modèle et dessin des profils de fréquence réelle

Des études récentes ont testé ce modèle et ont trouvé qu'il semble être extrêmement robuste, au moins en termes de caractérisation des populations d'apprenants. Milton (2006b) a mené une étude sur l'ensemble des 227 apprenants d'une école de langues en Grèce, avec des capacités allant de débutants à des apprenants intermédiaires supérieurs de l'anglais, expressément pour tester l'hypothèse de fréquence. Les apprenants ont été soumis à un test de

reconnaissance du vocabulaire orthographique Oui/Non avec 20 mots de test tirés de chacune des cinq premières bandes de fréquences de 1000 mots que Meara a incluses dans son modèle. Le test utilisé était X-Lex (Meara & Milton, 2003), qui utilise des listes de fréquences établies par Nation (1984) et Hindmarsh (1980). Les mots que les apprenants ont identifiés comme connus dans chaque bande ont été calculés séparément et les scores moyens pour chaque bande de fréquence ont été produits. Les résultats, lorsqu'ils sont représentés graphiquement, ont produit un profil haut à gauche et s'effilant vers la droite, comme le suggérait le modèle Meara. Ceci est illustré à la figure 2-3.

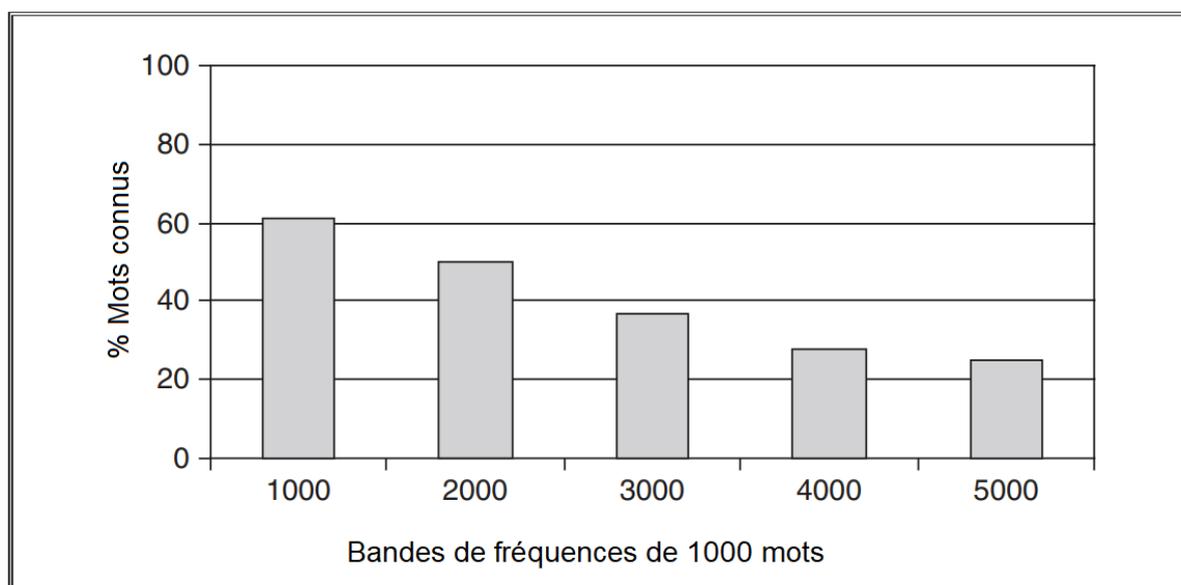


Fig. 2-3: Profil de fréquence pour les apprenants grecs d'EFL (Milton, 2006a)

Milton cite qu'une ANOVA confirme qu'il existe une relation statistiquement significative entre les bandes de fréquences et les scores de taille de vocabulaire. Ce type de preuve apporte un soutien considérable au modèle de fréquence de l'apprentissage et à l'idée que plus un mot est fréquent, plus il a des chances d'être appris, en règle générale.

Cet effet ne se limite pas aux apprenants de l'anglais comme langue étrangère (EFL), mais a également été observé chez les apprenants du français dans les écoles britanniques. Richards et Malvern (cité dans Milton 2009, 27) rapportent des données recueillies par Helen Bradley auprès d'un petit échantillon d'étudiants en français de 17 ans, ce qui est illustré à la figure 2-4.

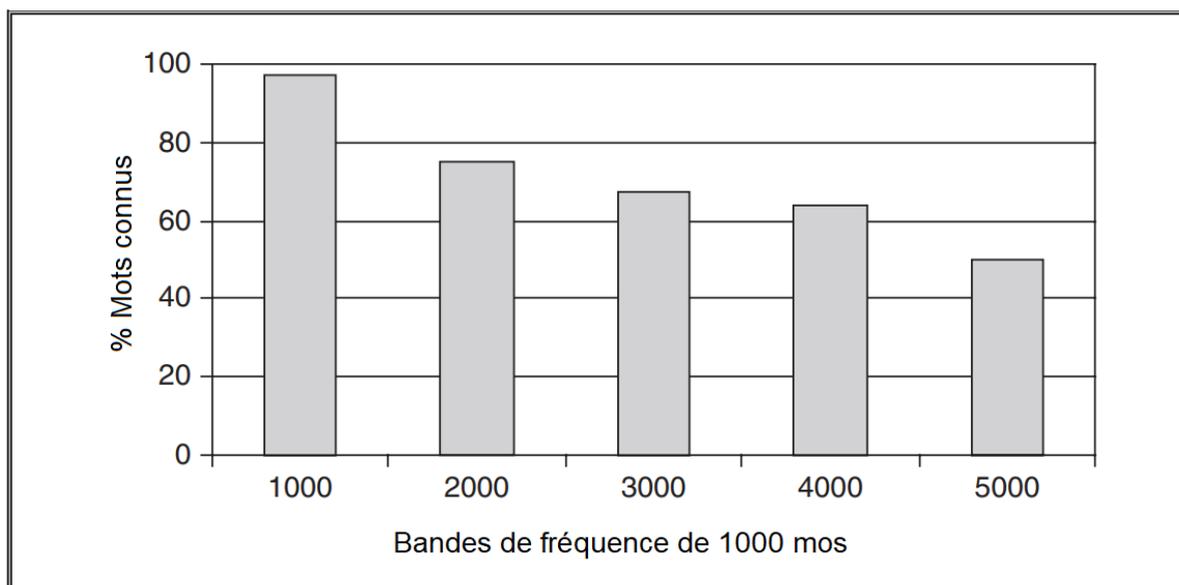


Fig. 2-4 : Profil de fréquence des apprenants britanniques du français (Richards & Malvern, 2007)

Les données ont été collectées avec une adaptation française du test X-Lex (Meara & Milton, 2003) utilisé dans l'étude grecque et conçu pour être structuré de la même manière à partir des informations de fréquence tirées de Baudot (1992).

Sur la base de ces graphiques, Milton (2009) explique que les profils, qui ne sont pas des droites, sont plus raides à gauche entre les bandes 1 et 2 et plus plats à droite. Cela devrait refléter les différences dans les fréquences des mots au sein de chaque bande différente. Les mots de la colonne 1 sont comparativement beaucoup plus fréquents que les mots de la colonne 2, et ainsi de suite, et dans les bandes les moins fréquentes, cette différence s'égalise et la fréquence des mots devient beaucoup plus similaire.

Deux mises en garde s'imposent ici, note Milton (2009, p. 29). La première est qu'il s'agit d'une tendance et non d'une règle absolument rigide. Un mot plus fréquent ne sera pas forcément appris avant un mot moins fréquent. Les matériels d'apprentissage des langues sont généralement organisés par thème et les mots à enseigner ne peuvent pas être sélectionnés uniquement en fonction de la fréquence. Inévitablement, les apprenants apprendront les mots thématiquement importants en plus des mots fonctionnels très fréquents. Ainsi, les jeunes apprenants ont tendance à apprendre des mots tels que girafe et tigre au début de l'apprentissage, même si ceux-ci sont très peu fréquents dans les corpus d'adultes. Ils font partie de l'image du monde de l'enfant et l'apprentissage des noms d'animaux est un point de départ commun pour

de nombreux cours de langue destinés aux jeunes apprenants. Le deuxième point est que les informations de fréquence ne fournissent pas d'informations sur la difficulté. En fait, il est possible de faire valoir que certains des mots les plus fréquents sont, à bien des égards, les mots les plus difficiles à utiliser pour les apprenants d'anglais avec une parfaite maîtrise. Ce que les informations sur la fréquence nous disent, c'est quelle est la chance qu'un apprenant rencontre un mot et le réemploie assez souvent pour qu'il soit appris.

2.5 Le développement des profils au cours de l'apprentissage

Les preuves suggèrent donc que le modèle de Meara (1992) d'un profil de fréquence est précis dans la façon dont il caractérise la croissance du vocabulaire dans les groupes d'apprenants, la connaissance des mots se concentrant dans les bandes les plus fréquentes. Son idée est-elle également correcte selon laquelle le profil sera similaire à tous les niveaux et augmentera avec des connaissances et des niveaux accrus jusqu'à ce qu'il atteigne 100 % de connaissance d'un groupe, lorsqu'un plateau émergera ? Il n'est pas inévitable que ce soit le cas. Le profil de vocabulaire a ses bandes de fréquences tirées de très grands corpus linguistiques. Mais les textes individuels au sein de chaque corpus peuvent varier de l'un à l'autre et de la liste de fréquences globale qu'un corpus produit. Les textes individuels varieront car ils sont susceptibles de traiter de sujets et de thèmes différents. Les éléments lexicaux présents dans un texte sur les animaux, par exemple, sont susceptibles d'être absents dans un texte sur la mécanique automobile. Lorsque beaucoup de ces éléments sont rassemblés dans un corpus, l'effet est que les éléments de vocabulaire structurel, qui apparaissent dans tous ou plusieurs textes, se frayent un chemin vers le haut de la liste de fréquence générale, tandis que les noms d'animaux et les pièces de voiture, absents de la plupart des textes, s'enfoncent quelque part vers le bas. L'effet est visible dans le tableau 2-2.

Tableau 2-2 : Comparaison des lexiques les plus fréquents des livres pour enfants (Vassiliu, cité dans Milton, 2009), des textes automobiles (Milton & Hales, 1997) et du BNC

Ordre	BNC	Manuels de voiture	Livres EFL pour enfants
1	the	and	is
2	be	the	the
3	of	to	a
4	and	of	your
5	a	in	no
6	in	is	look
7	to	or	where
8	have	with	for
9	it	remove	but
10	to	a	did
11	for	replace	from
12	I	for	say
13	that	oil	as
14	you	be	very
15	he	valve	dog

Alors que le BNC est un corpus plus important que les deux autres, ce qui est susceptible d'exagérer la tendance des mots fonctionnels à prédominer dans les matériaux tirés de tous les domaines thématiques, l'individualité des deux autres ensembles de matériaux est très claire.

Le corpus des manuels automobiles, d'environ 100 000 mots, comprend un vocabulaire tel que l'huile et la soupape avec des degrés de répétition très élevés qui ne seraient pas attendus en dehors de ce registre. Travailler dans ce genre implique également une répétition élevée d'actions particulières, d'où la fréquence remarquable de vérification, de suppression et de

remplacement. Ces mots, bien sûr, n'ont pas leur place dans les textes d'enseignement de l'EFL³ conçus pour les jeunes apprenants, et l'échantillon de ces matériaux de Vassiliu (cité dans Milton, 2009) produit des mots lexicaux très différents. Des mots tels que *regarder* et *dire* font clairement partie de la rubrique des manuels de cours et donnent des instructions aux apprenants en leur disant quoi faire. *Chien* est un produit du contenu thématique du matériel qui se concentre sur des idées appropriées pour les très jeunes apprenants, dans ce cas les animaux. Néanmoins, il est clair aussi que même dans ces petits échantillons de textes très restreints aux sujets, les mots fonctionnels de l'anglais commencent encore à prédominer.

Il semble donc possible qu'au début de l'apprentissage, le profil puisse changer par rapport au profil de Meara car l'exposition à la langue n'aura été qu'à un, ou à un petit nombre de textes linguistiques. L'exposition au langage n'a peut-être pas été suffisante pour que les caractéristiques du langage normal émergent. Cette variation se produit-elle et, si c'est le cas, à quel moment une exposition linguistique suffisante a-t-elle été atteinte pour que le profil normal se réaffirme ? Pour répondre à cette question, les données de 227 apprenants grecs de l'anglais (Figure 2.3) ont été réanalysées et des profils de vocabulaire ont été établis pour chacune des sept classes de junior, les débutants, jusqu'au *First Certificate in English* (FCE), en passant l'examen de Cambridge. Les résultats sont présentés à la figure 2-5.

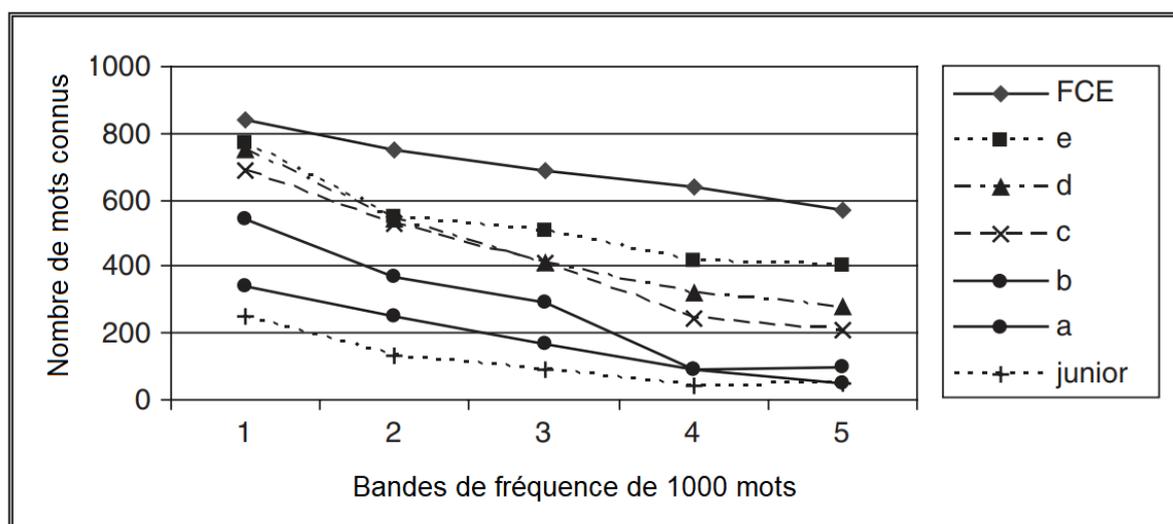


Fig. 2-5: Profils de fréquence pour les apprenants grecs d'EFL divisés par classe

³ English as a Foreign Language

De manière peut-être surprenante, ces données suggèrent que le profil de fréquence de Meara peut émerger même aux premiers stades de l'apprentissage. Après seulement 100 heures de classe dans ce cas, et lorsque l'on peut s'attendre à ce que l'entrée soit au moins typique, le profil de fréquence est presque complet. À partir de ce niveau bas, le profil remonte le graphique, comme l'avait prédit Meara. Il n'est pas clair à partir de ces données à quel moment ou si un plateau est atteint à 100% de la connaissance dans n'importe quelle bande de fréquence. Même le groupe de tête a une moyenne dans les bandes les plus fréquentes quelque peu inférieure au maximum.

Afin d'examiner si les apprenants atteignent régulièrement les 100 % dans les tranches les plus fréquentes, les profils individuels des 10 élèves les mieux notés à Milton (2006a) ont été établis et sont présentés dans la figure 2-6.

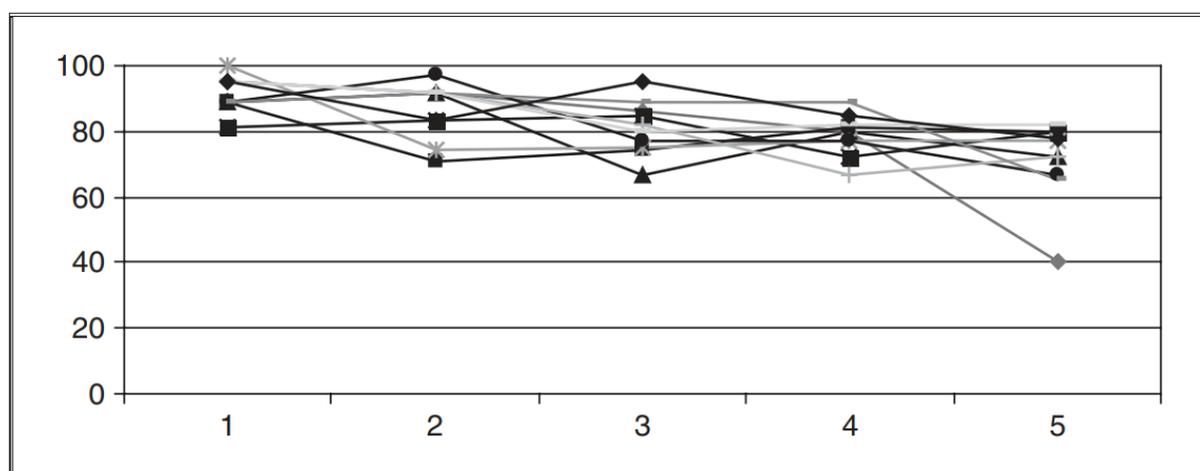


Fig. 2-6: Profils individuels des 10 étudiants les mieux notés à Milton (2006a)

Ce chiffre est, à bien des égards, très insatisfaisant, confus et peu clair avec de nombreuses lignes qui se chevauchent, mais il remplit son objectif. Ce qui ressort de ce fouillis, c'est que les apprenants les plus capables semblent atteindre un plateau dans les bandes les plus fréquentes, non pas à 100 % comme prévu, mais à un niveau plutôt inférieur, environ 85 à 90 % dans ces cas. Il n'est pas évident pourquoi le profil se stabilise à ce stade. Il se peut que certains des mots, même les plus fréquents, aient des qualités qui les rendent imperméables aux effets des processus normaux d'enseignement et d'apprentissage. Alternativement, il peut y avoir un problème dans la sélection du vocabulaire à inclure dans les manuels, ce qui signifie que certains vocabulaires très fréquents doivent être omis et ne peuvent donc pas être appris. Ou il se peut que cette

observation soit simplement le sous-produit des idiosyncrasies de l'apprenant et que chaque apprenant réussisse à éviter de rencontrer et d'apprendre un ensemble de mots différent. Pour répondre à cette question, un examen plus détaillé des mots inconnus dans ces bandes de fréquences serait nécessaire.

D'autres études ont étayé l'hypothèse stipulant que les tests de vocabulaire se concentrant dans ces bandes fréquentes fourniront des données utiles à tous les stades de l'apprentissage après les 100 premières heures de cours et que les entrées soient de bonne qualité.

2.6 Profils de vocabulaire chez les apprenants individuels

En s'appuyant sur des données provenant de groupes d'apprenants et basées sur le modèle de fréquence de l'apprentissage, Milton (2009) suggère que les apprenants, en tant que groupe, sont sensibles à la fréquence d'occurrence des mots qu'ils rencontrent. Cependant, les groupes sont constitués d'apprenants individuels, et il n'est en aucun cas inévitable que chaque apprenant soit identique dans sa réaction à la fréquence des mots. Chaque enseignant sait que si chaque élève d'une classe reçoit le même temps de classe, suit les mêmes manuels de cours, entend le même enseignant parler et fait les mêmes exercices, nul ne pourra prédire ce que les élèves apprendront et combien ils apprendront.

Milton (2009) note que si la fréquence des mots doit être utilisée comme base pour la construction des tests de vocabulaire et pour la mesure des connaissances, alors cette variation individuelle est potentiellement importante. Milton (2009) ajoute qu'il existe un risque que les tests fonctionnent bien pour certains ou la plupart des apprenants sensibles à la fréquence, mais qu'ils fonctionnent mal avec ceux qui apprennent différemment. Il vaut la peine de se demander si certains apprenants sont moins sensibles à la fréquence des mots que d'autres. La réponse à cette question suggérera à quel point le contenu et la validité de construction des mesures de vocabulaire basées sur les informations de fréquence seront forts.

Un test de Friedman sur les résultats des apprenants grecs de l'anglais à Milton (2006a), illustré à la figure 2-3, confirme l'impression que les apprenants individuels ont tendance à être influencés par la fréquence des mots et que la tendance globale est en effet très forte dans une population dans son ensemble. Mais même de fortes relations de ce type peuvent masquer une variation systématique et n'implique pas que l'assimilation du vocabulaire de chaque apprenant produise un profil de vocabulaire parfaitement régulier.

Milton (2009) déduit qu'environ 60% des apprenants, 132 sur 227, dans ce groupe ont exactement le type de profil de fréquence que le groupe affiche dans son ensemble. Ce n'est que dans les classes inférieures et supérieures que les apprenants avec des profils non fréquents sont plus nombreux que ceux avec des profils plus réguliers. Parmi les débutants, la présence du profil de fréquence est quelque peu surprenante étant donné le peu d'apports que ces apprenants auront reçu et la façon dont les petits corpus, tels que les manuels pour débutants, diffèrent des très grands corpus où nous prenons nos informations de fréquence. Dans la classe la plus élevée, il apparaît que les effets de plafond influencent les profils. Les apprenants les plus capables ont des scores dans les bandes les plus fréquentes qui ont atteint un maximum et le profil de fréquence n'est plus observable à mesure que l'apprentissage se déplace vers les bandes les moins fréquentes. Pourtant, une minorité d'apprenants n'affichent pas le profil de fréquence dans les mots qu'ils apprennent. Milton (2009) note qu'il est possible que cette variation indique des différences systématiques dans la façon dont les apprenants acquièrent le vocabulaire.

2.7 Difficulté des mots et comment cela influence le modèle de fréquence

Il est clair que pour Milton la fréquence des mots est le facteur le plus important pour aider à comprendre quels mots sont susceptibles d'être appris et quand. Les preuves qui ont été recueillies dans ce chapitre confirment à quel point la fréquence des mots semble être influente. Dans le même temps, il a été suggéré qu'il existe un certain nombre d'autres facteurs, associés à la difficulté des mots et à la charge d'apprentissage, qui contribueront également à influencer les types de mots appris. Potentiellement, ces facteurs pourraient déstabiliser le modèle de fréquence. Quels sont ces facteurs et quelle influence ont-ils sur l'apprentissage du vocabulaire ? Est-il possible, ou nécessaire, d'intégrer des caractéristiques de difficulté des mots dans un modèle général d'apprentissage des mots pour créer un test mieux dirigé de la connaissance du vocabulaire et permettre une meilleure compréhension des mesures qui ressortent de ces tests ?

La source de difficulté potentielle la plus évidente pour un apprenant est peut-être la forme du mot lui-même. Si le nouveau mot contient des combinaisons inhabituelles de lettres ou de sons, ou est difficile à prononcer, alors il semblerait moins susceptible d'être retenu en mémoire et rappelé pour utilisation, qu'un nouveau mot qui ne possède pas ces caractéristiques. En Algérie, ce problème est plus saillant quand on sait que le système phonologique du français est complètement différent de celui de l'arabe dialectal. Il semblerait évident que dans cas,

l'apprenant sera incapable de retenir des mots dont il ne sait pas reproduire les sons. Cette idée est soutenue par l'étude de Rodgers (1969) sur des étudiants anglophones de russe. Cette étude a rapporté que les mots avec des combinaisons de sons non anglais et qui étaient difficiles à prononcer, n'étaient pas appris aussi bien que les mots qui étaient plus faciles à prononcer. Les locuteurs de langues sémitiques, telles que l'arabe, ont tendance à se concentrer sur les consonnes en lecture et ont par conséquent des difficultés dans d'autres langues où les voyelles sont importantes. Ryan (cité dans Milton 2009) signale des confusions d'écriture en anglais, telles que *wheel* et *wells*, *left* et *lift*, et *present* et *prison*, où les consonnes restent relativement peu affectées par l'erreur, mais les voyelles sont souvent mal positionnées, omises ou substituées à l'un l'autre. Les mots qui sont similaires dans la forme, ou du moins qui semblent similaires à l'apprenant peuvent ainsi être confondus. Laufer (1997, p. 148) par exemple, se souvient de son expérience des apprenants que *available* est souvent interprétée à tort comme étant *valuable*, *embrace* comme *embarrass* et *simulate* comme *stimulate*. En production, elle souligne que des paires telles que *thinking* et *sinking*, *price* et *prize*, et *cute* et *acute*, peuvent transmettre en production un message très différent de celui qui était prévu. Un système d'écriture L1 différent peut également être responsable de certains problèmes d'apprentissage. Les locuteurs natifs des langues sémitiques qui accordent une grande importance aux consonnes et représentent à peine les voyelles dans l'écriture, ont tendance à confondre les mots avec des consonnes similaires et des voyelles différentes.

Un autre facteur largement pris en compte est la similitude entre le mot cible dans une nouvelle langue et son équivalent dans la propre langue de l'apprenant. Gairns et Redman (Cité dans Milton, 2009) illustrent cette idée avec des termes apparentés tels que *taxi*, *bar* et *hôtel*. Ce sont des éléments de vocabulaire courants dans de nombreuses langues européennes et les apprenants français, espagnols ou italiens de l'anglais auraient peu de difficulté à reconnaître ces mots s'ils les rencontraient sous forme écrite. La prononciation peut être différente dans ces langues, ce qui peut rendre difficile l'apprentissage d'une prononciation correcte, mais la forme écrite est souvent identique et l'apprentissage devrait, pourrait-on penser, nécessiter très peu d'efforts de la part de l'étudiant. En Algérie, les emprunts qu'on retrouve dans le dialecte arabe comme "ماصو" ou encore « "بلومي" ou "طاكسي" joueraient un rôle favorable à l'acquisition du vocabulaire. Tharp (1934, pp. 129-130), rapportant les conclusions d'une thèse rédigée par Limper, met en garde contre une généralisation trop large de l'effet sur l'apprentissage des cognates. Il semble que les apprenants peuvent être très idiosyncratiques dans leurs réactions

à ceux-ci et pas seulement dans la manière dont deux mots peuvent apparaître apparentés à un apprenant, mais complètement déconnectés l'un de l'autre. Il semble que les apprenants soient également imprévisibles dans la manière dont une dérivation d'un mot peut être reconnue comme apparentée alors qu'une dérivation différente du même mot n'est pas reconnue.

Gairns et Redman (cité dans Milton, 2009) suggèrent également que les éléments concrets qui peuvent être représentés visuellement ou simplement démontrés peuvent également être plus économiques à enseigner et à apprendre que les éléments et les idées abstraits. Un élément tel qu'une table, qui peut être vu et touché, peut être plus facile à apprendre que, disons, la philosophie ou l'intuition. Une partie du problème ici peut être de savoir s'il existe un équivalent de traduction directe entre la langue cible et la langue maternelle. Il y a aussi des domaines où il n'y a pas de traduction directe et claire et à mesure que les niveaux de sophistication augmentent, le traitement du sens et de la forme devient une activité qui prend beaucoup de temps. Les enseignants de langues, en particulier ceux qui enseignent une langue à des fins spécialisées ou académiques, seront habitués à faire face à des apprenants qui ne connaissent pas la matière de la langue qu'ils apprennent. Enseigner des mots tels que *carburateur* et *transmission* peut être relativement simple si les apprenants sont des mécaniciens automobiles, ils savent exactement de quoi on parle, mais pour quelqu'un qui n'a aucune connaissance des moteurs de voiture, ces termes deviennent des amas de lettres ou de sons dénués de sens, et donc, très oubliables.

Laufer (1997, p. 144) rapporte que, intuitivement, il semblerait que les mots plus longs devraient porter un fardeau d'apprentissage plus important que les mots plus courts.

Les preuves sont ambiguës et certaines études ne parviennent pas à trouver un tel lien (par exemple Rodgers, 1969). Comme le souligne Laufer, la longueur est un facteur difficile à isoler dans les études car il entre souvent en conflit avec la transparence morphologique des mots plus longs. Des mots tels que *mismanagement*, suggère-t-elle, bien qu'ils soient plus longs, comprennent plusieurs morphèmes familiers qui permettent aux apprenants de le reconnaître plus facilement et, vraisemblablement, de le mémoriser.

Une partie du discours peut également jouer un rôle en influençant la charge d'apprentissage d'un mot. Les noms, semble-t-il, sont plus faciles à apprendre que les verbes, qui sont plus faciles que les adjectifs. L'étude de Horst et Meara (1999) sur un seul apprenant faisant plusieurs lectures d'une bande dessinée a examiné les classes de mots que l'apprenant a acquis et ces résultats sont résumés dans le tableau 2-3.

Tableau 2-3 : Absorption du vocabulaire par classe de mots

Noms	Verbes	Adjectifs	Adverbes
n=51	n=32	n=20	n=8
63%	38%	35%	13%

Source : Adapté de Horst et Meara (cité dans Milton, 2009)

Tous ces facteurs peuvent affecter les mots susceptibles d'être retenus en tant que mots, de sorte qu'une forme correcte d'un mot dans une langue étrangère puisse être associée à un élément, un concept ou une traduction en première langue (L1). Un mot apparenté, même s'il est très peu fréquent et n'est rencontré qu'une seule fois, peut être appris parce qu'il est si similaire à l'élément de la langue maternelle, tandis que les éléments très fréquents, s'ils sont longs ou complètement différents de la langue maternelle, peuvent prendre beaucoup plus de rencontres avant d'être mémorisés correctement. Il existe d'autres causes potentielles de difficulté, telles que les différences de connotations, d'associations et de collocations entre les mots de la langue maternelle et cible, qui ne seraient pas détectées par un test de vocabulaire Oui/Non de connaissance passive des mots isolément. Ces types de facteurs ont tendance, comme le souligne Laufer (1990, p. 574) à être étudiés séparément et isolément les uns des autres. L'effet cumulatif de ces facteurs peut être observé. Milton (2009) souligne que très souvent aussi, les investigations dans ce domaine reposent sur des analyses de très petits nombres de mots et avec un petit nombre d'étudiants, et il se demande comment les conclusions auxquelles elles aboutissent sont généralisables dans le processus de construction d'un lexique complet avec, potentiellement, plusieurs milliers de mots. Même la contribution considérable de Laufer s'arrête avant d'essayer de construire la mesure de ces facteurs dans un modèle unique d'apprentissage des mots.

Cependant, il devrait être possible de prendre tous ces facteurs séparés et de les opérationnaliser dans un seul modèle. Milton et Daller (2007) ont tenté d'intégrer certains facteurs pour calculer dans quelle mesure chaque facteur contribue à déterminer si un mot est susceptible d'être appris ou non. Cette tentative est en quelque sorte une expérience, donc en plus de la fréquence, ils n'ont ajouté que la longueur du mot cible et le degré de parenté dans le modèle. Pour ce faire, 106 apprenants britanniques de français, dont les capacités vont du niveau débutant au niveau universitaire, ont été testés à l'aide d'un test orthographique Oui/Non à 100 items des 5000 mots les plus fréquents de la liste de vocabulaire français lemmatisé de Baudot (1992). Quatre séries de données ont été recueillies.

Les mots de test ont reçu un score pour leur fréquence. Cela a été fait de plusieurs manières. Une façon consistait à regrouper les mots de test dans leurs bandes de fréquences, et les mots ont ainsi obtenu le score de leur bande : 1 pour les mots de la bande de 1000 mots la plus fréquente, 2 pour les mots de la deuxième bande de 1000 mots, et ainsi de suite. Le test contenait 20 mots choisis au hasard dans chacun des cinq premiers niveaux de 1000 mots. Une deuxième façon de noter la fréquence était d'utiliser le nombre d'occurrences dans le corpus de Baudot. La troisième méthode consistait à utiliser les scores de fréquence du corpus de Baudot, mais à omettre les mots de niveau 0 du test, examinant ainsi l'effet de la fréquence lorsque seul le vocabulaire lexical est testé.

La longueur des mots a été mesurée en comptant le nombre de syllabes dans chaque mot de test. On suppose que plus un mot possède de syllabes, plus il est susceptible d'être difficile. Les mots de test en français variaient entre une et cinq syllabes et n'étaient pas également répartis entre les niveaux de fréquence. La longueur moyenne des syllabes pour les mots à chaque niveau de fréquence est indiquée dans le tableau 2-4.

Tableau 2-4: Scores moyens de longueur de syllabe et de parenté par niveau de fréquence

Niveau de fréquence	Nombre moyen de syllabes	Score moyen de parenté	Score moyen d'apprenabilité
1	1.9	0.4	23.70
2	2.4	0.6	10.85
3	2.6	0.6	11.80
4	2.4	0.5	9.40
5	2.6	0.7	8.65

Mesurer la parenté d'un mot est un peu plus compliqué. Lorsque cela est discuté, par des enseignants ou des universitaires, on a tendance à supposer qu'un mot est soit apparenté, soit qu'il ne l'est pas. Cependant, une division binaire ne fait pas très bien fonctionner ce type de modélisation et ne reflète pas le fait que les mots L2 peuvent varier dans leur degré de similitude avec les équivalents L1. Certains mots L2 ressemblent plus à la L1 que d'autres. Pour les apprenants anglophones du français, le mot *moment* est identique en orthographe et en apparence à un équivalent L1, mais le mot *heure* n'est pas identique, bien qu'il soit assez similaire. Dans le

but de mesurer ce degré de similitude, un calcul a été effectué en fonction du nombre de lettres dans la séquence correcte se produisant dans le mot cible, en proportion du mot L1 d'origine. *MOMENT* a six lettres sur six en commun entre l'anglais et le français et obtient un score de 1. En revanche, *heure* a trois lettres sur cinq en commun, HeURe, et un score de 0,6. On suppose que plus un mot a de lettres en commun, plus il sera apparenté et plus ce sera facile. Les mots de test en français variaient entre 0 et 1 selon leur degré de parenté et n'étaient pas également répartis entre les niveaux de fréquence. Le score moyen de parenté pour les mots à chaque niveau de fréquence est présenté dans le tableau 2.4.

À première vue, il semble que ces facteurs n'agissent peut-être pas entièrement indépendamment les uns des autres. Les mots de la bande la plus fréquente semblent en moyenne plus courts, ce qui devrait les rendre plus faciles, mais moins apparentés, ce qui devrait simultanément les rendre plus difficiles. Pour les mots de la bande la moins fréquente, l'inverse est vrai. Ces mots semblent avoir plus de syllabes, ce qui devrait les rendre plus difficiles, mais sont plus apparentés, ce qui devrait les rendre plus faciles. Cependant, les apparences peuvent être trompeuses et le calcul de la corrélation entre ces facteurs suggère qu'il y a peu ou pas de relation systématique, et aucune qui soit statistiquement significative. Les corrélations sont données dans le tableau 2-5.

Tableau 2-5: Corrélations entre le niveau de fréquence, le nombre de syllabes et la parenté

	Niveau	Appartenance
Syllabes	0.182	0.135
Appartenance	0.167	

Source (Milton, 2009, p. 40)

Milton (2009) ne s'attarde pas ensuite à « s'auto-critiquer » en qualifiant ce modèle d'apprenabilité de "très grossier". Même le classement des mots en fonction de leur fréquence dans un corpus doit souffrir d'une sorte d'effet plafond car, au-delà d'un certain niveau, si des mots doivent être appris, ils le seront et la répétition ultérieure deviendra redondante. Milton estime ensuite que le calcul qui omet le niveau 0 pourrait s'avérer plus utile que les autres méthodes d'évaluation de la fréquence. Les calculs de la longueur des mots et de la parenté

pourraient également être contestés et modifiés. Néanmoins, étant donné qu'il existe une relation si cohérente entre la fréquence et l'apprentissage des mots, il serait surprenant que la fréquence n'ait pas apporté une contribution quantifiable au résultat de l'apprentissage des mots et de ceux qui ne le sont pas. Milton conclut que les résultats de corrélation qui sont tous de l'ordre de 0,5 suggèrent que la relation est assez forte et qu'une analyse de régression devrait montrer le degré de variance des scores d'apprentissage que la fréquence peut expliquer. Mais, si les mesures de difficulté, telles que la parenté et les nombres de syllabes, ont également un impact sur la probabilité qu'un mot soit appris ou non, cela devrait également être visible dans une telle analyse. Si ces mesures supplémentaires s'avèrent fiables, elles affineront le système et contribueront en outre à expliquer la variation des scores d'apprentissage.

Les résultats d'une analyse de régression par étapes ont confirmé que la fréquence a un impact mesurable sur l'apprentissage du vocabulaire. Lorsque les mots sont placés dans des bandes de fréquences pour ne fournir qu'une estimation de la fréquence relative, la mesure de fréquence a pu expliquer environ 20 % de la variance du score d'apprentissage.

En utilisant les scores de fréquence réels, même cette prédictivité disparaît sous l'influence des mots de fonction très fréquents. En utilisant uniquement des éléments lexicaux et en excluant les mots du niveau 0 dans l'analyse, la prédictivité de la fréquence revient et explique environ 30% de la variance dans les scores d'apprentissage. Il s'agit toujours d'un score modéré, mais quelle que soit la manière dont la mesure de fréquence est modifiée, le chiffre d'environ 30% ne peut pas être amélioré avec cet ensemble d'apprenants. Ce résultat est probablement assez encourageant et pourrait cependant être amélioré dans un meilleur modèle capable de prendre en compte les différences réelles de fréquence entre les items du test. Il existe plusieurs manières évidentes de modifier et d'améliorer l'aspect fréquence du modèle, notamment en se référant aux matériels réellement utilisés par les apprenants. L'étude a inclus un grand nombre d'apprenants de français de niveau débutant, parmi lesquels il semble que plusieurs années d'études soient nécessaires pour que les effets de fréquence apparaissent.

L'analyse de régression raconte une histoire différente pour les facteurs de difficulté, la parenté et la longueur des mots. Dans aucun des cas, et dans aucune des analyses, ces facteurs n'ont pu être démontrés comme ayant un impact sur la capacité d'apprentissage. Les effets attendus pour les éléments de difficulté des mots n'ont pas émergé dans cette analyse et il pourrait y avoir plusieurs raisons pour lesquelles il en est ainsi. La première est que les systèmes utilisés pour mesurer les éléments de difficulté des mots sont en quelque sorte mal adaptés pour saisir

ces qualités de difficulté, et différentes formes de mesure peuvent donner des résultats différents. De plus, le choix de la langue à étudier peut affecter les résultats. Le français et l'anglais ont tant de mots apparentés, que l'effet pourrait être perdu dans ces circonstances. Les mots qui sont apparentés aux deux langues dans cette étude pourraient ne pas être apparentés dans d'autres et des résultats différents pourraient être produits si l'étude était reproduite avec, par exemple, des apprenants algériens, où l'on pourrait s'attendre à beaucoup moins de mots apparentés ou cognates.

Cependant, il faut également considérer la possibilité que les effets des facteurs de difficulté, tels que la parenté et la longueur des mots, ne soient pas aussi cohérents dans l'ensemble d'un lexique, ni aussi importants, que leur importance dans la littérature nous laisse supposer. Ces deux facteurs peuvent avoir été influencés par l'effet de la morphologie en anglais et en français. De nombreux affixes dans ces langues sont si similaires qu'il peut être possible de réduire un mot long inconnu à des composants plus courts connus ou devinables. C'est clairement un sujet à approfondir. Néanmoins, cette conclusion sert à indiquer l'importance de la fréquence d'occurrence en tant qu'influence sur les mots qu'un apprenant est susceptible de rencontrer et d'apprendre. Il sert à mettre en contexte l'effet des différents types de difficultés qui peuvent également influencer si un mot est susceptible d'être appris. Des facteurs tels que la longueur des mots, la partie du discours ou le caractère concret des mots et les particularités du manuel ne semblent pas inverser l'influence de la fréquence générale. Il ne s'agit pas de nier que des facteurs tels que la longueur et la parenté n'ont pas d'effet, mais il semble que cela ne ressemble en rien à l'effet de la fréquence. Ces facteurs de difficulté peuvent avoir un impact sur l'apprentissage de petits groupes ou de mots individuels, mais le type d'effets généraux créés par la fréquence semble être absent des facteurs de difficulté.

2.8 Conclusion

La conclusion que l'on peut tirer de ce chapitre, et le but de ce chapitre, est qu'il existe un système pour apprendre le vocabulaire, ou au moins reconnaître le vocabulaire d'une langue étrangère, et qu'il est déterminé par la fréquence des mots. Des facteurs tels que la difficulté des mots peuvent opérer au niveau des mots individuels, mais ne semblent pas avoir un impact comparable à la fréquence et ne semblent pas déstabiliser le modèle fréquentiel de l'apprentissage. Comme Palmer l'a souligné, les mots les plus fréquents dans la langue ont vraiment tendance à être appris le plus tôt. Cette conclusion permet de cibler les tests et la mesure

du vocabulaire. Nous pouvons utiliser les informations de fréquence pour étudier la connaissance d'une sélection minutieuse de mots que les apprenants sont susceptibles d'avoir rencontrés et ont eu l'occasion d'apprendre, sans craindre que de telles mesures manquent gravement l'essentiel en ne prenant pas en compte d'autres facteurs et en étudiant un grand nombre de vocabulaire peu fréquent. Les mesures utilisant uniquement des données basées sur la fréquence comme base d'une estimation de la taille totale du vocabulaire seront déstabilisées par un effet de plafond si les personnes testées sont très douées. Plus le niveau des apprenants est élevé, plus il sera probable d'altérer la qualité des mesures utilisant uniquement les mots les plus fréquents.

Le profil de fréquence montre que le vocabulaire acquis par les apprenants de L2 a tendance à se concentrer dans les bandes de vocabulaire les plus fréquentes. Cela facilite le processus de test de la connaissance du vocabulaire d'un apprenant, car un test peut se concentrer sur des mots dans ce domaine. Si un test répartit ses éléments de test au hasard sur toutes les dizaines ou centaines de milliers de formes de mots dans une langue, alors, là où les connaissances sont concentrées de cette manière, il y a un danger que les mots qu'un apprenant connaît soient manqués. Mais, en sachant où se concentre l'apprentissage, le test pourrait également concentrer les éléments du test dans ce domaine et formerait ainsi une bonne estimation des connaissances. Ce ne sera cependant qu'une estimation. C'est en partie parce que même avec cette connaissance, tous les mots des bandes les plus fréquentes ne peuvent pas être testés, seul un échantillon sera utilisé. Mais c'est aussi parce qu'il y aura toujours une connaissance des mots en dehors des bandes les plus fréquentes. En fonction de l'expérience individuelle de l'apprenant des thèmes abordés en classe, ou des intérêts individuels de l'apprenant, la connaissance des mots de contenu thématiquement liés sera dispersée parmi les bandes les moins fréquentes et un test de vocabulaire ne peut raisonnablement espérer les échantillonner de manière adéquate. Malgré cela, comme le suggère Aizawa (2006), un test basé sur les bandes les plus fréquentes devrait donner une bonne estimation de la connaissance du vocabulaire, ce qui devrait, à son tour, nous permettre d'examiner le développement du vocabulaire tout au long du processus d'apprentissage de la langue.

À d'autres fins, cela ne doit pas être un tel problème. Aizawa (cité dans Milton 2009) conclut que pour les tests de placement, l'utilisation de profils et de données de fréquence, comme dans, par exemple, X-Lex de Meara et Milton (2003), pourrait être limitée aux seuls niveaux les plus fréquents. Le fait que de telles mesures puissent ne pas donner un score absolu tout à fait exact pour la connaissance du vocabulaire ne les empêche pas de donner de bonnes indications sur les connaissances globales et des indications précises sur les niveaux comparatifs

de connaissances. Il semble inévitable qu'une mesure de la connaissance du vocabulaire qui puisse être comparée entre les apprenants, les systèmes éducatifs, les langues et les pays doit s'appuyer sur des informations de fréquence.

CHAPITRE 3 :

RELATION ENTRE FREQUENCE ET COUVERTURE

3.1 Introduction

En connaissant le vocabulaire le plus fréquent dans une langue, nous pouvons construire des tests de vocabulaire bien orientés qui peuvent mesurer les connaissances dans un domaine essentiel de la langue. Il a été souligné que ce n'est pas une idée nouvelle et que Harold Palmer a décrit cette relation entre fréquence et apprentissage. Mais l'observation de Palmer (Palmer, 1917) selon laquelle les mots les plus fréquents seront appris le plus tôt, va plus loin et suggère que les mots les plus fréquents sont aussi les plus utiles à l'apprenant ; ce sont les mots qui permettront à l'apprenant de se comprendre et de s'exprimer le plus efficacement.

« It is evident that it is more useful for a foreigner to be able to use words like go, is, here, and I than words like fidelity, quarry, or transit » (Palmer, 1917, p. 123)

Il y a donc une deuxième raison de vouloir tester et mesurer la connaissance des mots les plus fréquents dans une langue, c'est que les informations obtenues sont susceptibles de nous renseigner sur la capacité de l'apprenant à fonctionner, et à communiquer, dans la langue étrangère. Il est suggéré qu'un apprenant qui ne connaît que des mots très peu fréquents sera moins bien équipé pour fonctionner qu'un apprenant qui connaît les mots les plus fréquents.

L'hypothèse de Palmer selon laquelle les mots les plus fréquents seront appris le plus tôt s'est avérée globalement correcte. Sa seconde hypothèse, que les mots les plus fréquents sont les plus utiles, résiste-t-elle également à l'examen ?

3.2 La fréquence des mots et la couverture : la loi de Zipf

Il a déjà été noté au chapitre précédent que dans une langue, il y a toujours des mots très fréquents, des mots comme *le* et *et* en français ou comme *the* et *of* en anglais. À l'autre extrémité de l'échelle de fréquence, il y a beaucoup de mots, comme *apurer*, *œcuménique* et *tabouer*, qui semblent être très rares et n'apparaissent que quelques fois dans un corpus de langage normal.

Entre les deux, il y a un nombre moyen de mots avec des scores de fréquence intermédiaires quelque part entre les deux extrêmes. Ce type de distribution est connu sous le nom de distribution *Zipf* et a donné naissance à la loi de *Zipf*. La loi de *Zipf* permet de décrire plus systématiquement et de représenter graphiquement la relation entre le rang d'un mot dans une liste de fréquences et le nombre de fois où il apparaît. La loi de *Zipf* stipule que dans un corpus de langage naturel, la fréquence d'un mot est à peu près inversement proportionnelle à son rang dans le tableau des fréquences. Ainsi, le mot classé premier dans le tableau est susceptible d'apparaître environ deux fois plus souvent que le mot classé deuxième, ce qui est susceptible d'être deux fois plus fréquent que le mot classé quatrième et ainsi de suite. Pour aider à illustrer cela, le tableau 3-1 montre le rang et la fréquence des huit mots les plus fréquents dans les corpus anglais et français.

Tableau 3-1 : Rangs et fréquences des mots en anglais et en français (Milton, 2009)

	Anglais (Kilgariff, 2006)		Français (Baudot, 1992)	
1	the	6,187,267	de	68,373
2	be	4,239,632	le	42,419
3	of	3,039,444	être	26,897
4	and	2,687,862	un	26,613
5	a	2,186,369	avoir	23,570
6	in	1,924,315	à	23,475
7	to	1,620,850	et	23,325
8	have	1,375,636	les	19,230

Ce n'est clairement pas une loi parfaite, ni une relation parfaite. Dans ces deux exemples, le mot du deuxième rang peut être sensiblement moins fréquent que le mot du premier rang, mais il est supérieur aux 50 % suggérés par la loi de Zipf. De même, dans les deux exemples, le quatrième mot classé a une fréquence bien supérieure à 50 % de la seconde. Dans le corpus anglais, la fréquence du huitième mot classé, *have*, est près de la moitié de celle du quatrième, *and*, mais dans le corpus français cette régularité fait encore défaut.

Dans d'autres corpus, cette relation peut être plutôt nette. Dans le Brown Corpus (Kučera & Francis, 1967), par exemple, le mot anglais le plus fréquent, *the*, représente près de 7 % de

l'ensemble du corpus (69 771 occurrences sur plus d'un million de mots). Le deuxième mot classé, *of*, a 36 411 occurrences, soit près de 3,5 %. (Milton, 2009)

Webb et Macalister (cité dans Nation, 2013) ont constaté que 42% des familles de mots à moyenne et basse fréquence dans les journaux scolaires de Nouvelle-Zélande (écrits pour les enfants) ne se présentaient qu'une seule fois, et 47% des mots à moyenne et basse fréquence dans une collection de textes de journaux et de fiction ne s'est produite qu'une seule fois. Si nous nous intéressons aux mots avec une fréquence de 2, la formule nous dit qu'un sixième se produira deux fois. La loi de *Zipf* décrit une distribution qui ne se limite pas au vocabulaire, mais s'applique à de nombreuses occurrences naturelles, comme la répartition de la richesse et les effets des répétitions.

La loi de *Zipf* n'est pas une description parfaite du langage, c'est donc une loi empirique et non théorique. Néanmoins, la relation entre la fréquence d'un mot et sa position dans le tableau des fréquences est assez claire et a une implication importante. Cela signifie qu'un petit nombre de mots ont tendance à constituer une très grande partie d'un texte normal. Dans le *Brown Corpus*, quelque 135 mots représentent la moitié de l'ensemble du corpus. Une liste plus complète des bandes de fréquences et de la couverture qu'elles sont susceptibles de fournir, en anglais, est présentée dans le tableau 3.2, qui montre également que les mots les plus fréquents contribuent très fortement à la couverture du texte et que moins un mot est fréquent, moins il y contribue.

Tableau 3-2 : Chiffres de couverture types pour différentes bandes de fréquences

Nombre de mots	Couverture de texte (%)
10	24
100	49
1,000	74
2,000	81
3,000	85
4,000	88
5,000	89
6,000	95
7,000	99
8,000	100

Source Carroll et al. (1971, cité dans Nation, 2001)

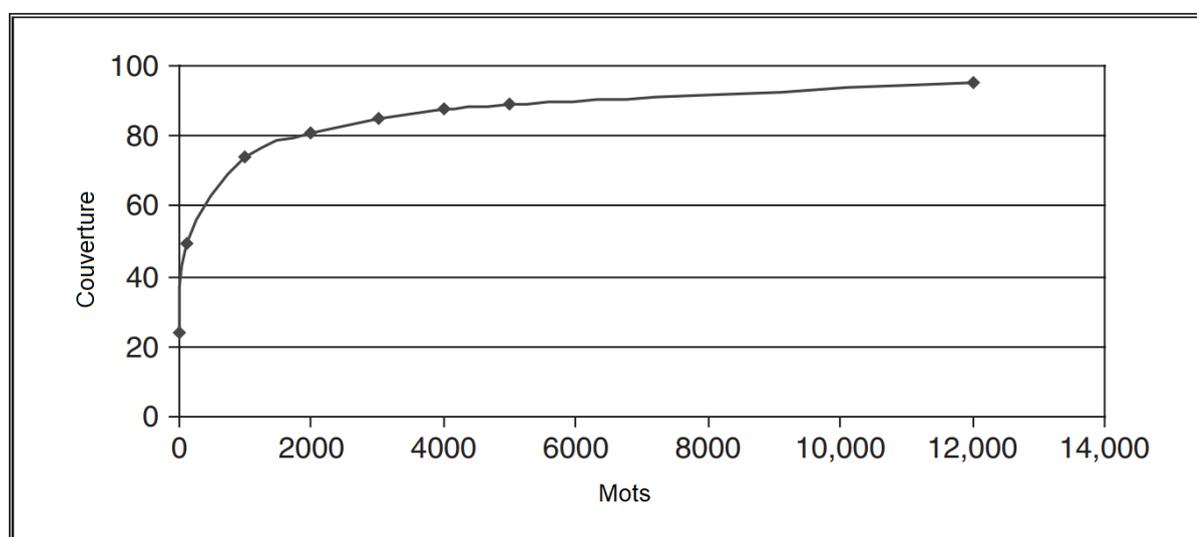


Fig. 3-1: Les bandes les plus fréquentes du tableau 3.2 présentées sous forme de graphique

Dans la figure 3-1, la courbe monte fortement sur le côté gauche, et dans cette zone, chaque mot supplémentaire contribue de manière significative à la couverture du texte. La connaissance d'environ 1000 mots en anglais devrait signifier qu'un apprenant reconnaîtrait et comprendrait environ les trois quarts des mots du texte normal. D'après Milton (2009), une connaissance

d'environ 2000 mots en anglais devrait signifier qu'environ 80% des mots du texte normal seraient compris. Ainsi, si les apprenants apprenaient ces mots, ils connaîtraient une grande partie des textes qu'ils lisent ou entendent et auraient par conséquent de meilleures chances de les comprendre. La logique derrière l'affirmation de Palmer selon laquelle les mots les plus fréquents sont aussi les plus utiles est très claire. Nation suggère que ces 2000 mots les plus fréquents sont si importants pour l'apprentissage et la capacité linguistiques que presque tout ce qui peut être fait pour s'assurer qu'ils sont appris vaut la peine d'être fait (Nation, 2013, p. 16). Cela pourrait impliquer qu'ils devraient être enseignés explicitement, malgré l'énorme engagement en temps de classe que cela impliquerait. Après environ 2000 mots, la courbe s'aplatit considérablement et chaque mot supplémentaire contribue beaucoup moins à la couverture globale. Si vous connaissez déjà les 2000 mots les plus fréquents en anglais, alors pour l'effort d'apprentissage des 2000 mots les plus fréquents suivants, seulement 7 ou 8% de couverture supplémentaires pourraient être obtenus. En français, certains parlent de 600 mots qui couvriraient 90 % de n'importe quel texte comme cité dans le site <http://www.encyclopedie-incomplete.com/?Les-600-Mots-Francais-Les-Plus>.

Malgré ces réserves, il y a plus qu'une part de vérité à l'idée que la couverture est importante pour la compréhension et que la connaissance des mots les plus fréquents, qui contribuent tant à la couverture, est également importante. Milton donne l'exemple d'un discours bien connu qui simule à quel point il serait compréhensible pour un apprenant qui ne connaît que les 10 mots d'anglais les plus fréquents. Ces mots ont été laissés en place dans le texte et tous les autres mots ont été remplacés par une série de XXXX.

XXXX XXXX and XXXX XXXX XXXX, XXXX XXXX XXXX XXXX on XXXX XXXX a XXXX XXXX, in XXXX, and XXXX to the XXXX XXXX XXXX XXXX XXXX XXXX XXXX.

Même si ces 10 mots donnent environ 20 à 25% de couverture dans un texte normal, ce qui semble beaucoup, il est impossible de tirer quoi que ce soit de ce texte. Le discours d'où vient cette phrase est méconnaissable. Le deuxième exemple porte les connaissances du lecteur entre 100 et 150 mots, ce qui est suffisant pour couvrir environ 50 %. Les mots supplémentaires dans cette plage ont été insérés dans le texte et le reste laissé sous forme de XXXX.

Four XXXX and seven years XXXX, our fathers XXXX XXXX on this XXXX a XXXX XXXX in XXXX, and XXXX to the XXXX that all men are XXXX XXXX..

Encore une fois, une couverture de 50% semble être assez bonne, mais il est toujours impossible de gagner beaucoup en compréhension. La plupart des gens ne peuvent pas suggérer quels pourraient être les mots manquants, à moins qu'ils ne reconnaissent le discours, car il n'y a pas assez d'informations pour donner un indice sur le sujet du matériel. Même les lecteurs qui connaissent ce discours, dans de nombreux cas, ne le reconnaissent pas à ce niveau de couverture. Milton (2009, p. 48) cité l'exemple suivant qui élève les connaissances du lecteur à environ 2000 mots, ce qui est suffisant pour une couverture d'environ 80%, et ces mots supplémentaires ont, encore une fois, été insérés aux endroits appropriés.

Four XXXX and seven years ago, our fathers brought forth on this continent, a nation, XXXX in XXXX and XXXX to the XXXX that all men are created equal..

À ce stade, la plupart des gens, s'ils connaissent le discours de Gettysburg⁴, le reconnaîtront comme la première phrase de ce discours. Le texte est évidemment beaucoup plus compréhensible à ce niveau de couverture, et les lecteurs qui ne connaissent pas le discours auront désormais une bonne idée de ce dont il s'agit. Mais la compréhension est loin d'être parfaite et la série de quatre mots inconnus très proches dans la seconde moitié du passage est une véritable barrière à la prise de sens de l'ensemble. Il est toujours étonnamment difficile, même avec autant d'informations contextuelles, de deviner les mots manquants si vous ne les connaissez pas déjà. Le dernier exemple porte les connaissances du lecteur à environ 6 000 mots, suffisamment pour une couverture de plus de 90 %, en anglais et réinsère les mots supplémentaires dans cette plage.

Four score and seven years ago, our fathers brought forth on this continent, a nation, XXXX in liberty and dedicated to the XXXX that all men are created equal.

À ce niveau de couverture, la plupart des lecteurs pensent qu'ils peuvent comprendre à peu près tout. La couverture à 95% ne laisse que quelques mots inconnus dans un passage de cette longueur, et la plupart des lecteurs ont la possibilité de les ignorer et de prendre le sens général de l'article sans avoir besoin de reconnaître ou de deviner chaque mot. Si le passage devait être étudié de manière intensive, pour un examen par exemple, alors le lecteur pourrait

⁴ Le discours de Gettysburg est un discours prononcé par le président américain Abraham Lincoln pendant la guerre de Sécession lors de l'inauguration du cimetière national des soldats à Gettysburg, en Pennsylvanie, dans l'après-midi du 19 novembre 1863, quatre mois et demi après la défaite des armées de l'Union. ceux de la Confédération à la bataille de Gettysburg. C'est l'un des discours les plus connus de l'histoire américaine.

s'inquiéter des mots inconnus, mais pour une lecture générale, ce niveau de connaissance semble permettre une assez bonne compréhension. Il est suggéré à partir de ces informations que de grandes quantités de vocabulaire et une couverture presque complète sont nécessaires pour tout ce qui approche la compréhension et l'utilisation normales de la langue.

Cela vaut la peine de considérer ce qui se passe si seuls les mots les moins fréquents sont connus. On pourrait faire valoir que si ces mots de contenu peu fréquents sont si importants, alors peut-être que les apprenants seraient mieux employés à les apprendre. Un apprenant peut-il fonctionner dans une langue étrangère sans les mots de fonction et de structure les plus fréquents ? Pour répondre à cette question, Milton a supprimé, dans l'exemple final, les 150 mots environ les plus fréquents en anglais et remplacé les XXXX comme précédemment.

*XXXX score XXXX XXXX XXXX ago, XXXX XXXX brought forth XXXX XXXX continent
XXXX nation conceived XXXX liberty, XXXX dedicated XXXX XXXX proposition XXXX XXXX
XXXX XXXX created equal*

Milton constate qu'il est possible d'avoir une idée de ce dont parle le discours avec seulement cette information mais, en raison de l'information manquante, la compréhension est fracturée et le détail du message est difficile à saisir.

Connaître seulement quelques mots d'une langue étrangère n'aide pas un apprenant à achever sa compréhension même si les mots connus sont les éléments lexicaux qui ont le plus de poids sémantique (Milton, 2009). Quels que soient les mots que vous choisissiez, presque tous les mots d'un texte doivent être connus avant que le message complet puisse être transmis au lecteur ou à l'auditeur.

Cela remet en question de nombreuses hypothèses sur le nombre de mots nécessaires pour exceller dans une langue. Ils remettent en question le genre d'hypothèses attribuées à Ogden et aux autres champions de l'anglais de base selon lesquelles un vocabulaire aussi petit que 850 mots est suffisant pour comprendre et pour sonner normalement dans la parole. La plupart des lecteurs seraient poussés jusqu'au point de rupture s'ils devaient fonctionner normalement dans leur vie professionnelle avec le type de couverture que fournit moins de 1000 mots. Cela suggère également que les tests de vocabulaire qui se concentrent sur les mots les plus fréquents de l'anglais peuvent, en effet, fournir des informations utiles sur la façon dont un apprenant peut fonctionner dans la langue étrangère.

3.3 La couverture requise pour la compréhension

S'il semble qu'il y ait du mérite à l'idée que plus la proportion d'un texte connu est grande, meilleure sera la compréhension, alors cela soulève la question de savoir exactement quelle couverture est nécessaire avant que la compréhension ne soit complète. C'est un sujet qui a reçu une certaine attention de la part des chercheurs, et en particulier de Laufer. Ces chercheurs se sont préoccupés d'établir le type de proportions d'un texte, et les ressources de vocabulaire, nécessaires pour gérer la lecture nécessaire dans l'étude académique dans une langue étrangère.

Laufer et Sim (1985) ont étudié les connaissances nécessaires pour bien comprendre un texte d'anglais à des fins académiques lors de l'examen du *Cambridge First Certificate in English*. L'étude est parvenue à la conclusion que la connaissance du vocabulaire est le domaine de connaissance le plus important requis pour la compréhension ; plus important que la connaissance du sujet et plus important que la connaissance syntaxique. Laufer (1989) est allé plus loin et a calculé que les élèves devaient connaître et comprendre au moins 95 % des mots courants dans un texte afin de s'assurer une compréhension « raisonnable ». La compréhension raisonnable a été évaluée par la capacité d'obtenir au moins 55 % à un examen de compréhension en lecture. Bien qu'il ne soit pas tout à fait clair ce que cela signifie en termes de compréhension, cela suggère qu'il existe un chiffre pour la connaissance du vocabulaire nécessaire pour répondre aux exigences des études universitaires par le biais de l'anglais comme langue étrangère (EFL). Hu et Nation (2000) ont rapporté des niveaux de couverture encore plus exigeants requis pour d'autres types de compréhension et suggèrent qu'une couverture de 98% serait un seuil à partir duquel les apprenants pourraient comprendre suffisamment un texte pour pouvoir le lire pour le plaisir. Il n'est pas nécessaire qu'il y ait contradiction entre ces deux chiffres. La lecture pour l'étude et la lecture pour le plaisir peuvent simplement nécessiter des niveaux de connaissances différents. La lecture pour l'étude n'a pas besoin d'être inhibée par une pause une fois tous les 20 mots environ pour vérifier un sens, tandis que la lecture pour le plaisir suggère que le lecteur n'a pratiquement pas besoin de faire une pause (Carver, 1994).

Un seuil de 95 ou 98 % peut impliquer qu'il y a un tout ou rien pour ce type de connaissances sans lesquelles les apprenants peuvent tirer peu d'un texte et échoueraient aux examens conçus pour évaluer la compréhension. Cependant, les résultats de Laufer (1989) indiquent qu'il existe une certaine flexibilité autour de ce chiffre. Certains apprenants avaient moins de 95 % de

connaissances et obtenaient tout de même un score de compréhension en lecture, ce qui suggérerait qu'ils pouvaient comprendre les textes utilisés. Il s'agit peut-être d'un artéfact des tests qui n'ont probablement pas mesuré les mots que les élèves connaissaient ou leur compréhension avec une précision absolue. Milton reconnaît qu'aucun test ne peut être parfait. Mais peut-être aussi, cela peut simplement refléter des différences dans la façon dont les individus peuvent déduire le sens d'une connaissance partielle d'un texte. Certains lecteurs peuvent être particulièrement doués dans ce domaine et peuvent se débrouiller avec des vocabulaires légèrement plus petits que ceux requis par d'autres lecteurs. Nation (2013, p. 147) suggère que ces seuils à 95 et 98 % sont donc probabilistes et ne sont pas des niveaux de connaissances tout ou rien requis pour la compréhension. La compréhension, comme le souligne Ward (1999, cité dans Milton, 2009), dépendra probablement autant de l'importance, de la position exacte, de la possibilité de deviner et de la réapparition des mots inconnus que du pourcentage de mots connus.

Nation (2013, p. 146) suggère que l'analyse des corpus pourrait bien produire de meilleurs résultats et plus d'informations. Les graphiques qui tracent la fréquence par rapport à la couverture montrent leur valeur ici et permettent d'estimer facilement la couverture produite par différents volumes de vocabulaire. Dans la figure 3-2, la couverture est établie et les chiffres du corpus de Carroll et al. (1971) suggèrent combien de mots les plus fréquents de l'anglais qu'un apprenant devrait connaître avant de pouvoir s'attendre à une couverture de 95 % du texte normal.

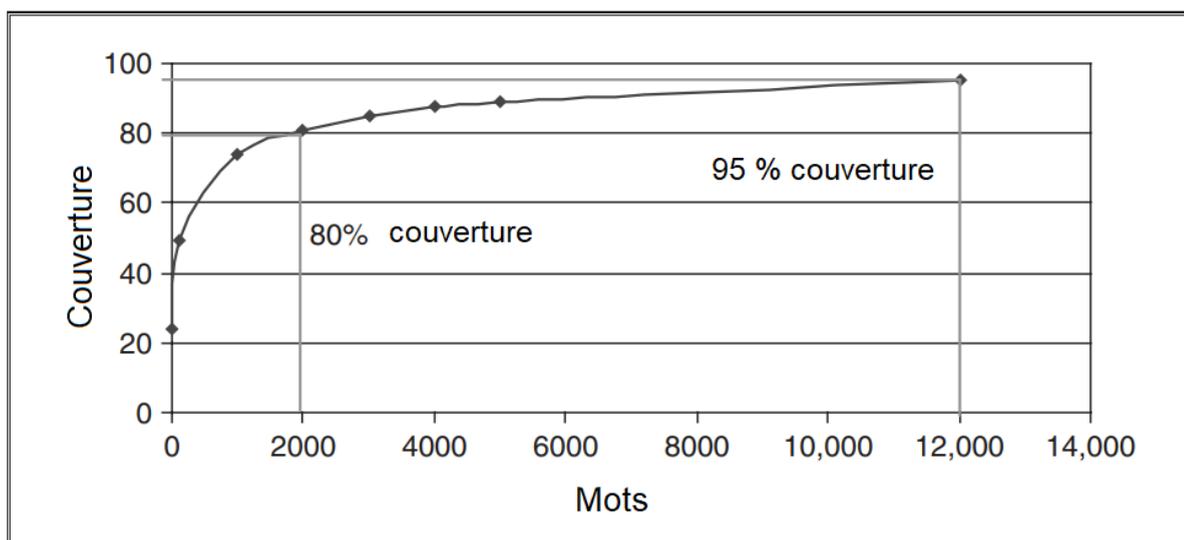


Fig. 3-2 : Connaissance des mots impliquée par les corpus pour 80 et 95% de couverture du texte Carroll et al. (1971)

Le chiffre suggéré par ce corpus est d'environ 12 000 mots lemmatisés, nettement plus que ce que Laufer a estimé. Mais, comme le souligne Nation, l'analyse du corpus suggère une explication à cela, à savoir que le nombre de mots nécessaires à la couverture et à la compréhension peut varier en fonction d'un certain nombre de facteurs. L'un est le type de texte. Les articles de journaux, par exemple, différeront par les mots qu'ils utilisent, des essais universitaires et des lettres informelles écrites à la maison. Certains types de texte sont susceptibles de nécessiter un vocabulaire plus important pour la compréhension que d'autres. Un deuxième facteur est la longueur du texte lu. Une brève note pour le laitier aura non seulement moins de mots, mais moins de mots différents qu'une bibliothèque de textes sur le marketing d'entreprise. Enfin, il y a l'homogénéité du texte. Un recueil d'écrits sur différents sujets et par plusieurs auteurs différents peut très bien utiliser une plus grande variété de mots que les écrits sur un seul sujet par un seul auteur, même si la longueur totale du recueil est la même. Des corpus très volumineux, tirés de sources disparates telles que celle utilisée dans la figure 3-2, peuvent donc suggérer qu'un vocabulaire beaucoup plus important est nécessaire que ce ne serait le cas pour un domaine plus spécialisé conclut Milton (2009).

Il peut, bien sûr, y avoir une différence entre ce que les enseignants et les universitaires considèrent comme une compréhension raisonnable et la façon dont les apprenants eux-mêmes perçoivent leurs capacités dans une langue étrangère. Seaton (cité dans Milton, 2009) a collecté des données auprès de ses étudiants en anglais académique, où il a testé leur connaissance des

5000 mots les plus fréquents en anglais (à l'aide de Meara & Milton, 2003) et leur a également demandé d'évaluer sur une échelle de 1 à 5 dans quelle mesure ils ont estimé que leur compréhension était, séparément, en lecture et en communication orale. Un score de 1 signifiait qu'ils n'avaient rien compris et 5 indiquaient qu'ils pensaient avoir tout compris. Milton (2009) a résumé les résultats qu'a obtenus Seaton dans la figure 3-3.

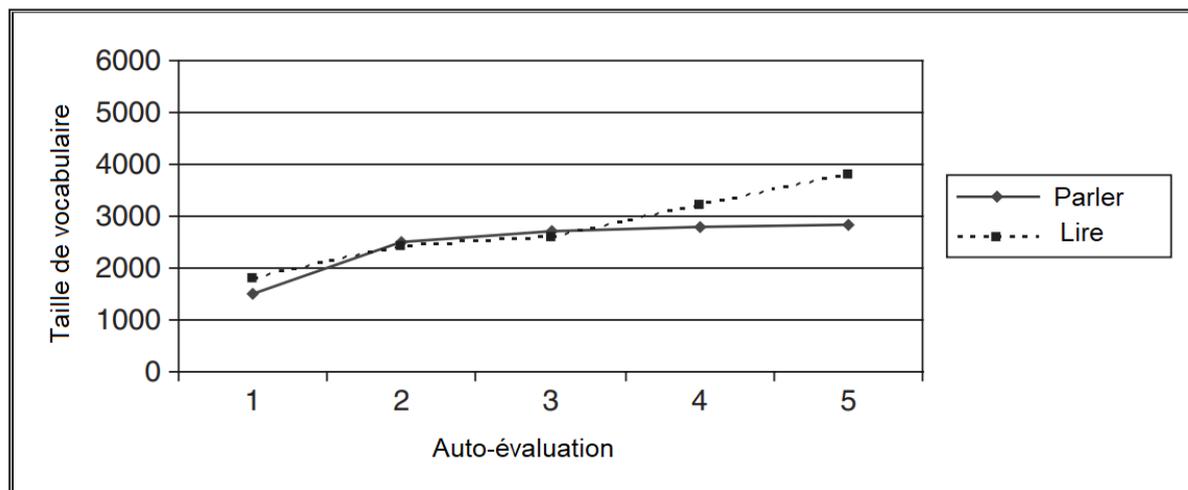


Fig. 3-3 : Auto-déclarations des élèves sur la compréhension et la connaissance du vocabulaire (Milton, 2009, p. 53)

Pour Milton (2009), cela suggère que plus les apprenants connaissent de vocabulaire, plus ils pensent comprendre à la fois en lecture et en parole. Néanmoins, Seaton a pu tirer des données qu'il avait collectées, un certain nombre d'observations perspicaces. La première est que les apprenants qui ont marqué moins de 1000 mots dans le test de taille de vocabulaire se sont toujours déclarés au niveau 1 en compréhension, et les apprenants notant sous 2000, ont fait principalement la même chose. Seaton suggère que ces apprenants évaluent leurs capacités sur la base de leur capacité à performer en classe, où les textes et la langue ont tendance à être notés et ont peu d'interaction avec les locuteurs natifs en dehors de la classe. Plus généralement, les apprenants à ce niveau de connaissance du vocabulaire ont tendance à déclarer qu'ils ne comprennent que peu.

Un deuxième point qui ressort de ces données se situe à l'autre extrémité de l'échelle de compréhension, chez ceux qui déclarent comprendre tout ce qu'ils lisent et entendent. Ici, il y a un écart évident entre les tailles moyennes de vocabulaire en lecture et en écoute. Les apprenants semblent avoir besoin d'un vocabulaire plus large entre 3 500 et 4 000 sur ce test de 5 000 mots avant de déclarer systématiquement une compréhension complète des documents

qu'ils lisent. Ce chiffre correspond probablement à l'estimation de Laufer d'environ 5000 mots lemmatisés au total requis pour la compréhension. Les apprenants à ce niveau de capacité auront une certaine connaissance du vocabulaire en dehors des 5000 mots les plus fréquemment testés dans l'expérience de Seaton. Les apprenants dont le vocabulaire est un peu plus petit sont capables de déclarer une compréhension totale. Une connaissance de 2500 3000 mots lemmatisés est systématiquement rapportée par les apprenants comme suffisante dans ce domaine de la langue. Milton (2009) pense qu'il est possible que cela reflète les différences de couverture que décrivent les corpus tirés de sources écrites uniquement et de sources orales uniquement. Dans la langue parlée, les mots les plus fréquents sont généralement considérés comme encore plus fréquents qu'ils ne le sont dans la langue écrite.

3.4 Couverture dans les corpus écrits et parlés

La différence dans la connaissance du vocabulaire des apprenants déclarant une compréhension orale et écrite complète peut être expliquée par la variation des fréquences des mots en anglais oral et écrit. Une grande partie du discours conversationnel a tendance à être moins formel, moins académique que l'écriture, de sorte que les mots que nous utilisons varieront en conséquence. La conversation, par exemple, les salutations que nous échangeons tous les jours avec des amis et des collègues, peuvent être plus stéréotypées et répétitives que tout ce que nous pourrions mettre par écrit. La conversation orale, en particulier, peut également être fortement sollicitée pour la communication sur les gestes, les expressions faciales et d'autres indices non linguistiques qui devraient réduire le fardeau de la connaissance du vocabulaire nécessaire. Ceux-ci sont susceptibles d'affecter les distributions de fréquence des mots dans le langage écrit et conversationnel, ce qui renforce l'idée que les mots nécessaires à la compréhension varieront en fonction de ce que les apprenants veulent faire du langage. Les apprenants seront-ils obligés de lire et d'écrire ou de parler abondamment dans la langue étrangère ?

Jusqu'à présent, la discussion sur la couverture s'est largement intéressée à la capacité de comprendre un texte écrit et les listes de fréquences utilisées ont été dérivées de corpus qui dépendent fortement de matériel écrit assez formel. Ce n'est peut-être pas surprenant lorsque la plupart des apprenants que nous étudions apprennent dans un environnement académique formel qui met fortement l'accent sur la capacité d'interagir et d'apprendre à partir de la langue écrite. Les différences peuvent paraître très grandes en effet et une seule étude, maintenant

assez ancienne, a beaucoup fait pour favoriser cette idée. Une étude de Schonell et al. (cité dans Milton, 2009) qui a été réalisée dans un milieu d'ouvriers australiens qualifiés et non qualifié a suggéré qu'il est possible d'obtenir des niveaux significatifs de couverture en anglais parlé avec remarquablement peu de mots. Quelque 200 mots fourniraient une couverture d'environ 80% et environ 800 mots environ fourniraient une couverture de 95%, ce qui pourrait impliquer une assez bonne compréhension sur la base des chiffres indiqués ci-dessus.

Adolphs et Schmitt (2003) ont tenté de mettre à jour cette étude et de la mettre en conformité avec la pratique moderne, afin que les résultats soient plus comparables avec les informations tirées d'autres corpus, et de s'appuyer sur un corpus plus large. Une liste de fréquences et des chiffres pour la couverture ont été dérivés en utilisant un nombre de mots basé sur des familles de mots, et non sur des lemmes, afin d'être aussi similaire que possible à ceux de Schonell et al. Les résultats qu'ils ont obtenus suggèrent qu'il faut beaucoup plus de vocabulaire pour la communication orale qu'on ne le suppose souvent, mais suggèrent tout de même que des niveaux de couverture plus élevés peuvent être obtenus avec des tailles de vocabulaire plus petites qu'il n'est possible avec la langue écrite. Quelques centaines de familles de mots suffisent pour fournir une couverture d'environ 80 %, mais plus de 2 000 familles de mots sont nécessaires pour une couverture de 95 % ou plus, ce qui pourrait donner une compréhension raisonnable d'une conversation normale et non technique. Cette conclusion, 2 000 familles de mots nécessaires pour les niveaux de connaissance communicatifs, semble bien correspondre aux chiffres de 3 000 à 3 500 lemmes que les apprenants ont rapportés pour leur propre compréhension dans l'étude de Seaton.

La conversation générale n'est pas nécessairement typique de chaque partie du discours parlé. Comme le soulignent Adolphs et Schmitt (2003, p. 433), si un locuteur non natif écoutait un cours universitaire sur un sujet assez technique sans support visuel et sans la possibilité d'interrompre la suite du discours pour poser des questions et confirmer sa compréhension, alors il y aurait une charge beaucoup plus lourde placée sur la connaissance du vocabulaire que dans une conversation générale. Cependant, la plupart des grands corpus modernes contiennent généralement des sous-sections importantes de texte transcrit à partir d'une variété de sources orales. Les informations de fréquence fournies par ces sous-corpus peuvent différer les uns des autres et du corpus écrit. Ceci est perceptible même dans quelques-uns des mots non lemmatisés les plus fréquents.

3.5 La conclusion à en tirer n'est pas seulement que la compréhension

La compréhension de supports oraux peut nécessiter moins de ressources de vocabulaire que la compréhension de textes écrits. C'est aussi que la capacité de bien comprendre aussi bien le texte écrit que parlé, si cela est à peu près normal, nécessite une connaissance très importante du vocabulaire. Connaître moins de 1000 mots d'une langue étrangère est probablement insuffisant pour la compréhension, même dans la langue parlée, à moins que la communication ne soit du type le plus stéréotypé, comme les salutations, et dépourvue de tout contenu spécifique ou spécialisé. Une fois qu'un apprenant se lance dans une conférence ou dans tout type d'engagement détaillé, l'apprenant a besoin d'un vocabulaire plus spécialisé et moins fréquent. La plupart des apprenants dans des contextes éducatifs formels auront besoin de la langue écrite et orale en classe et pour les examens, et ils auront besoin de ressources de vocabulaire substantielles pour faire face dans la plupart des cas. Cependant, certains apprenants, les personnes qui suivent des cours du soir avant de partir en vacances par exemple, peuvent souhaiter quelque chose de tout à fait différent, où quelques phrases parlées pour être poli et commander un repas seraient suffisantes. Beaucoup dépendra des types de communication dans lesquels un apprenant devra s'engager et du degré auquel le vocabulaire dont il aura besoin peut être prédit avec un certain degré de certitude. Le sujet des lexiques spécialisés, anticipant les mots nécessaires à une activité particulière plutôt que d'obliger les apprenants à acquérir de grandes quantités de vocabulaire général dont ils n'auront jamais besoin, a récemment attiré l'attention et c'est le sujet du chapitre suivant.

3.6 Couverture dans les lexiques spécialisés

Les chiffres de couverture et de compréhension discutés jusqu'à présent reposent sur des corpus linguistiques généraux et supposent, par conséquent, que les apprenants devront être capables d'utiliser leur langue dans une grande variété de domaines et de registres pour parler couramment. C'est en partie cette étendue de matériel dans les corpus qui donne lieu aux grands volumes de vocabulaire requis pour quelque chose comme une couverture complète. Apprendre ces grandes quantités de mots prend du temps et, pour l'apprenant, une tâche intimidante. Cependant, de nombreux utilisateurs de langues n'ont peut-être pas besoin de fonctionner dans une variété de paramètres. Le voyageur en vacances en Méditerranée n'aura probablement pas besoin des noms d'oiseaux et d'arbres en langue étrangère, des noms de produits chimiques ou du vocabulaire académique nécessaire pour rédiger un essai dans une université, afin de

fonctionner et de survivre. Si ces choses ne sont pas nécessaires, alors il ne devrait pas être nécessaire de les apprendre. Une grande partie de la valeur des lexiques spécialisés et techniques réside dans la perspective d'identifier des mots spécifiques que les apprenants doivent utiliser dans divers types d'activités afin de maintenir une couverture et une compréhension élevées de ces documents spécialisés dans ces domaines spécialisés, mais avec un vocabulaire global plus restreint. Il y a de nombreux avantages à réduire la charge d'étude de cette manière ; c'est moins intimidant et chronophage pour les apprenants et moins coûteux pour leurs parents ou parrains.

L'une de ces listes spécialisées qui est largement utilisée est la liste de mots académiques (Coxhead, 2000) qui est destinée à être utilisée par les étudiants universitaires qui étudient en anglais. Coxhead a dérivé cette liste de 570 mots-clés à partir d'un corpus de 3,5 millions de mots de matériel académique tirés d'une variété de disciplines académiques dans les arts, le droit, le commerce et les sciences naturelles. Ces quatre grands domaines étaient composés de 28 disciplines spécifiques, telles que l'histoire, l'économie, la géographie et la biologie. Les mots-clés incluent toutes les formes dérivées et fléchies, et constituent donc une analyse basée sur la famille de mots plutôt que sur le lemme, et comprennent environ 3 100 formes de mots au total. Pour être inclus dans la liste de mots académiques, ces mots devaient remplir un certain nombre de critères :

- Ils devaient être en dehors des 2000 mots les plus fréquents de l'anglais tels que définis par *West's General Service Word List* (West, 1953).
- Ils devaient se produire au moins 100 fois dans l'ensemble du corpus de 3,5 millions de mots (le critère de fréquence).
- Ils devaient se produire dans au moins 14 des 28 disciplines spécifiques (le critère de portée).
- Ils devaient, en outre, se produire au moins 10 fois dans chacun des quatre grands domaines.

La liste de mots académiques, lorsqu'elle est combinée avec les 2000 mots les plus fréquents, s'est avérée avoir une couverture étonnamment bonne d'une grande variété de textes académiques, équivalent à un nombre beaucoup plus important de mots tirés d'un corpus écrit plus général. Le tableau 3-3 donne les chiffres fournis par Coxhead (2000, p. 225) pour la couverture fournie par ces deux listes dans son corpus (*West's General Service Word List* est

divisée en deux ; les 1000 premiers mots les plus fréquents et le second 1000 mots les plus fréquents).

Tableau 3-3 : Couverture des textes académiques fournis par le GSL et l'AWL

GSL 1 K	71.4 %
GSL 2 K	4.7 %
AWL	10.0 %
Total	86.1 %

Selon le texte académique, les deux listes de mots peuvent donner une couverture d'environ 85,5 91,4 % (Nation, 2004, p. 8), un chiffre qui nécessiterait 4 000 à 5 000 familles de mots d'après les chiffres de couverture de Carroll et al. (1971) dans un corpus. Bien que cela fournirait à un apprenant les connaissances nécessaires pour lire et reconnaître la plupart des mots d'un texte, il est encore loin des chiffres de couverture de 95 et 98 % qui seraient nécessaires pour une compréhension satisfaisante. Il peut y avoir deux raisons à cela. La première est que ce genre particulier, le texte académique, peut être particulièrement exigeant en termes de connaissance du vocabulaire et qu'il serait impossible d'expliquer ou de comprendre la profondeur et la complexité de nombreuses idées académiques sans avoir accès à un très grand nombre de mots pour exprimer ces idées. La deuxième raison peut être que la liste de mots académiques s'inspire d'une grande variété de textes et que des disciplines académiques spécifiques peuvent nécessiter leur propre vocabulaire spécialisé. Les étudiants d'une spécialité particulière auraient besoin de ce vocabulaire en plus des listes de mots académiques et générales afin de comprendre les textes spécialisés. Les compilateurs de dictionnaires et les rédacteurs de manuels EFL créent souvent de telles listes de mots ou fournissent des glossaires, et un exemple typique pourrait être *l'Oxford Business Dictionary of English*, qui identifie une liste de base de 1000 éléments qui sont prétendument « particulièrement importants en anglais des affaires » (Parkinson, 2005, p. Préface).

Konstantakis (cité dans Milton, 2009) tente de construire exactement une telle liste de mots spécialisés ou techniques dans le domaine des études commerciales afin non seulement de fournir une liste potentiellement utile, mais aussi de tester la taille d'une telle liste avant qu'elle puisse atteindre le seuil de 95% de couverture. En utilisant une combinaison de critères de gamme et de fréquence et une méthodologie similaire à celle de Coxhead, Konstantakis dérive 498 mots avec une couverture inhabituellement large dans les manuels d'anglais des affaires, puis dérive des listes de familles de mots pour chacun d'eux afin de créer une liste de mots d'affaires. Il est démontré que les mots font partie d'un domaine spécialisé en comparant leur couverture dans un corpus d'un million de mots tirés de textes académiques commerciaux, avec une couverture dans des textes littéraires sans rapport avec les affaires. Dans les textes non commerciaux, la couverture fournie par cette liste est beaucoup plus faible. Le chiffre de la couverture dans ces différents domaines est fourni dans le tableau 3-4.

Tableau 3-4 : Couverture dans *business academic* et littéraires

	Business text	General fiction	Lord of the Rings
GSL 1 k (%)	73.70	82.01	82.95
GSL 2 k (%)	6.13	5.23	5.41
AWL (%)	11.15	1.31	0.52
BWL (%)	2.60	0.71	0.41
Sous-total (%)	93.58	89.26	89.29
Abréviations etc. (%)	0.83	0.01	0.00
Total (%)	94.41	89.27	89.29

Dans les textes académiques commerciaux, la liste de mots académiques et la liste de mots commerciaux ont une couverture considérablement plus large que dans les textes tirés de la fiction. Les mots de la liste de mots académiques sont 10 fois plus fréquents et ceux de la liste de mots commerciaux six fois plus fréquents dans les textes commerciaux que dans *Le Seigneur des anneaux*. La couverture globale de ces listes, dans ce domaine spécialisé, est également plus large ; quelque 5 % de plus que dans l'un ou l'autre des deux corpus de fiction. Néanmoins, la couverture supplémentaire fournie par même une liste de mots spécialisés très soigneusement construite telle que la liste de mots commerciaux semble assez modeste à environ 2,5%, et bien que la couverture globale approche les 95%, elle est encore juste en dessous de ce chiffre. Konstantakis a supprimé, tout comme Coxhead, les abréviations, les noms propres, les éléments

avec trait d'union et autres bizarreries qui sont normalement considérés comme assez transparents dans leur sens pour les lecteurs. Avec ceux-ci inclus dans le compte, ce chiffre s'élève à plus de 95%. Il semble qu'un large vocabulaire, par milliers, puisse être requis pour la compréhension même dans ce qui semble être un domaine restreint et discret.

Bien entendu, tous les domaines spécialisés peuvent ne pas être les mêmes, et certains peuvent nécessiter beaucoup plus de vocabulaire que d'autres.

Cela soulève des doutes quant à savoir s'il est toujours possible, voire généralement possible, de réduire très sensiblement la charge d'apprentissage du vocabulaire en limitant soigneusement le domaine de la langue à étudier. Construire un vocabulaire large de plusieurs milliers de mots apparaît comme une condition absolue pour pouvoir bien fonctionner dans une langue étrangère.

3.7 Couverture dans différentes langues

La loi de Zipf et les commentaires de Palmer sur l'utilité d'un vocabulaire très fréquent ne sont pas destinés à être limités à la langue anglaise uniquement, mais devraient être applicables à toutes les langues. La question se pose donc de savoir dans quelle mesure ces idées se traduiront dans d'autres langues et si les mêmes types de chiffres pour une compréhension minimale et plus générale peuvent être produits, par exemple, en français, en arabe ou en chinois.

Il n'est pas inévitable que chaque langue produise des résultats comparables à ceux de l'anglais. Il peut y avoir plusieurs raisons à cela. La première est que les mots les plus fréquents en anglais incluent des pronoms et des prépositions, mais toutes les langues ne partagent pas cette qualité. Les langues agglutinantes, comme le hongrois, le finnois et le turc, gèrent de nombreuses fonctions et significations que ces mots véhiculent en anglais, de manière assez différente. En règle générale, ces significations sont transmises par l'ajout de suffixes à la forme racine d'un verbe ou d'un nom, de sorte qu'une seule famille de mots peut inclure beaucoup plus de formes de mots que ce ne serait le cas en anglais.

Une deuxième raison pourrait résider dans le développement historique de l'anglais. Les locuteurs de l'anglais se targuent souvent d'avoir une langue avec un vocabulaire particulièrement riche et l'anglais semble souvent avoir une variété de mots disponibles pour à

peu près la même idée. Pour des raisons historiques, l'anglais différencie par exemple de nombreux animaux de basse-cour et la viande qui en provient, entre *porc* et *cochon*. Si c'est vraiment le cas, la compréhension de l'anglais peut nécessiter un vocabulaire plus large que ce qui serait nécessaire pour une compréhension équivalente dans une autre langue.

De plus, les langues peuvent différer considérablement dans leurs processus de formation des mots. Certaines langues, comme l'allemand, favorisent la composition et combinent des mots existants pour former un nouveau mot. D'autres langues comme l'anglais peuvent le faire aussi, bien sûr ; *brief*, par exemple, est combiné à *case* pour créer *briefcase*, pour décrire la mallette particulière utilisée pour transporter des papiers commerciaux. Mais, en anglais, ce procédé n'est pas utilisé de la même ampleur qu'en allemand, et cela donne à l'allemand sa particularité de mots très longs. En revanche, d'autres langues favorisent la création ou la dérivation de mots pour de nouveaux concepts et idées. L'effet de ce processus de combinaison de mots peut être aggravé, car toutes les langues ne sont pas aussi claires que les langues d'Europe occidentale pour signaler les limites des mots.

Jusqu'à récemment, les corpus bien construits dans des langues autres que l'anglais étaient rares. Mais à mesure que le texte devient de plus en plus disponible sous forme numérique, il devient possible de construire une plus grande variété de corpus linguistiques et de comparer le nombre de mots dans différentes langues. Milton (2009) estime qu'il y a deux principales questions qui doivent être posées sur l'applicabilité de la loi de *Zipf*. La première est de savoir si la loi de *Zipf* est vraiment valable, ou fonctionne de manière équivalente, dans d'autres langues. En anglais, environ 2 000 mots donneront une couverture d'environ 80 % : est-ce le même ou sensiblement différent en français, en grec ou en chinois, par exemple ? La seconde est de savoir si la couverture et la compréhension seront toujours équivalentes. Une couverture à 80 % vous offrira-t-elle plus de compréhension et de communicabilité en français, par exemple, qu'en anglais ou en grec ?

Pour Milton (2009), la réponse à la première question est que, lorsqu'il existe des comptes équivalents pour différentes langues, la loi de *Zipf* apparaît remarquablement robuste même s'il existe des variations entre les langues. Une caractéristique de toutes les langues semble être qu'un petit nombre de mots sont très fréquents et couvrent une grande partie d'un texte.

Quelques exemples permettront d'illustrer cela. La figure 3-4 superpose la ligne de couverture du corpus d'anglais de Carroll et al. (1971) avec la couverture du corpus national

hellénique (Hatzigeorgiu et al., 2001) lorsqu'elle est lemmatisée pour fournir quelque chose comme une liste équivalente. Au départ, les premiers mots sont comparativement plus fréquents

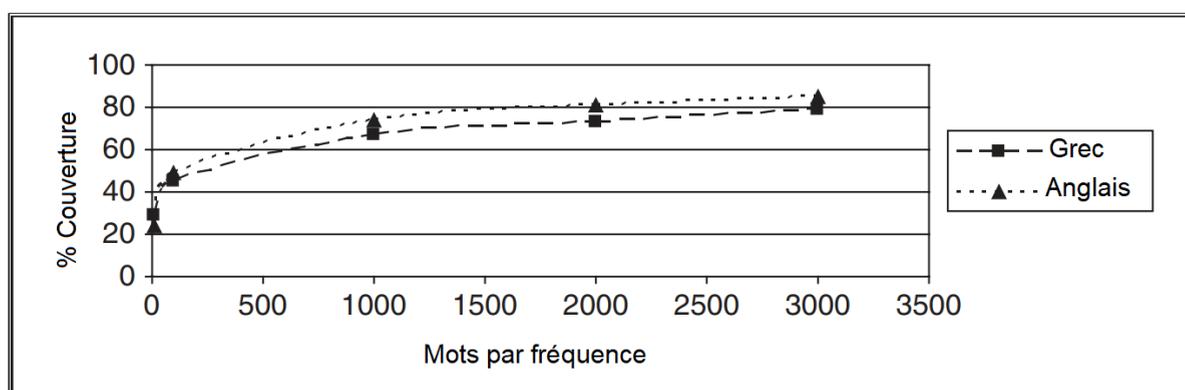


Fig. 3-4 : Comparaison de la couverture entre Carroll et al. et le Corpus National Hellénique (Milton, 2009)

en grec qu'en anglais ; en grec, l'article défini est très fréquent même par rapport à l'anglais. Par la suite, le vocabulaire grec offre proportionnellement moins de couverture et les deux lignes se croisent. Les 5 000 mots les plus fréquents du corpus grec offrent une couverture de 82,6 %, ce qui est nettement moins que les 5 000 mots les plus fréquents en anglais. Une particularité du grec est le nombre très élevé de hapax⁵, qui constituent 49,4 % du corpus en grec, mais se rapprochent des 30 % en anglais et en français. Néanmoins, les détails de la couverture, et les lignes du graphique, sont globalement similaires et la conclusion à tirer est que la loi de Zipf fonctionne en grec comme en anglais même si elle ne fonctionne pas de manière identique dans les deux langues, concluent (Hatzigeorgiu et al., 2001).

Le français présente des caractéristiques similaires et les chiffres de couverture sont souvent encore plus proches de l'anglais. Encore une fois, les quelques mots les plus fréquents en français sont plutôt plus fréquents que les plus fréquents en anglais mais, contrairement au grec, cette tendance ne se perd pas aux niveaux les moins fréquents. Milton précise que, dans le contexte du type de graphiques utilisés, jusqu'à présent les différences sont faibles et la superposition des deux lignes de couverture sur le même graphique, comme dans la figure 3-5, donne deux lignes presque indiscernables. Néanmoins, le même nombre de mots donne

⁵ Du grec hapax legomenon, « chose dite une seule fois ». Fait linguistique isolé, c'est-à-dire réduit à une seule attestation dans un état de langue ou bien dans un corpus donné. » (Neveu, 2004, p. 148)

systématiquement une couverture légèrement plus importante en français qu'en anglais et dans certains domaines, il existe des différences potentiellement importantes.

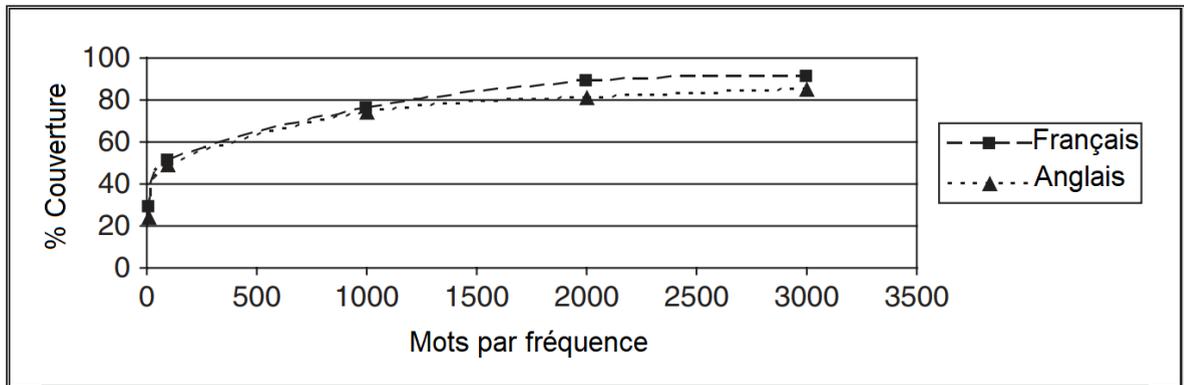


Fig. 3-5 : Comparaison de la couverture entre le corpus anglais de Carroll et al. (1971) et le corpus français de Baudot (1992)

Cobb et Horst (2004) soulignent la couverture des textes académiques par les 2000 mots français les plus fréquents. Le chiffre qu'ils citent de près de 89 % (Cobb & Horst, 2004, p. 30) équivaudrait à la *General Service Word List* de 2000 mots plus la *Coxhead's Academic Word List*, quelque 2600 mots soigneusement sélectionnés plutôt que des mots purement fréquents, en anglais. Il apparaît qu'en français, le vocabulaire le plus fréquent est au service du langage courant et du vocabulaire académique spécialisé qu'exige l'anglais. Encore une fois, la loi de Zipf selon laquelle les mots les plus fréquents fourniront une couverture disproportionnée dans un texte est valable, mais les détails de cette couverture semblent différer d'une langue à l'autre.

La réponse à la deuxième question, à savoir si la même quantité de texte ou le même nombre de mots fournira le même niveau de compréhension dans toutes les langues, est beaucoup plus difficile à répondre. Il y a un manque de recherche directe dans ce domaine sur laquelle s'appuyer.

Il n'y a pas de réponses faciles à la question de savoir si la même couverture ou la même connaissance du vocabulaire est nécessaire pour une compréhension équivalente dans différentes langues, et cela nécessite une enquête plus approfondie.

3.8 Conclusion

Ce chapitre a examiné si les mots les plus fréquents dans une langue sont aussi les plus utiles, comme le suggère Palmer (1917). Cette idée semble avoir un mérite considérable. Les mots les plus fréquents dans n'importe quelle langue contribuent énormément à la couverture dans un texte, et une couverture considérable est nécessaire avant qu'un lecteur ou un auditeur ait suffisamment d'informations pour être capable de saisir le sens dans la plupart des langues normales. Ceci fournit une deuxième raison pour concentrer les tests de connaissance du vocabulaire sur les mots les plus fréquents de la langue : cette information doit nous dire si les mots que connaît un apprenant sont utiles et peuvent contribuer à la communication.

Lors de l'examen de la couverture, un certain nombre de chiffres ont émergé et sont fréquemment répétés. L'un d'eux est le chiffre de 95 % de couverture pour une compréhension raisonnable ou complète. Malgré les difficultés qui entourent ce chiffre, comme ce que l'on entend exactement par compréhension, il continue d'être utilisé, probablement parce qu'il y a là une sorte de vérité, que nous comprenons tous. Les apprenants devront connaître la quasi-totalité d'un texte avant de pouvoir en tirer un sens réel. Cela souligne encore l'importance des mots les plus fréquents, car ces mots sont si fréquents qu'il est impossible d'atteindre une couverture de 95 % sans eux. Un deuxième chiffre de 80 % de couverture est apparu comme un seuil en dessous duquel la compréhension et la communication deviennent presque impossibles dans autre chose que les circonstances les plus artificielles et les plus limitées. Cela suggère que les apprenants ont besoin d'une connaissance considérable du vocabulaire avant de pouvoir commencer à fonctionner de manière autonome. L'idée que la communication peut être réalisée avec un très petit nombre de mots soigneusement sélectionnés semble fautive malgré sa popularité.

En anglais, ces chiffres suggèrent que les tests qui évaluent la connaissance et l'utilisation des 2000 premiers et 5000 mots les plus fréquents peuvent être particulièrement utiles pour une évaluation des compétences générales ou académiques. Il semble probable que dans des langues similaires à l'anglais, des tests de taille similaire devraient donner des résultats tout aussi perspicaces, mais comme les langues varient dans la fréquence de leurs mots et la couverture qu'elles fournissent, comparer le vocabulaire entre différentes langues est un exercice semé d'embûches. Ce n'est qu'en prenant ces mesures et en essayant de les appliquer à notre compréhension de la compréhension, cependant, que nous pourrions clarifier le fonctionnement des langues et faire le genre de comparaisons inter-langues qui, dans une

société multilingue, seraient si utiles. Le chapitre qui suit rendra compte des mesures de vocabulaire et tentera de donner un sens aux résultats qui en ressortent.

CHAPITRE 4 :

MESURE DE L'ÉTENDUE DU VOCABULAIRE PASSIF

4.1 Introduction

Alderson (2005) a souligné que nous en savons vraiment très peu sur ce qui est normal et ce qui est anormal dans le développement d'une langue étrangère. Pour mieux comprendre le processus de la langue étrangère, la collecte de données normatives est essentielle. Ce n'est qu'alors que nous pourrions commencer à comprendre non seulement ce qui est appris et quand, mais aussi comment les individus et les groupes varieront. La mesure de la connaissance du vocabulaire chez les apprenants de langues étrangères devrait faire partie de ce processus. Dans les deux chapitres précédents, il a été soutenu qu'il est possible de construire un test bien dirigé de la connaissance du vocabulaire en ciblant la connaissance d'un apprenant des mots les plus fréquents dans une langue. Cela devrait permettre de se faire une estimation des mots que l'apprenant est le plus susceptible d'avoir rencontré fréquemment, et est donc susceptible de connaître. Il doit permettre d'estimer l'utilité du vocabulaire de l'apprenant à des fins de compréhension et de communication. Nous évoquerons dans ce chapitre les tests d'étendue du vocabulaire qui sont conformes à cette spécification ainsi que les mesures qu'ils nous donnent lorsqu'ils sont appliqués aux apprenants.

4.2 Tests d'étendue du vocabulaire

Certains des tests de vocabulaire les mieux documentés sont des tests de liste de contrôle de reconnaissance passive du vocabulaire, conçus pour donner une estimation de l'étendue ou de la taille du vocabulaire. Milton (2009, p. 71) reconnaît la reconnaissance passive comme étant probablement la définition la plus basique de la connaissance des mots ; l'apprenant reconnaît la forme d'un mot et qu'il s'agit d'un mot plutôt que d'un fouillis insignifiant de symboles ou de sons. Toute autre qualité de connaissance devrait relever d'une telle estimation, car vous ne pouvez pas traduire ou connaître les détails de la collocation d'un mot que vous ne pouvez même pas reconnaître comme un mot. Le format des tests de liste de contrôle est trompeusement simple.

L'apprenant est présenté avec une série de mots et est invité à cocher ceux qu'ils reconnaissent. Un exemple du format est donné dans la figure 4-1.

Please look at these words. Some of these words are real French words and some are invented but are made to look like real words. Please tick the words that you know or can use. Here is an example.

chien

Thank you for your help.

de distance abattre absurde achevé manchir

Fig. 4-1 : Exemple d'un test de liste de contrôle (version française de Meara et Milton, 2003)

Il est vrai qu'il n'existe pas de méthode de test parfaite et cette méthode de liste de contrôle a ses avantages et ses inconvénients. Les avantages comprennent la vitesse et facilité relatives avec lesquelles ces tests peuvent être construits sous la base d'une liste de fréquences convenablement construite. Les résultats sont susceptibles d'être plus fiables à la fois parce qu'un échantillon plus grand est toujours susceptible de donner de meilleurs résultats qu'un plus petit échantillon, et parce que le test peut être relativement bref et par conséquent, il y a moins de chance pour que les apprenants s'ennuient et perdent leur concentration. Le test sélectionne généralement des mots dans toutes les bandes de fréquences et ce format a également un énorme avantage parce qu'il permet de créer rapidement et facilement plusieurs versions du test, avec différents mots sélectionnés parmi les bandes de fréquences. L'inconvénient est le degré auquel les apprenants sont capables de deviner, quand ils ne reconnaissent pas vraiment un mot ou ne sont pas sûrs.

Milton (2009) a décrit ce type de test comme étant d'une simplicité trompeuse car, dans ces tests, les apprenants sont souvent confrontés à une décision délicate. S'ils pensent reconnaître un mot mais ne sont pas sûrs, comment faut-il le noter ? Un Oui direct peut entraîner l'inclusion de mots qui ne sont pas du tout connus dans l'estimation de la taille, ce qui entraîne une surestimation, mais l'omission peut entraîner une sous-estimation, car les mots qui sont connus,

au moins partiellement, sont omis. Certains de ces tests incluent une catégorie *Pas sûr* pour aider à séparer des mots comme celui-ci. Plus généralement, de faux mots sont souvent inclus dans ces tests pour permettre une compensation pour ce type d'incertitude et pour de pures conjectures. Les faux mots sont des mots construits pour se lire et sonner comme de vrais mots, mais qui n'existent pas vraiment. Dans la figure 4-1, *manchir* est un faux mot. Ce sont des mots que les apprenants ne peuvent pas avoir rencontrés et qu'ils ne connaissent pas et qu'ils ne doivent pas cocher. La proportion de réponses Oui, ou de coches, à ces mots faux permet d'estimer le degré de surestimation qu'un apprenant fait et les scores peuvent être ajustés sur cette base. L'idée est très séduisante et on pourrait penser qu'un système qui vous permet de faire une sorte de compensation pour les conjectures ou la surestimation serait une bonne chose. Cependant, il semble que les apprenants réagissent très différemment face à un test de liste de contrôle. Shillaw (1999) rapporte, par exemple, que les apprenants japonais qu'il a étudiés étaient si prudents dans leurs estimations de leurs propres connaissances que ces faux mots étaient très rarement cochés. Al-Hazemi et Vassiliu (cité dans Milton, 2009) rapportent que les apprenants en Arabie Saoudite et en Grèce prennent beaucoup plus de temps de réflexion dans ce type de test, peut-être parce qu'ils sont formés à l'examen technique, mais que les tests semblaient néanmoins fonctionner de manière fiable. Eyckmans (cité dans Milton, 2009), dans une étude sur les apprenants belges, signale d'énormes surestimations ; jusqu'à 60% des mots faux ont été, en moyenne, identifiés comme réels. Une partie du problème identifié est due à la tendance des personnes, confrontées à des questions Oui/Non de ce type, à répondre par *Oui* quelle que soit la question. Il avertit cependant que tous les tests peuvent ne pas fonctionner aussi bien avec tous les apprenants, et si les apprenants ne veulent pas ou ne peuvent pas interagir avec le test de la manière attendue, alors les données qu'ils fournissent peuvent être sans valeur.

Deux de ces tests sont délibérément construits pour donner une estimation de la taille du vocabulaire, dans des limites de fréquence strictes, et sembleraient très appropriés pour des mesures qui cherchent à évaluer la croissance du vocabulaire au fil du temps. L'un est le Vocabulary Size Test de l'Eurocentre (Meara & Jones, 1990), qui est auto-adaptatif, teste environ 150 mots et forme une estimation de la connaissance d'un apprenant des 10 000 mots les plus fréquents. X-Lex (Meara & Milton, 2003) teste 120 mots, 20 mots choisis au hasard dans chacune des cinq premières bandes de fréquence de 1000 mots et 20 autres pseudo-mots.

Le nombre de réponses Oui à ces pseudo-mots permet d'ajuster le score sur les vrais. La connaissance du vocabulaire d'un apprenant est calculée en comptant le nombre de réponses Oui à des mots réels et en le multipliant par 50 pour donner un score brut sur 5000. Ce chiffre est

déduit du score brut pour donner un score ajusté, également sur 5000, qui comprend ainsi une compensation pour les conjectures. Il n'y a pas de limite de temps pour le test, qui ne prend généralement que 5 ou 10 minutes.

Un deuxième type de méthode de test, très répandu, pour estimer la connaissance du vocabulaire consiste à demander aux apprenants de démontrer qu'ils connaissent des traductions ou des explications de mots en langue étrangère. Les tests peuvent être passifs et fournir aux apprenants des traductions ou des explications parmi lesquelles choisir, ou plutôt plus productifs en exigeant de l'apprenant qu'il produise un mot en langue étrangère en réponse à un stimulus dans sa langue maternelle. Dans la littérature, ces deux tests sont appelés tests de reconnaissance et de rappel. Le test des niveaux de Nation (1990) est un exemple de ce type de test de reconnaissance où les apprenants reçoivent des mots de test dans la langue étrangère et une sélection d'explications qui doivent être mises en correspondance. Un exemple est donné à la figure 4-2.

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning.	
1 business	
2 clock	_____ part of a house
3 horse	_____ animal with four legs
4 pencil	_____ something used for writing
5 shoe	
6 wall	

Fig. 4-2 : Exemple de test de niveau tiré de Nation (2001, p. 416)

Ce format permet plutôt de tester la reconnaissance passive de la forme des mots et cette forme de test devrait permettre de former une estimation de la connaissance des mots et de leurs significations. Il s'agit cependant d'un test assez complexe, où la réussite ne repose pas seulement sur la connaissance par l'apprenant des mots du test (sur le côté gauche), mais aussi sur la connaissance des mots dans les explications (sur le côté droit), et il n'est pas tout à fait clair quels éléments sont testés. De plus, chaque question contient plusieurs éléments de test, et la connaissance de certains éléments par l'apprenant est susceptible d'avoir un impact sur la capacité de trouver les réponses à d'autres éléments lorsque ceux-ci ne sont pas connus. Nous savons que les apprenants essaient souvent de maximiser leurs scores en faisant des suppositions

éclairées dans ces circonstances, mais il est beaucoup moins facile de déterminer les effets des conjectures dans un test de ce genre et il n'y a pas de mécanisme dans le test pour faire ce genre de calcul.

Il n'est pas certain que connaître 1000 mots en allemand, par exemple, fournira autant de compétences linguistiques potentielles que connaître le même nombre de mots en anglais. Les estimations de la connaissance du vocabulaire dans ces différentes langues peuvent nous dire des choses très différentes sur les connaissances et les performances linguistiques globales. Néanmoins, il devient possible de tenter ces investigations et de considérer, à la lumière des données qui se dégagent, à quel point ce type d'analyse est informatif.

4.3 Prédications de croissance de l'étendue du vocabulaire

Il semble qu'il soit possible de tester la connaissance du vocabulaire de manière significative, de tester les apprenants du niveau débutant jusqu'à des niveaux de connaissance assez élevés, et de faire de bonnes estimations de leur connaissance du vocabulaire dans les bandes les plus fréquentes ; alors à quoi s'attendre à voir dans une étude longitudinale d'un apprenant de langue, ou dans une étude transversale d'apprenants de différents niveaux ? Certes, nous nous attendrions à voir une croissance du vocabulaire. Vous ne pouvez pas être bon dans une langue étrangère sans un vocabulaire important. Il se peut que l'inverse soit vrai, connaître beaucoup de mots dans une langue étrangère mais être incapable de les utiliser. Mais, pour qu'un apprenant soit capable de fonctionner de manière fluide et idiomatique dans la grande variété de registres et d'environnements linguistiques que la plupart des locuteurs natifs gèrent avec facilité, un vocabulaire très étendu est nécessaire. Les chiffres sont considérables. S'ils veulent bien lire et écrire, alors les apprenants de langues étrangères capables devront probablement maîtriser des milliers de mots.

Cela implique que, dans un bon cours d'instruction, nous pourrions nous attendre à voir un apport régulier de vocabulaire soigneusement sélectionné. Et nous espérons, vraisemblablement, une croissance assez régulière du vocabulaire de la part des apprenants. Un cours qui se concentrerait uniquement sur l'apprentissage du vocabulaire, ou l'omettrait entièrement, serait un cours étrange et probablement un échec. Avec une tâche importante telle que l'apprentissage du vocabulaire, il serait beaucoup plus logique de diviser la tâche en éléments plus petits et plus faciles à gérer et de les aborder régulièrement tout au long de l'enseignement.

C'est l'hypothèse de nombreux écrivains. Scholfied (cité dans Milton, 2009), par exemple, dresse des graphiques montrant l'apport idéalisé d'un manuel de langue étrangère, ce qui présuppose qu'une charge de vocabulaire égale par unité ou heure d'apprentissage pourrait être une norme intelligente.

Il y a aussi une idée que l'assimilation du vocabulaire peut varier selon le niveau des apprenants ou les besoins des apprenants, et cet apport peut être adapté en fonction des capacités des apprenants. Gairns et Redman (1986, p. 66), par exemple, suggèrent que l'apprentissage de 8 à 12 nouveaux mots productifs par heure serait une aspiration raisonnable pour les apprenants, mais ils supposent que plus l'apprentissage de la langue progresse, plus les apprenants seront capables d'apprendre efficacement. Le chiffre supérieur qu'ils suggèrent convient aux apprenants plus avancés et le chiffre inférieur aux niveaux élémentaires. Il n'est pas difficile de trouver une justification à cela, bien que Gairns et Redman n'en suggèrent pas eux-mêmes. Les apprenants les plus avancés pourraient être en mesure d'utiliser leurs connaissances existantes du lexique et de la morphologie pour mémoriser de nouvelles dérivations ou combinaisons. Les apprenants de niveau inférieur n'auraient pas cette connaissance ou cette capacité. Il s'agit toutefois de spéculations et il n'est pas certain que ce soit ce que font réellement les apprenants.

Il existe un contre-argument potentiel à cela, qui suggère qu'un chargement en amont des mots les plus fréquents dans un cours serait souhaitable. Nation (2001, p. 16) reconnaît l'importance de ces mots fréquents à la communication et au sens et, par conséquent, conseille que tous les efforts soient faits pour enseigner ces mots dès le début de l'apprentissage.

« *Almost 80% of the running words in the text are high frequency words.* » (Nation, 2001, p. 16)

Potentiellement, les apprenants peuvent connaître une poussée d'apprentissage du vocabulaire au tout début, car ils sont gavés de force avec ces mots. La concentration explicite sur le vocabulaire diminuerait plus tard au cours de l'apprentissage. Il n'est pas difficile de trouver une justification à l'idée d'accentuation du vocabulaire et d'avancer l'idée d'une forte concentration sur le vocabulaire dès les premiers stades. Comme le chapitre précédent l'a montré, la connaissance du vocabulaire le plus fréquent peut permettre aux apprenants d'atteindre une certaine indépendance dans la compréhension et l'utilisation de la langue. L'avantage pour la motivation d'amener les apprenants le plus rapidement possible à ce niveau d'indépendance devrait être considérable. Par la suite, les apprenants pourraient prendre la responsabilité et

élargir davantage leur connaissance du vocabulaire. Il a été également suggéré que si les apprenants ne progressaient pas dans la construction de cette connaissance de base du vocabulaire, cela devrait être une source de préoccupation. Les apprenants de bas niveau qui n'acquerraient pas de vocabulaire supplémentaire n'avanceront probablement pas, car un vocabulaire de plusieurs milliers de mots semble être une condition de pleine compréhension de tout texte normal. Cependant, il est concevable que des apprenants très avancés, qui ont déjà un vocabulaire étendu, puissent développer la profondeur et la fluidité d'utilisation au détriment de la largeur, afin de faire des gains et de devenir plus compétents.

4.4 Attentes en matière de connaissances du vocabulaire dans les examens et autres programmes d'études

Certaines informations sur la charge de vocabulaire des cours sont fournies dans les programmes fournis par les ministères de l'éducation et par les organismes d'examen. Certains d'entre eux sont étonnamment détaillés. Par exemple, le programme national de base hongrois (Krizsán, cité dans Milton, 2009) donne des chiffres précis sur le vocabulaire anglais actif et passif que les élèves devraient acquérir à la fin des différentes années dans les écoles primaires et secondaires. Un résumé est fourni dans le tableau 4-1.

Tableau 4-1 : Objectifs de croissance du vocabulaire actif et passif dans le programme national de base hongrois

	3 ^{ème} année	4 ^{ème} année	5 ^{ème} année	6 ^{ème} année	7 ^{ème} année	8 ^{ème} année
Vocabulaire actif	200	350	500	600	800	1200
Vocabulaire passif	150	150	200	250	300	400
Vocabulaire actif et passif	350	500	700	850	1100	1600

Source : Krizsán (cité dans Milton, 2009)

Cela suggère que l'apprentissage du vocabulaire ne sera pas nécessairement un processus

régulier. Il semble que les apprenants devraient gagner 350 mots au cours de leur première année, mais n'en ajouteront 150 que lors de leur deuxième année d'études. Le contact en classe dans la langue étrangère reste cependant le même.

En 8e année, l'absorption devrait être dix fois plus importante qu'en 4e année, quelque 500 mots devraient être ajoutés, et cela ne s'explique qu'en partie par les heures de classe supplémentaires. Au moment où les apprenants passent les examens de niveau de maturité (niveau B1), un vocabulaire cible d'environ 3000 mots sont attendus (Orosz, 2009). Globalement, cela anticipe un apprentissage de vocabulaire d'environ 260 à 270 nouveaux mots par an.

Tous les ministères de l'Éducation ne sont pas aussi explicites dans leurs objectifs curriculaires. Comme Häcker (2008) le note, les niveaux de réussite du *British National Curriculum* omettent toute mention de vocabulaire jusqu'au niveau 8, et décrivent ensuite la réussite uniquement dans les termes les plus généraux. Les jurys d'examen formels, cependant, fixent souvent des chiffres utiles pour guider les apprenants, les enseignants et les rédacteurs du matériel de test qu'ils utilisent. Le *Cambridge First Certificate in English* (FCE) suppose une connaissance d'environ 4 500 mots pour que les apprenants puissent passer et réussir cet examen (Hindmarsh, 1980), qui est lancé au CECR B2, et suppose que les apprenants auront eu un minimum de 600 contacts heures de cours pour atteindre ce niveau de connaissances (University of Cambridge Local Examinations Syndicate, 2001, p. 14). Au Royaume-Uni, les jurys d'examen de français langue étrangère produisent des listes de vocabulaire de base minimum pour les examens du GCSE de 16 ans, CECR B1, contenant environ 1000 à 1500 items (Milton & Meara, 1998). Cela implique que les apprenants sont généralement censés en savoir plus que cela, mais on ne sait pas combien plus. Dans ces cas, les mots font clairement référence à des lemmes ou à des familles de mots, car les pluriels et autres inflexions et dérivations sont exclus des listes, tout comme les éléments tels que les nationalités et certains parents. Quelles que soient les limites de ce type de liste et de taille de vocabulaire cible, il doit être utile d'avoir des cibles par rapport auxquelles les enseignants peuvent programmer leur enseignement de la langue et les apprenants peuvent mesurer leurs progrès.

Milton (2009) estime que ce sont des considérations importantes dans la construction d'un programme bien conçu, mais il semble y avoir un manque de données qui peuvent suggérer comment le vocabulaire se développe au fil du temps et de l'apprentissage, et qui pourraient fournir un contexte au contenu et au chargement du programme. Burns (1951) et Robson (1934), par exemple, se concentrent sur la première année d'apprentissage du français dans les écoles

britanniques. Ils ne donnent pas une image de la façon dont le lexique se développe en largeur, ou toute autre caractéristique de la connaissance, au cours de l'apprentissage et tout au long d'un système éducatif. Ces chercheurs ont rencontré de très réelles difficultés, en l'absence de corpus fiables, pour construire des tests à différents niveaux de connaissances et, presque certainement, des contraintes de temps les ont empêchés d'étendre leur étude au-delà d'une seule cohorte pendant un an. Mais, en utilisant des techniques de test basées sur la fréquence, il est désormais possible de réaliser des études transversales et d'examiner ce qui se passe au cours de l'apprentissage.

4.5 Études sur la croissance de l'étendue du vocabulaire

Les tests de vocabulaire basés sur la fréquence existent maintenant depuis suffisamment longtemps pour qu'un certain nombre d'études dans différents pays aient été entreprises. A notre connaissance, ce genre d'initiatives n'a jamais été envisagé ou projeté en Algérie.

Jusqu'ici, les études qui ont été faites concernent les pays où les objectifs sur le vocabulaire sont clairs et où les résultats des tests peuvent être comparés. En Grèce et en Hongrie, il y a les chiffres du FCE et du programme de base national hongrois et l'impression est donnée d'un environnement d'apprentissage des langues étrangères robuste. Au Royaume-Uni, il y a aussi des objectifs, mais des doutes réguliers sont exprimés dans la presse et même par la *Qualifications and Curriculum Authority* (QCA) quant aux standards et niveaux atteints par les apprenants. (Milton, 2009, p. 79)

Milton (2009) cite l'exemple des écoles privées d'anglais langue étrangère (EFL) en Grèce (appelées *frontisteria*). Selon lui, le progrès y ressemble beaucoup à l'idéal. La figure 4.3 présente la taille moyenne du vocabulaire de sept classes sur sept niveaux annuels, des débutants aux étudiants se préparant à passer le FCE au niveau B2 du CECR. Chaque apprenant de cette école a été testé à la fin de l'année d'enseignement en juin à l'aide du test X-Lex (2003) de Meara et Milton, fournissant une estimation du nombre de mots connus parmi les 5000 mots lemmatisés les plus fréquents en anglais. Les 227 apprenants de cette étude ont reçu 100 heures d'enseignement en classe par an au cours des cinq premières années (Junior à niveau D) et 125 heures d'entrée en années 6 et 7 (classe E et la classe FCE). Généralement, les classes progressent à travers l'école et la série de niveaux sur un cycle annuel sans processus de sélection. Cependant,

la classe FCE est légèrement différente des autres en ce sens que seules les personnes jugées capables de passer le FCE et de réussir seront admises.

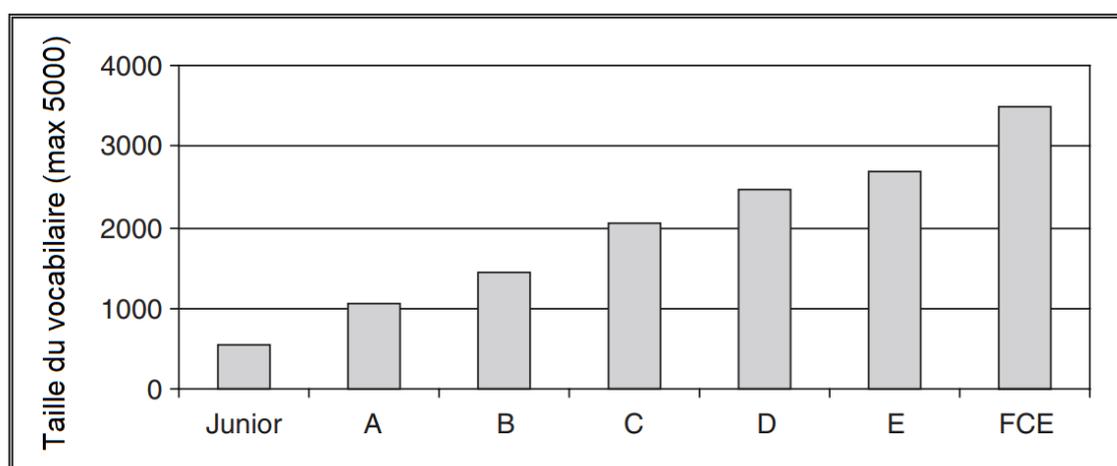


Fig. 4-3 : Croissance lexicale des apprenants d'EFL en Grèce (Milton, 2006a)

Milton relève un certain nombre de caractéristiques de l'apprentissage du vocabulaire dans une des écoles qu'il a visitée en Grèce. Ces caractéristiques comprennent une croissance de vocabulaire très régulière et une croissance assez substantielle. Environ 500 mots parmi les 5000 les plus fréquents sont appris chaque année au cours du cycle de sept ans jusqu'à ce qu'ils atteignent le FCE niveau avec une connaissance d'environ 3 500 mots sur les 5 000 les plus fréquents. Ce chiffre correspond bien aux propres estimations de Cambridge sur la taille probable du vocabulaire nécessaire pour réussir cet examen. Les listes de Hindmarsh (1980) comprennent environ 4 500 éléments avec une concentration, sans surprise, aux niveaux les plus fréquents, mais aussi une gamme de vocabulaire de basse fréquence thématiquement lié qui n'aura pas été inclus dans le test de 5 000 mots. (Milton 2009). L'impression pour les apprenants en tant que groupe est très régulière. Il n'y a aucune preuve du type de chargement initial du vocabulaire suggéré par Nation. En vérité, ces données sont tirées de ce qui est probablement une très bonne école. C'est une école populaire avec d'excellents résultats en des examens externes et une forte implication des parents dans la progression des apprenants qui commencent généralement les apprentissages à 7 ou 8 ans. Les propriétaires et les enseignants utilisent des techniques assez sophistiquées pour faire correspondre les livres de cours et le niveau de l'apprenant avec la connaissance du vocabulaire.

Alors que la plupart des apprenants progressent clairement et régulièrement dans le développement de leur vocabulaire, il est également clair que de nombreux élèves ne le font pas. Il y a un chevauchement considérable entre les classes. Certains élèves font ce qui semble être des progrès spectaculaires. Les résultats suggèrent que les meilleurs étudiants peuvent acquérir plus de 1000 nouveaux éléments lexicaux en première année et continuer à faire des progrès considérables par la suite. Ces apprenants, à la fin de ce programme d'études, semblent avoir une connaissance très réelle de la grande majorité des mots courants en anglais et sont vraisemblablement bien placés pour lire avec compréhension et bien communiquer en anglais. Tous les élèves ne sont pas aussi bons, bien sûr, et les élèves les moins bien notés ont des vocabulaires estimés dans la région de 400 à 500 mots à travers tous les niveaux jusqu'au groupe FCE.

Les étudiants les moins capables ont été omis de la classe FCE où l'étudiant le moins bien noté connaît environ 3000 mots. Cela reflète probablement une vérité générale selon laquelle, même dans les meilleures écoles, certains apprenants trouveront les exigences de l'apprentissage académique des langues plus qu'ils ne peuvent facilement y faire face et feront peu ou pas de progrès. Il est moins clair que le chiffre de 400 à 500 représente réellement avec précision les connaissances minimales des apprenants de ce groupe. Il y a toujours des problèmes d'échantillonnage où les apprenants ont très peu de connaissances sur le matériel testé et le format du test encourage les apprenants à s'accorder le bénéfice du doute. Le taux de fausses alarmes (le nombre de pseudo-mots identifiés comme réels) dans cet échantillon est toutefois relativement faible, ce qui suggère que le test fonctionne bien. Les enseignants de cette école ont estimé qu'il s'agissait d'une estimation réaliste, car même l'élève le moins performant aurait un peu de vocabulaire anglais acquis à partir de programmes télévisés, de films, de chansons et de publicités en Grèce où l'utilisation de l'anglais est courante.

La croissance du vocabulaire dans le système scolaire public hongrois semble présenter des caractéristiques similaires à celles des écoles grecques, avec des augmentations régulières et substantielles à chaque niveau scolaire à mesure que les apprenants progressent dans le système scolaire. Milton (2009) rapporte que Dans cette étude de l'Université de Szeged (Orosz, 2009), 726 apprenants de 2 écoles hongroises ont été testés à la fin de l'année scolaire. L'écart entre les deux écoles, lorsque les élèves passent de l'une à l'autre, se situe entre la 8e et la 9e année. Les données ont été recueillies auprès de débutants, commençant l'anglais en 3e année, jusqu'aux jeunes qui ont terminé l'examen de maturité en 12e année, alors que les apprenants devraient être au CECR Niveau B1. Les heures d'apprentissage ne sont pas tout à fait régulières tout au long

du programme d'études, les apprenants de 3e, 4e, 9e et 10e années recevant 55 heures d'enseignement, les années 5, 6, 7, 8 et 11 recevant 83 heures d'enseignement, et les apprenants au cours de leur dernière année recevant 100 heures au fur et à mesure qu'ils se préparent aux examens finaux. Les résultats d'Orosz, les scores moyens, maximum et minimum, sont présentés à la figure 4-4.

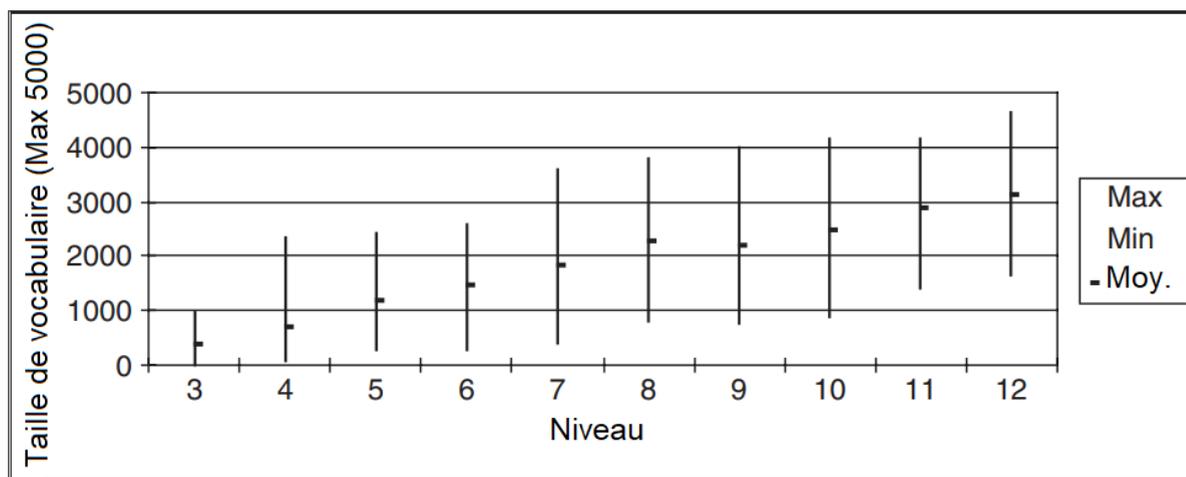


Fig. 4-4: Croissance lexicale des apprenants d'EFL en Hongrie (Milton, 2009, p. 80)

L'impression générale, comme pour les données grecques, est que l'apprentissage du vocabulaire parmi les classes d'apprenants est, en général, une activité très régulière.

Ceci est assez différent du programme qui anticipait des progrès loin d'être réguliers. Le seul coude dans le modèle de croissance se produit entre la 8e et la 9e année, lorsque les élèves terminent leurs études dans une école et progressent vers une nouvelle école à un niveau supérieur. Généralement, les apprenants, en moyenne, semblent ajouter 300 à 400 éléments de vocabulaire à leurs lexiques anglais chaque année à un rythme et un volume de progrès qui semblent être environ 50 % supérieurs aux exigences du programme national. Les apprenants passent leur examen de maturité B2 avec environ 3 500 mots ; presque identique aux apprenants grecs de l'anglais au même niveau B2. Comme pour les données grecques, cependant, il existe des variations individuelles considérables. Les élèves les moins bien notés au début des apprentissages semblent acquérir très peu de vocabulaire, à peine quelques centaines de mots en trois ou quatre années d'études. Cela contraste fortement avec les élèves les plus performants qui, comme dans les données grecques, semblent capables d'acquérir jusqu'à 1000 mots par an au début de l'apprentissage. Comme pour les résultats en grec, il n'y a pas de signe évident d'une

mise en avant du vocabulaire dans les premières classes suggérées par Nation, ni d'une assimilation plus rapide dans les classes plus avancées, comme suggéré par Gairns et Redman.

Toutes les classes d'apprenants de langues étrangères n'affichent pas le genre de régularité rassurante dans leurs progrès que les systèmes grec et hongrois ont montré. Les apprenants du français langue étrangère dans les écoles britanniques, par exemple, affichent un schéma de progrès très différent (Milton, 2006b). La figure 4-5 présente la taille moyenne du vocabulaire de 449 apprenants dans une école secondaire au Royaume-Uni et entrant à l'université. Les données sont une image transversale de l'apprentissage, avec des apprenants à chacune des sept années d'études, des débutants à 11 ans jusqu'aux examens nationaux, qui sont passés à 16 ans, GCSE (CEFR B1) et niveau « avancé » (CECR B2) à 18 ans. Les apprenants ont été testés en fin d'année d'enseignement. Ils ont reçu un peu moins de cours de langue étrangère que les apprenants grecs de la figure 4-3 avec environ 80 heures de cours chaque année en première, quatrième et cinquième années, environ 60 heures par an en années 2 et 3, et 100 heures ou plus pour les apprenants qui conservent le français en 6e et en 7e. Comme pour les apprenants grecs, un processus de sélection est en place aux niveaux d'études supérieurs, certains étudiants abandonnant la matière après la 4e année, et seuls les plus apprenants capables progressant vers

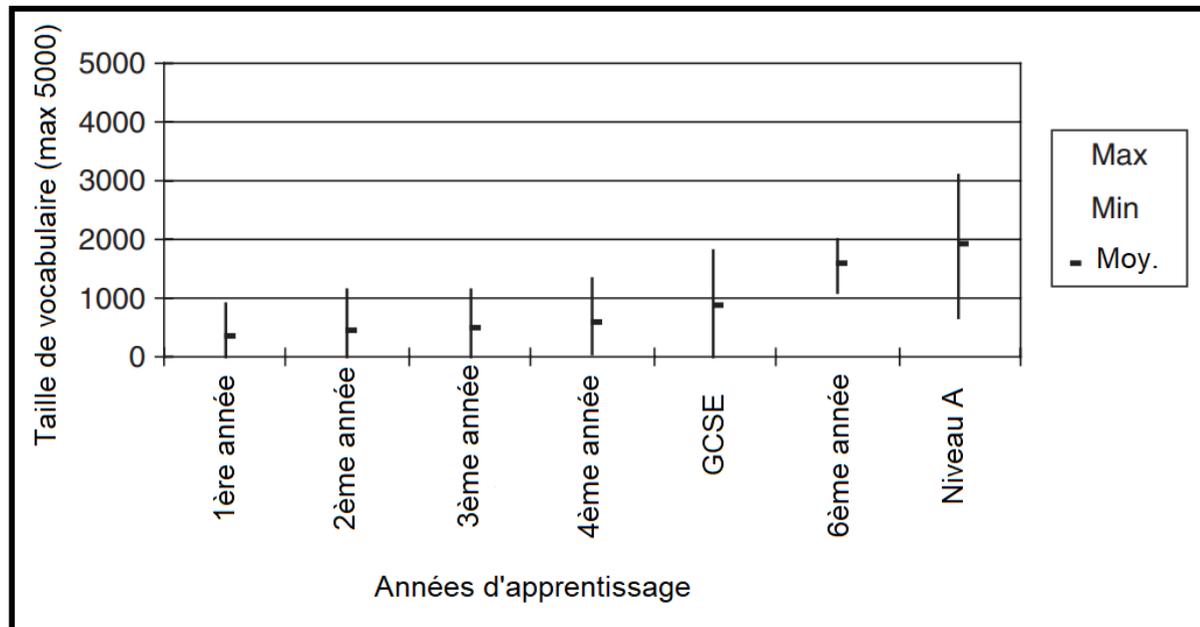


Fig. 4-5: Croissance lexicale des apprenants de français langue étrangère au Royaume-Uni (Milton, 2006b)

les 6 et 7 années, le niveau avancé. Ces apprenants ont été testés en utilisant une version française du X-Lex de Meara et Milton (2003), construit à partir des données fréquentielles du corpus de Baudot (1992) et en utilisant les mêmes critères de sélection des mots que pour le test d'anglais. Les caractéristiques de l'apprentissage du vocabulaire dans les écoles au Royaume-Uni comprennent croissance irrégulière et où, sur plusieurs années au moins, peu de progrès notables dans la connaissance du vocabulaire. Les apprenants semblent prendre un bon départ dans le développement du langage, sachant en moyenne environ 300 mots à la fin de la première année, mais par la suite, les connaissances semblent plafonner. Moins de la moitié de ce nombre de mots, environ 150 mots, semble être ajouté en moyenne au cours des deux prochaines années d'étude. Si cette baisse de la progression coïncide avec une baisse des heures d'entrée, il n'y a pas de raison évidente pour une baisse aussi marquée de la progression. Les apprenants de ce niveau n'ont que quelques centaines de mots dans leur langue étrangère et ne peuvent pas être communicatifs. On pourrait s'attendre à ce que la croissance du vocabulaire soit une priorité du programme à ce stade. Ainsi, si la très grande majorité des apprenants n'apprennent pas le vocabulaire dans leurs cours de langues étrangères, on peut supposer que le nouveau vocabulaire n'est pas contenu dans les matériels qu'ils rencontrent en classe. Häcker (2008) observe dans le programme du Royaume-Uni un effet de retour négatif lorsque les enseignants limitent l'input, y compris le vocabulaire, aux sujets limités qui apparaissent dans les examens.

Milton (2009) voit que cela soutient l'idée que les étudiants au Royaume-Uni n'obtiennent pas la gamme ou le volume de vocabulaire auquel les apprenants ailleurs sont exposés. Cette observation nécessite une enquête plus approfondie. Quelle qu'en soit la raison, ce n'est qu'à partir de la cinquième année d'apprentissage, lorsque certains apprenants sont autorisés à abandonner la matière, que l'apprentissage du vocabulaire semble reprendre. Les apprenants passent leurs examens B1 GCSE avec, en moyenne, 850 mots de français et le niveau B2 Advanced avec un peu moins de 2000 mots. (Milton, 2009, p. 84)

Ces chiffres sont reproduits dans une étude similaire menée ailleurs au Royaume-Uni (David, 2008a). Ces résultats sont considérablement inférieurs aux chiffres de vocabulaire de base minimum spécifiés par les examens d'État britanniques à ces niveaux. Même en tenant compte de la présence dans les listes de vocabulaire de base de mots peu fréquents en dehors du contenu du test utilisé dans cette étude, le chiffre semble petit et les apprenants ayant ce niveau de connaissance du vocabulaire le plus fréquent auraient du mal à communiquer. Placés à côté des résultats des apprenants de niveau B2 en Grèce et en Hongrie, ces chiffres apparaissent

également très faibles. Les apprenants au Royaume-Uni apprennent une langue différente, mais il n'est pas clair que cela explique la différence.

Comme pour les apprenants grecs, il existe une variation individuelle très considérable. Au cours de la première année d'études, certains apprenants semblent avoir acquis près de 1000 mots de français. Le même genre de chiffre que les apprenants les mieux notés en Grèce ont réussi, mais en seulement la moitié des heures de contact. Mais, systématiquement, les apprenants les moins bien notés semblent ne connaître aucun mot français. Ce n'est que lorsque les meilleurs apprenants sont sélectionnés pour l'étude de niveau avancé que la queue des apprenants à faible score disparaît. Probablement, ces chiffres reflètent le fait que certains apprenants, en effet, retirent peu ou rien de l'expérience d'apprendre le français à l'école. Ces apprenants ne sont pas entourés de français en dehors des cours, à travers des films, des chansons et des publicités, et il est beaucoup plus difficile pour ces apprenants d'accéder à la langue étrangère. Le français langue étrangère n'est pas considéré comme un atout éducatif ou professionnel par de nombreux élèves britanniques, voire la plupart, et leurs parents. Il semble tout à fait possible pour les apprenants de ne rien reconnaître visuellement même après plusieurs années de cours. Mais, plus curieux que cela est le plateau parmi les apprenants les mieux notés, qui se produit dans les deuxième, troisième et, dans une moindre mesure, quatrième années d'études. Il s'agit d'apprenants qui pourraient apparemment faire d'énormes gains de vocabulaire dès leur première année, mais qui semblent arrêter d'apprendre un aspect essentiel de leur langue étrangère après la première année, et pendant plusieurs années. Ce n'est que lorsque les élèves les plus faibles décrochent que ces apprenants semblent reprendre leur progression à environ 500 par an en moyenne, ce qui semble beaucoup plus satisfaisant. Encore une fois, cette observation mériterait une enquête plus approfondie.

Que peut-on faire de ces chiffres ? Quelles conclusions générales peut-on dessiner ? Dans les systèmes que Milton a qualifiés d'assez bons, la Grèce et la Hongrie, les apprenants se retrouvent avec un vocabulaire étendu. Les chiffres suggèrent des progrès réguliers, et des progrès annuels assez importants. Les objectifs curriculaires, fixés par les ministères ou les organismes d'examen, sont atteints ou dépassés. Certes, il y a des variations, mais globalement, une bonne progression régulière, comme on l'espérait, est observable. Dans le système britannique, note Milton (2009) , les progrès ne sont pas réguliers. L'apprentissage du vocabulaire dans certaines périodes d'enseignement peut être très limité et des plateaux de connaissances peuvent apparaître. Les objectifs de vocabulaire fixés par les organismes d'examen semblent ne pas être atteints. Bien sûr, en Grèce et en Hongrie, il existe des

environnements extrêmement favorables à l'apprentissage de l'EFL. Les parents et les apprenants veulent tous deux réussir en anglais, et la connaissance de l'anglais ouvre des voies d'avancement éducatif et professionnel. Au Royaume-Uni, le système est régulièrement remis en cause et les normes sont souvent dites en baisse. De plus, au Royaume-Uni, il n'existe aucun environnement favorable à l'apprentissage des langues étrangères, les langues ne sont pas valorisées. Il peut être difficile d'avoir un accès régulier et communicatif au français en dehors de la salle de classe.

Ces résultats confirment cette impression et commencent même à quantifier l'ampleur de l'apprentissage du vocabulaire et des progrès réalisés dans les écoles.

Cependant, ces études sont à relativement petite échelle et limitées à une ou deux écoles seulement. Aussi, quels que soient les progrès globaux dans ces écoles, il existe d'énormes variations, certains apprenants faisant progrès incroyablement bons, et d'autres peu ou pas du tout. Néanmoins, ces chiffres fournissent le type de données normatives qui permettent de fixer des normes, de suivre les progrès et d'effectuer des comparaisons, avec d'autres écoles ou avec d'autres systèmes.

Les progrès dans l'apprentissage du vocabulaire en Grèce et en Hongrie semblent remarquablement réguliers. En théorie, il peut y avoir des effets de plafond avec des tests de 5000 mots parmi les apprenants au sommet du système et ceux-ci peuvent masquer l'ampleur de l'apprentissage dans les meilleures classes.

4.6 Temps passé en classe et taux d'apprentissage

Les comparaisons entre systèmes éducatifs, voire les comparaisons de progrès d'une année à l'autre au sein d'un même système, sont souvent difficiles car le temps passé en classe, et donc l'opportunité d'apprendre, varie. Sans une sorte d'ajustement ou de correction pour le temps passé en classe et le volume d'entrée, alors les comparaisons peuvent devenir odieuses. (Milton, 2009, p. 86). Les apprenants au Royaume-Uni apparaissent souvent sous un jour particulièrement mauvais et il est facile de mal caractériser les apprenants. Un exemple de ceci serait de comparer les progrès annuels dans la connaissance du vocabulaire parmi les apprenants de français au Royaume-Uni avec d'autres apprenants. Les progrès parmi les apprenants britanniques sont beaucoup plus faibles que parmi les apprenants EFL en Grèce, et le niveau final de réalisation est plus petit. Mais le temps disponible pour l'apprentissage dans le système britannique est

également beaucoup plus petit. Les apprenants grecs reçoivent 100 à 125 heures de contact par an, tandis que le Royaume-Uni, pour une grande partie de leur éducation, reçoit la moitié de cette entrée. Encore une fois, Milton et Meara (1998) passent en revue un certain nombre d'articles qui font état des taux annuels de croissance du vocabulaire en Europe et au Japon. En règle générale, ces apprenants peuvent s'attendre à apprendre 500 à 600 mots par année d'études formelles. Par comparaison, les apprenants équivalents au Royaume-Uni ont été laissés languissant au pied du tableau qui a été produit (Milton & Meara, 1998, p. 74), car ils n'ont appris qu'environ 200 mots. Mais les apprenants britanniques avaient passé beaucoup moins de temps en classe que les Européens ou les Japonais avec lesquels ils étaient comparés. Les apprenants britanniques ne sont peut-être pas les mauvais apprenants que suggèrent ces comparaisons, mais le système éducatif dans lequel ils apprennent peut leur avoir refusé le temps et la possibilité d'apprendre, ce que d'autres pays offrent.

Une solution à la difficulté de comparaison, créée par le temps de classe très variable, consiste à examiner l'assimilation du vocabulaire par heure de contact. Il est possible de déduire la participation par heure ou la participation par classe à partir des chiffres des examens et du programme du ministère, qui ont été décrits précédemment. Milton a pris les informations que fournit le tableau 4-2 sur les programmes et systèmes scolaires de Hongrie, de Grèce et du Royaume-Uni pour calculer le type d'apprentissage par heure qui est, apparemment, attendu. Le système hongrois donne des objectifs annuels, tandis que pour les autres systèmes, nous n'avons que des objectifs finaux associés aux examens sur lesquels travailler.

Tableau 4-2 : Consommation de vocabulaire anticipée par heure en Hongrie, dans la frontière grecque et au Royaume-Uni (Milton, 2009, p. 87)

	3 ^{ème} année	4 ^{ème} année	5 ^{ème} année	6 ^{ème} année	7 ^{ème} année	8 ^{ème} année	Examen niveau B1	Examen niveau B2
Hongrie	6.3	0.9	3.6	1.8	3	6	4.5	
Grèce								5-6
Royaume Uni							2.8-4.3	

Les chiffres du programme hongrois suggèrent une consommation annuelle de mots par heure et ceux-ci peuvent varier de moins d'un mot par heure en 4e année, à plus de six mots par heure en 3e et 8e années. Les apprenants hongrois ont été beaucoup plus réguliers dans leur apprentissage du vocabulaire. En Grèce, le Syndicat des examens locaux de l'Université de Cambridge (UCLES) est inhabituellement précis en déclarant qu'environ 600 heures de contact seraient nécessaires pour les apprenants capables se préparant à passer l'examen FCE, ce qui impliquerait une absorption de vocabulaire d'environ 7,5 mots par heure de contact, y compris les mots peu fréquents en dehors de la plage de test X-Lex, et probablement cinq à six mots par heure de contact tels que mesurés par X-Lex. L'objectif britannique du GCSE d'environ 1 000 à 1 500 mots comme vocabulaire de base minimum suggère une utilisation plus modeste comprise entre 2,8 et 4,3 mots par heure de contact.

Ce sont des moyennes utiles, bien que très générales, et peuvent fournir un critère auquel n'importe quel cours ou apprenant pourrait être comparé. Comment ces chiffres se comparent-ils aux chiffres réels que les apprenants affichent réellement ? Milton et Meara (1998) ont ensuite examiné des articles dans lesquels l'apprentissage du vocabulaire est calculé et le nombre d'heures de cours est également spécifié, ce qui a permis de calculer l'absorption de vocabulaire par heure. Ces résultats ont été comparés aux taux d'utilisation du vocabulaire dans leur propre étude d'apprenants au Royaume-Uni, en Allemagne et en Grèce. Ils ont conclu que dans l'apprentissage normal en classe, suivi d'un programme d'études et où les apprenants ne sont pas sélectionnés pour leurs capacités, l'assimilation du vocabulaire pour une classe d'apprenants semble remarquablement régulière à environ trois ou quatre mots par heure de contact. Un résumé de leurs résultats est présenté dans le tableau 4-3.

Tableau 4-3: Vocabulaire appris par heure de contact

Apprenants	Langue étrangère	Consommation de vocabulaire par heure	Source
Hongrois	Anglais	5.4	Milton 2009
Grecs	Anglais	4.7	Milton 2009
Grecs	Anglais	4.4	Milton and Meara (1998)
Allemands	Anglais	4	Milton and Meara (1998)
Britanniques	Français	3.8-4.3	Milton and Meara (1998)
Grecs	Anglais	2.8	Vassiliu (1994)
Britanniques	Français	2.4	Milton 2009
Indiens	Anglais	1.7-3.3	Barnard (1961)
Indonésiens	Anglais	1.7-3.3	Quinn (1968)

Vu sous cet angle, les progrès des apprenants semblent généralement plus cohérents et les apprenants britanniques, en particulier, se comparent légèrement mieux avec les apprenants d'ailleurs. Les apprenants hongrois semblent apprendre le vocabulaire régulièrement sur six ans à un rythme approchant le maximum suggéré par leur programme. Les apprenants grecs, qui ont eu un peu plus que les 600 heures de contact suggérées comme minimum pour la préparation au FCE, semblent avoir acquis le vocabulaire de manière cohérente à un rythme approchant les cinq et six mots par heure de contact suggérés par le programme. Les apprenants britanniques ont appris le vocabulaire français en moyenne à un taux inférieur à l'objectif d'examen suggéré et, même après ajustement pour les différences de temps de contact, à un taux inférieur à celui des autres apprenants étudiés dans ce chapitre. Les apprenants grecs au niveau FCE sont une sélection des plus capables et ne comprennent que ceux qui avaient une chance raisonnable de réussir l'examen. Les apprenants qui n'étaient pas censés réussir, et qui auraient vraisemblablement appris le vocabulaire à un rythme plus lent, ont été exclus de la classe et de l'échantillon. Cela peut expliquer le taux d'assimilation relativement élevé du vocabulaire. Mais le groupe britannique a également subi une certaine forme de sélection, car les apprenants sont autorisés à abandonner l'étude du français l'année précédant les examens GCSE (niveau B1) et le groupe testé était, vraisemblablement, constitué d'apprenants les plus capables et les plus rapides.

Malgré ces considérations, le résultat de cette analyse confirme ce que Milton et Meara ont écrit en 1998, que l'absorption de vocabulaire par heure de classe semble remarquablement cohérente entre les classes et les systèmes. Et là où de grandes différences dans le niveau global de connaissances et de réussite apparaissent, entre les apprenants et les examens, cela semble être le produit des quantités très différentes de temps mis à disposition pour l'étude formelle. Le chiffre qu'ils suggèrent d'environ quatre mots appris par heure de contact semble être un critère extrêmement utile pour les enseignants et les apprenants. Il y aura des variations, bien sûr, et les bons apprenants apprendront plus vite tandis que d'autres apprendront plus lentement. Néanmoins, cela semble être une marque par rapport à laquelle ils peuvent évaluer si l'apprentissage semble normal. Cela suggère qu'avec du temps et de bonnes pratiques en classe, il semble possible de développer un vocabulaire important, suffisant pour la compétence même s'il ne s'agit pas d'une performance native dans une langue étrangère.

Mais même cette approche ne peut pas aplanir toutes les anomalies dans les données britanniques. Les plateaux, où très peu de croissance du vocabulaire semblait se produire, peuvent coïncider avec des réductions des heures de contact, mais les progrès semblent toujours incohérents. Les apprenants britanniques peuvent faire le genre de progrès que les autres apprenants affichent, mais ne le font pas avec la même régularité. Une comparaison des données grecques et britanniques extraites de la section précédente, converties en participation par heure de classe, le démontre et est illustrée à la figure 4-6.

Bien que les apprenants grecs de l'anglais terminent leurs sept années d'études avec le double du vocabulaire de langue étrangère que leurs homologues britanniques obtiennent, pendant quatre des sept années, les progrès sont très similaires lorsqu'ils sont mesurés en termes d'absorption par heure. Au cours des trois années restantes, les deuxième, troisième et quatrième années, les apprenants britanniques semblent faire peu ou pas de progrès en vocabulaire, moins de 50 nouveaux mots en troisième année, moins d'un mot par heure.

Les tailles de vocabulaire comme celle-ci peuvent-elles être comparées de manière significative dans deux langues ? Les langues diffèrent et il peut être possible de faire plus avec moins de mots dans une langue que dans une autre. L'article de Cobb et Horst (2004) sur la nécessité ou la faisabilité d'une liste de mots académiques en français suggère que le français et l'anglais diffèrent légèrement.

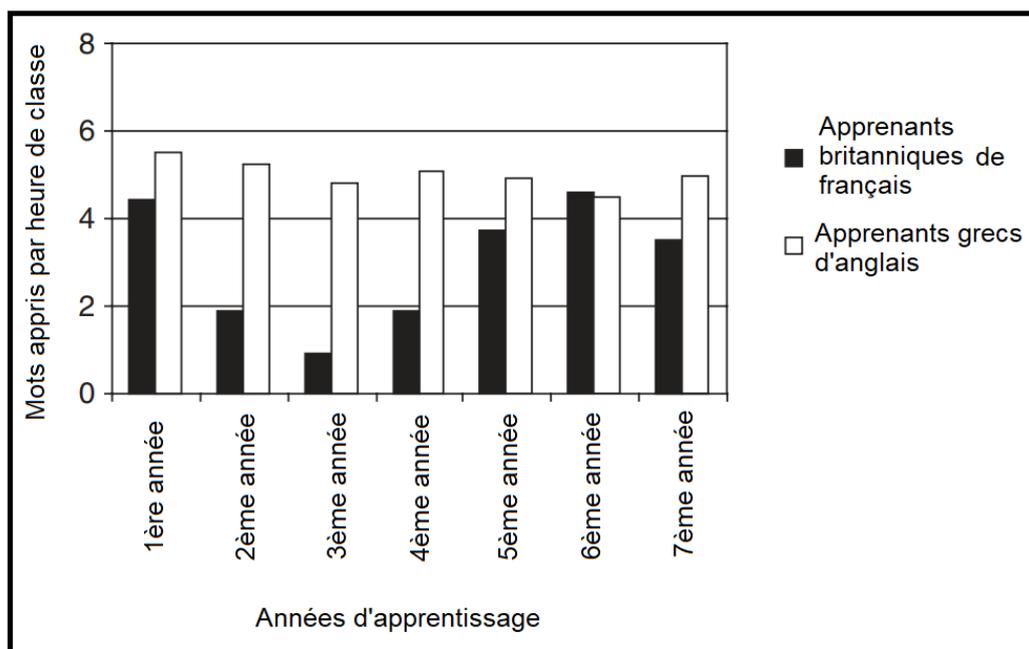


Fig. 4-6: Comparaison des mots appris par heure de classe au cours de chaque année d'études au Royaume-Uni et dans une école privée en Grèce (Milton, 2009, p. 89)

La comparaison de la couverture par les listes de fréquences française et anglaise au chapitre 3 confirme qu'il est possible d'obtenir une plus grande couverture avec moins de mots en français qu'en anglais. Mais, dans l'ensemble, le français et l'anglais sont très similaires, surtout aux niveaux les plus fréquents, il n'y a donc aucune raison majeure pour que ces deux ensembles de chiffres, basés uniquement sur le vocabulaire le plus fréquent dans les deux langues, ne soient pas similaires. Le fait que pendant quatre des sept années, les apprenants puissent apprendre au même rythme d'acquisition de vocabulaire tend à soutenir cette idée. Cela peut ne pas être le cas dans d'autres langues.

4.7 Conclusion

Les tests d'étendue du vocabulaire réceptif passif décrits dans ce chapitre ont permis d'obtenir des estimations crédibles de la connaissance du vocabulaire et des progrès des apprenants. Dans les exemples tirés de la Grèce et de la Hongrie, les résultats suggèrent que le vocabulaire anglais est appris de manière cohérente tout au long de l'enseignement et en quantités relativement prévisibles. Ces apprenants pris en groupe apprennent systématiquement quelque

chose comme quatre mots par heure de contact, ou légèrement mieux, sur des centaines d'heures en classe. Les tests utilisés ont été des tests de reconnaissance de la forme écrite des mots et cela suggère une bonne progression vers le volume de vocabulaire, plusieurs milliers de mots au moins, requis pour une bonne communication, y compris l'écriture et la lecture. Bien sûr, les échantillons sont encore petits, ce sont des instantanés d'une poignée d'écoles dans seulement trois pays, mais, néanmoins, il n'y a aucun signe du type de chargement de vocabulaire au début de l'apprentissage qui, selon Nation, pourrait être utile, ni de la vitesse croissante à laquelle l'apprentissage du vocabulaire pourrait se produire.

Les systèmes de test ont permis de comparer directement les résultats des apprenants d'autres langues, du français au Royaume-Uni, avec ceux des apprenants d'anglais. Le système britannique d'apprentissage des langues étrangères avait fait l'objet de certaines critiques vu que l'assimilation du vocabulaire semble plus lente qu'ailleurs, lente même par rapport aux objectifs relativement modestes fixés pour les apprenants, et il existe des plateaux inexplicables dans l'apprentissage où peu ou pas de progrès sont visibles.

Les observations de la croissance du vocabulaire et des progrès faites par Milton confirment également l'observation d'Alderson sur l'importance d'établir des données normatives dans ces domaines d'apprentissage. Ce n'est qu'en collectant des données de ce type que l'on peut déterminer dans quelle mesure la connaissance du vocabulaire est normale à n'importe quel stade de l'apprentissage et quels progrès sont possibles au cours de l'étude. Les données de ce type permettent de porter des jugements sur d'autres systèmes et apprenants : si les progrès semblent normaux ou s'ils sont meilleurs ou pires. Ils aident également à établir quels sont les niveaux de vocabulaire appropriés pour les apprenants dans l'éducation et quelles connaissances de vocabulaire sont nécessaires pour la réalisation générale ou pour réussir les examens.

CHAPITRE 5 :

TAILLE DU VOCABULAIRE ET COMPETENCES LINGUISTIQUES

5.1 Introduction

Il a été démontré que diverses mesures de la connaissance du vocabulaire sont en corrélation avec des facteurs tels que le temps passé à apprendre et le niveau de la classe. Cela démontre une caractéristique plus générale d'un apprentissage organisé, systématique et réussi d'une langue étrangère, à savoir que pendant la majeure partie du cours d'apprentissage, la connaissance du vocabulaire augmentera avec le niveau général de la langue. Les apprenants de niveau intermédiaire auront tendance à en savoir plus que ceux de niveau débutant, et les apprenants de niveau avancé auront tendance à en savoir plus que les apprenants intermédiaires. Milton (2009, p. 170) note que les tests de taille de vocabulaire standard, tels que le test de niveau de vocabulaire de Nation (1990) et le X_Lex de Meara et Milton (2003), se décrivent en fait comme des tests de niveau ou des tests de placement, et leur fonction est de permettre aux utilisateurs d'attribuer rapidement et facilement un apprenant à la bonne classe, ou d'évaluer si un apprenant pourrait faire face à une activité telle que l'étude académique dans la langue étrangère, sur la base du volume de vocabulaire que les apprenants connaissent et peuvent utiliser.

Néanmoins, les utilisateurs de ces tests, enseignants et apprenants, expriment parfois leur surprise et leur doute quant au fait que cette seule caractéristique des connaissances linguistiques, le vocabulaire, puisse être un bon indicateur général de la compétence linguistique globale. Cela semble être un problème de vocabulaire, d'autant plus que les utilisateurs sont souvent beaucoup plus à l'aise avec les tests de classement basés sur d'autres caractéristiques de la langue, tout aussi restreintes, telles que la grammaire. En principe, les connaissances linguistiques et les compétences linguistiques communicatives sont des qualités distinctes, et il devrait être possible pour les apprenants de varier, au moins dans une certaine mesure, dans leur capacité à utiliser leurs connaissances dans leur quête de compétences linguistiques et de communication. Le niveau de langue ne concerne pas seulement les connaissances linguistiques, qu'elles soient grammaticales ou lexicales, mais concerne l'utilisation de ces connaissances de manière communicative et la manière dont un apprenant peut être performant dans diverses circonstances.

Il faut donc s'attendre à ce que les mesures de la connaissance du vocabulaire soient liées à des suites d'examens ou à des hiérarchies formelles de niveau de langue et de performances, telles que les examens *Cambridge First Certificate and Proficiency* et le CECR.

5.2 Les quatre compétences et la taille du vocabulaire

La connaissance du vocabulaire est liée aux performances linguistiques globales et des organismes d'examen comme Cambridge lient explicitement la connaissance du vocabulaire à leurs critères d'évaluation, y compris l'écriture et l'expression orale mais des preuves suggèrent que la taille du vocabulaire est bien corrélée avec certaines compétences telles que l'écriture et beaucoup moins bien avec d'autres compétences. Milton (2009) attribue cela au caractère empirique et non théorique de l'idée qui avance que le vocabulaire joue un rôle important dans la détermination du niveau de langue. Elle repose sur le constat que les mesures de la taille du vocabulaire sont bien corrélées avec les niveaux progressifs d'aptitude dans les examens impliquant l'écrit en particulier.

Stæhr (2008), par exemple, examine la relation entre les notes d'examen sur les épreuves d'écoute, de lecture et d'écriture, et la taille du vocabulaire des personnes testées. Les résultats qu'il a obtenus, résumés dans le tableau 5-1, suggèrent un lien entre la connaissance du vocabulaire et les trois éléments de la performance à l'examen, et un lien fort avec la lecture en particulier.

Tableau 5-1; Corrélations de Spearman entre les scores de taille de vocabulaire et les scores de lecture, d'écoute et d'écriture (n=88) (Stæhr, 2008, p. 144)

	Ecoute	Lecture	Ecriture
Taille du vocabulaire	0.69**	0.83**	0.73**

Stæhr conclut que jusqu'à 72 % de la variance dans la capacité d'obtenir une note moyenne ou supérieure au test de lecture peut s'expliquer par la taille du vocabulaire. Le vocabulaire peut être moins important que cela dans l'écriture et l'écoute, mais l'apport de la connaissance du vocabulaire apparaît toujours important. Il conclut également que la connaissance des 2000 mots les plus fréquents en anglais représente un seuil à franchir si les apprenants veulent obtenir un

score moyen ou supérieur à ces tests. Il est à noter que, encore une fois, l'importance d'apprendre ces 2000 mots les plus fréquents en anglais est réapparue. Ces résultats renforcent la conclusion selon laquelle le vocabulaire est d'une importance vitale pour atteindre des niveaux de performance élevés et obtenir des notes élevées aux examens.

Il est possible que la connaissance du vocabulaire fonctionne différemment dans différentes compétences et cela peut être lié aux différences de couverture dans les corpus écrits et parlés. Il peut être possible de parler beaucoup plus couramment avec moins de ressources de vocabulaire dans la parole, obtenant ainsi des notes plus élevées que celles nécessaires à l'écrit. Les tests de taille et d'étendue du vocabulaire se concentrent souvent sur un vocabulaire relativement moins fréquent, ce qui est beaucoup moins utile dans la communication orale.

Milton (2009) évoque une deuxième raison pour laquelle les recherches à ce jour n'ont révélé qu'un lien fort entre le vocabulaire et les compétences communicatives impliquant l'écrit. Cela peut être le produit des tests utilisés dans ces quelques études publiées, qui ont tendance à utiliser des tests sous forme écrite de mots plutôt que de s'appuyer sur la connaissance du vocabulaire phonologique de l'apprenant.

Une étude de Milton a révélé qu'il existe des corrélations statistiquement significatives entre le test de taille de vocabulaire écrit et les sous-tests de l'IELTS (*International English Language Testing System*), qui obligent les apprenants à lire en anglais. Les résultats de l'étude de Milton sont résumés dans le tableau 5-2.

Tableau 5-2 : Corrélations entre les scores de taille de vocabulaire et les scores IELTS

	A-Lex	Lire	Ecouter	Ecrire	Parler	Tout
X-Lex	0.456*	0.699**	0.479**	0.761**	0.347	0.683**
A-Lex		0.217	0.676**	0.441*	0.713**	0.546**

Source : Milton et al.

Note. **La corrélation est significative au seuil de 0,01

*La corrélation est significative au seuil de 0,05

La connaissance du vocabulaire orthographique semble être un contributeur important à la capacité de gérer les exigences du test d'écriture en particulier, mais prédit également les scores aux tests de lecture et d'écoute. Des corrélations statistiquement significatives sont également apparues entre les scores de taille de vocabulaire phonologique et les sous-compétences impliquant la communication auditive.

Cela suggère une conclusion qui, rétrospectivement, n'est pas surprenante, à savoir que les tests de vocabulaire basés sur l'orthographe prédisent le mieux les éléments de ces tests qui reposent sur la capacité de gérer un texte écrit, tels que des exercices d'écriture et de compréhension en lecture. La connaissance du vocabulaire phonétique est le meilleur prédicteur de la capacité à traiter les éléments de test impliquant la capacité d'expression orale, tels que les tests d'expression orale.

Néanmoins, le rapport au vocabulaire est fort et la relation entre les compétences en lecture et en écriture et la taille du vocabulaire semble relativement simple : plus vous connaissez de mots, jusqu'à plusieurs milliers, mieux vous êtes susceptible de bien lire ou écrire. En règle générale, la taille du vocabulaire phonologique et les performances dans les tests d'expression orale et d'écoute sont moyennement bien corrélées.

Nonobstant les autres facteurs qui contribueront à obtenir des notes linguistiques élevées, la connaissance du vocabulaire semble donc être peut-être le plus grand élément de connaissance et de compétence dans l'ensemble des performances linguistiques. Cette analyse suggère également que la relation entre la taille du vocabulaire et la capacité d'effectuer des tâches d'écriture et de lecture est essentiellement linéaire : plus vous connaissez de mots, au moins jusqu'à 10 000 ou plus, mieux vous êtes susceptible de réussir. Nation (2006, p. 59) suggère que 8000 ou 9000 mots sont nécessaires pour la compréhension en lecture. Les performances sur des tâches orales telles que le test d'expression orale sont également liées au vocabulaire, mais la relation peut ne pas être linéaire : il existe une limite au-delà de laquelle l'augmentation du nombre de mots peut ne pas vous aider à améliorer vos performances.

Comme pour tant de travail sur le vocabulaire, l'analyse de la façon dont le vocabulaire et les quatre compétences interagissent a dû s'appuyer presque exclusivement sur l'étude des apprenants de l'anglais comme langue étrangère (EFL). Ce serait une grande surprise, cependant, si l'importance du vocabulaire pour la performance aux examens et les progrès dans d'autres langues devait différer du rôle qu'il joue en anglais.

5.3 Taille du vocabulaire et performance à l'examen

Dans les compétences liées à l'écriture en particulier et dans les notes globales, plus la connaissance du vocabulaire d'un apprenant est grande, meilleure est la note qu'il est susceptible d'obtenir. Cela implique que des niveaux particuliers de connaissance du vocabulaire seront associés au fait de passer et de réussir des examens particuliers. Alors, qu'en est-il pour ces niveaux de connaissance du vocabulaire.

Meara et Milton (2003) relient explicitement les scores X-Lex (max 5000) à une variété de niveaux d'examen et de scores d'examen. Selon Milton, les suggestions selon lesquelles des volumes assez substantiels de connaissances du vocabulaire sont nécessaires pour progresser au-delà des niveaux élémentaires de performance linguistique, et pour passer d'une capacité très limitée dans des environnements limités et prévisibles à une position où la variété et l'imprévisibilité du langage normal peut être surmontée. La connaissance de plus de 3 500 mots, soit 70 % du corpus de test de 5 000 mots, est associée aux étudiants qui peuvent passer et réussir l'examen du *Cambridge First Certificate in English* (FCE). Les étudiants qui passent des examens de niveau avancé devraient probablement reconnaître plus de 4 500, ou 90 % ou plus, de ce corpus. Milton (2009) conclut que pour passer et réussir le Cambridge FCE, les apprenants devront probablement connaître environ 3 500 des 5 000 mots les plus fréquents en anglais. Pour réussir Cambridge Proficiency, ils auront probablement besoin de 4 500 mots sur 5 000. Et pour obtenir une note A à *Cambridge Proficiency*, la connaissance d'environ 8 500 mots sur les 10 000 les plus fréquents est nécessaire.

Au Royaume-Uni, la connaissance du vocabulaire des apprenants qui passent les examens GCSE de 16 ans et les examens de niveau « A » de 18 ans en français ont été mesurées à l'aide d'une version française de X-Lex (Meara & Milton, 2003). Les résultats sont donnés dans le tableau 5-3.

Tableau 5-3 : Détails de la taille moyenne du vocabulaire aux niveaux GCSE et « A »

	Nombre	Moyenne	Max	Min	Ecart-type
GCSE	49	852	1800	0	440
Niveau A	69	1930	3100	650	475

Source : Milton (2006, p. 192)

Il existe une grande variation ici, mais l'attente générale est claire, que la connaissance d'environ 1000 mots de français permettra à un apprenant d'obtenir un laissez-passer au GCSE et 2000 gagnera un laissez-passer au niveau « A ». Une analyse plus approfondie dans le même article suggère que la taille du vocabulaire peut être liée au niveau de français « A » et que, comme on pouvait s'y attendre avec ce niveau de connaissance, une meilleure connaissance du vocabulaire est associée à un niveau supérieur. Encore une fois, ceux-ci peuvent être représentés graphiquement et les résultats de 69 apprenants qui ont obtenu le niveau « A » sont présentés à la figure 5-1.

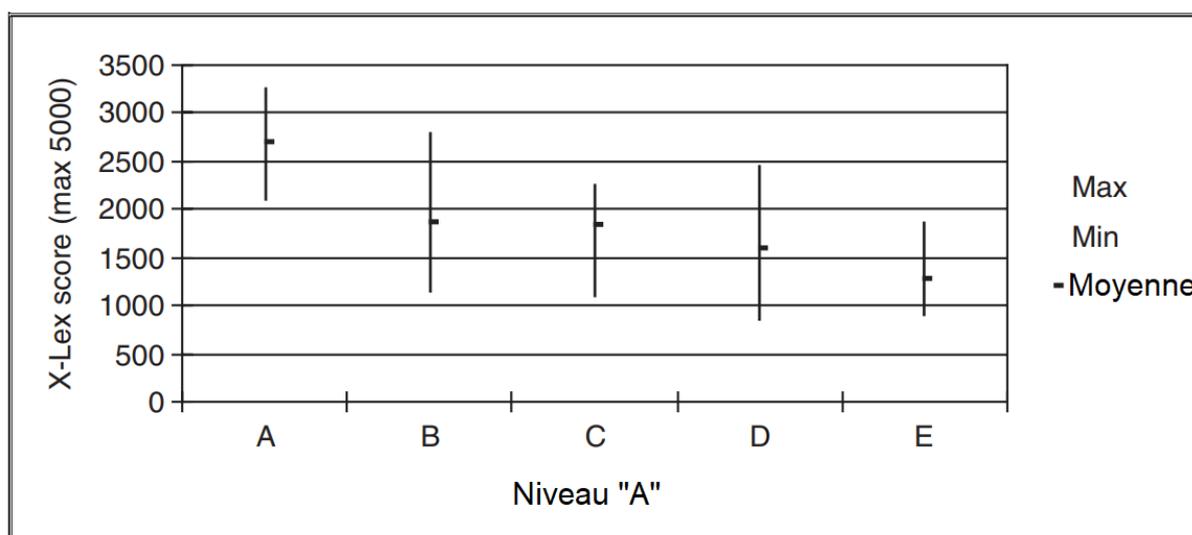


Fig. 5-1 : Taille du vocabulaire et note de niveau « A » (Milton, 2006b : 194)

Malgré cela, il existe encore un chevauchement considérable des scores de vocabulaire. Cela est peut-être le produit de la variation du degré de compétence avec lequel les apprenants peuvent rassembler leurs connaissances face au type de tâches de communication et académiques que présentent les examens formels. De même, cela pourrait aussi être le résultat du type d'imprécision que les critères évalués subjectivement introduisent dans le système de test.

Richards et al. (2008) étudient plus en détail cette relation parmi les apprenants du français langue étrangère. Ils comparent non seulement les notes globales du test de taille de vocabulaire X-Lex aux notes qui émergent au GCSE et sont prédites au niveau « AS », mais examinent également comment chaque bande de 1000 mots interagit avec ces notes. Leurs résultats sont résumés dans le tableau 5-4.

Tableau 5-4: Coefficients de classement de Spearman et signification statistique des corrélations entre les scores X-Lex et le GCSE et les notes AS prévues

	Note GCSE	Note AS prévu
Score total ajusté	0.43 (0.020)	0.43 (0.020)
Score brut sur les 1000 premiers	0.55 (0.004)	0.35 (0.049)
Score brut au deuxième 1000	0.29 (ns)	0.43 (0.021)
Score brut sur le troisième 1000	0.32 (ns)	0.35 (0.050)
Score brut au quatrième 1000	0.32 (ns)	0.58 (0.002)
Score brut au cinquième 1000	0.23 (ns)	0.49 (0.009)

Source: Richards et al. (2008, p. 204)

Superficiellement, il semble que la taille du vocabulaire prédit les notes dans les deux examens aussi bien, mais la rupture de la relation avec les tranches de 1 000 mots suggère que la relation est plus subtile. La note GCSE est prédite par la connaissance des 1000 premiers mots uniquement. Avec une taille de vocabulaire si petite parmi les apprenants du GCSE, généralement moins de 1000 mots, et avec des connaissances concentrées dans la première bande de fréquence, ce résultat n'est peut-être pas surprenant. Les notes AS sont prédites par toutes les bandes de fréquences, mais la relation la plus forte est avec les quatrième et cinquième bandes de fréquences de 1000 mots. Il semble que la connaissance d'un vocabulaire relativement peu fréquent soit une condition pour réussir l'examen AS.

Richards et al. (2008) parviennent également à démontrer des corrélations très modestes, mais néanmoins statistiquement significatives entre D, une mesure de la richesse lexicale, et les notes de niveau GCSE et « A ». Cette étude montre tout simplement que mieux un apprenant connaît les 1000 mots les plus fréquents en français, meilleur sera son score en français GCSE au Royaume-Uni.

5.4 Taille du vocabulaire et le CECR

La taille du vocabulaire étant liée au niveau d'examens particuliers, et comme ces examens sont liés au cadre du CECR, il est possible de lier la taille du vocabulaire au CECR. Meara et Milton (2003, p. 8) ont intégré des scores de taille de vocabulaire EFL, basés sur les résultats de tests utilisant X-Lex avec un plafond de 5000 mots. Les guides de vocabulaire qu'ils produisent semblent robustes par rapport aux données des apprenants passant les examens EFL en Grèce et en Hongrie. Les scores de vocabulaire à chaque niveau du CECR sont présentés dans le tableau 5-5.

Tableau 5-5 : Scores moyens de taille de vocabulaire EFL et CECR

Niveau CECR	Taille de la liste de mots	X-Lex	EFL Greece	EFL Hungry
A1		<1500	1477	
A2	1000	1500-2055	2156	
B1	2000	2500-3250	3264	3136
B2		3250-3750	3305	3668
C1		3750-4500	3691	4340
C2		4500-5000	4068	

L'EFL et le français peuvent être assez similaires, aux niveaux supérieurs. L'absence en français d'un équivalent à la liste de mots académiques en anglais suggère que les apprenants de français n'ont peut-être pas besoin d'ajouter du vocabulaire de la même manière que les apprenants d'EFL afin d'augmenter leur compétence. Les locuteurs de français langue étrangère peuvent utiliser leur vocabulaire français plus fréquent dans les environnements formels et académiques d'une manière impossible en anglais. Des estimations de la taille du vocabulaire, utilisant une version française de X-Lex, ont été faites parmi les apprenants au Royaume-Uni (Cité dans Milton, 2009) et en Grèce et en Espagne (Milton & Alexiou, 2009), et ont été liées par le niveau d'examen et les supports de cours au CECR. Ces chiffres sont présentés dans le tableau 5-6.

Tableau 5-6 : Scores de taille du vocabulaire français langue étrangère et CECR (Milton, 2009, p. 187)

Niveau CECR	Taille de la liste de mots	Français au Royaume-Uni	Français en Grèce	Français en Espagne
A1			1160	894
A2	1000	850	1650	1700
B1	2000	850	2422	2194
B2		1920	2630	2450
C1			3212	2675
C2		3300	3525	3721

Milton (2009) note que les chiffres de la Grèce et de l'Espagne en particulier racontent une histoire convaincante avec une taille de vocabulaire augmentant à chaque niveau successif du CECR. Les chiffres grecs et espagnols sont également notablement similaires et, aux niveaux A2 et B1, assez différents des chiffres des apprenants de français au Royaume-Uni. Le fait que A2 et B1 partagent le même score est un artefact du système d'examen britannique, qui permet aux apprenants de passer l'examen GCSE à 16 ans à l'un de ces niveaux et la collecte de données à ce jour n'a pas fait de distinction entre les deux. Au Royaume-Uni, cependant, le classement des examens nationaux aux niveaux B1 et B2 dans le CECR a été critiqué comme étant trompeur (Milton, 2006b) et sur la base de ces chiffres, il est possible de comprendre pourquoi.

Les tailles de vocabulaire français sont globalement similaires au niveau A1, A2 et, dans une moindre mesure, au niveau B1 aux scores EFL aux mêmes niveaux du CECR. Mais, par la suite, le français semble exiger moins de vocabulaire pour progresser que l'anglais. Les apprenants de français langue étrangère auront probablement besoin de connaître 2000 ou plus des 5000 mots les plus fréquents pour passer du CECR A2 à B1. Ils auront probablement besoin de 3000 mots ou plus pour passer du CECR B2 au C1. (Milton, 2009, p. 188)

Il existe quelques chiffres disponibles pour les apprenants du grec comme langue étrangère, apprenant en Grèce, et ceux-ci sont présentés dans le tableau 5-7.

Tableau 5-7 : Scores de taille du vocabulaire grec langue étrangère et CECR

Niveau CECR	Grec comme langue étrangère
A1	1492
A2	2238
B1	3338
B2	4013
C1	
C2	

Source: Milton et Alexiou (2009)

Encore une fois, ces chiffres suggèrent que le vocabulaire augmente avec le niveau du CECR. Les apprenants, comme en EFL, semblent avoir besoin d'environ 3000 mots, 60% du corpus de test, avant de passer du niveau élémentaire (A1 et A2) à intermédiaire (B1 et B2). Il n'est pas clair à partir de ces données combien de vocabulaire est requis pour les utilisateurs de niveau avancé (C1 et C2) en grec, mais il semblerait qu'il y ait plus de 4000 mots ou 80% du corpus de test. Il semblerait que les apprenants du grec langue étrangère soient plus exigeants en termes de vocabulaire que ce n'est le cas pour l'anglais ou surtout pour le français.

Milton (2009) soutient qu'en anglais et en français, il semble y avoir un certain accord quant à la quantité de vocabulaire requise aux divers niveaux du CECR. Néanmoins, la disparité entre les différentes langues examinées est assez nette. Elle s'éclaire lorsque les scores moyens obtenus dans tous les pays et pour toutes les langues sont regroupés dans un même tableau, comme dans le tableau 5-8.

Tableau 5-8 : Synthèse des scores moyens pour chaque niveau du CECR dans trois langues étrangères

Niveau CECR	Français au Royaume Uni	Français en Espagne	Français en Grèce	EFL en Grèce	EFL en Hongrie	Grec en Grèce
A1		1160	894	1477		1492
A2	850	1650	1700	2156		2238
B1	850	2422	2194	3264	3136	3338
B2	1920	2630	2450	3305	3668	4013
C1		3212	2675	3691	4340	
C2	3300	3525	3721	4068		

Source (Milton, 2009)

5.5 Conclusion

Les données présentées dans ce chapitre suggèrent très fortement que la taille du vocabulaire peut être liée aux niveaux d'examen et au CECR d'une manière plausible. En fait, la taille du vocabulaire apparaît très fortement liée à ces niveaux. En tant que tel, il semble que cela puisse être un moyen rapide et utile d'évaluer le niveau global de connaissance et de compétence d'un apprenant d'une langue étrangère, donnant du crédit à des tests de classement tels que le test de niveau de vocabulaire de Nation (1990) et le X-Lex de Meara et Milton. (2003). Il ne semble pas y avoir de bonne raison pour que les écoles, les apprenants et les commissions d'examen n'utilisent pas davantage ce type d'informations.

La taille du vocabulaire peut aider à suggérer des niveaux beaucoup plus appropriés pour différents examens. Les estimations de la taille du vocabulaire peuvent manifestement être très utiles pour traiter les systèmes d'examen et les grands groupes d'apprenants. La relation entre les groupes d'apprenants et la hiérarchie des examens et des niveaux est relativement prévisible.

CHAPITRE 6 :

EXPOSITION DU VOCABULAIRE DANS LES MANUELS SCOLAIRES

6.1 Introduction

Comme le souligne O'Dell (1997, p. 264), les auteurs de manuels de cours à succès ont suivi une approche plus éclectique de la sélection de vocabulaire que ne le suggère la théorie. Ce qui ressort du matériel didactique est un compromis tiré de priorités concurrentes dans le programme et de l'aspect pratique pur de la rédaction de matériel attrayant et motivant. Aussi souhaitable qu'il soit d'introduire un bon volume du vocabulaire le plus fréquent dans les matériaux pour débutants, ces matériaux doivent également avoir un contenu thématique. Les listes de fréquences ne sont pas commodément organisées par thème ou idée et si un texte d'enseignement doit avoir une certaine cohérence, son vocabulaire devra être sélectionné parmi les niveaux de fréquence. Le manuel et la langue de l'enseignant contiendraient donc une combinaison de vocabulaire fréquent et structurel et de vocabulaire thématique peu fréquent. Ce matériel devra être équilibré avec d'autres éléments du programme, forçant l'introduction d'un nouveau lexique à être progressive et limitée.

Le volume de vocabulaire introduit chaque année varierait de 562 à près du double du chiffre qu'à préconisé Scholfield et qui avoisine 1065 éléments. Il relève surtout du nombre d'heures d'apprentissage. Le taux d'exposition du vocabulaire aux apprenants semble raisonnable, et même assez constant, s'il se situe dans la partie supérieure de la fourchette suggérée par Gairns et Redmond. Tschichold (cité dans Milton, 2009) note que les glossaires peuvent inclure un volume substantiel de vocabulaire qui n'est pas réellement présent dans le texte du manuel de cours lui-même.

Les manuels modernes varient dans les volumes de vocabulaire qu'ils choisissent d'introduire et il semble qu'ils varient également dans le choix du vocabulaire. Ce constat ne se limite pas aux cours d'EFL, mais s'étend au français langue étrangère. Il semble que les énormes différences notées par Milton et Benn (1933) il y a près de 80 ans existent encore entre les manuels de cours destinés aux mêmes apprenants à un même niveau.

6.2 Sélection du vocabulaire dans les manuels

Milton et Vassiliu (2000, p. 448) examinent le contenu de trois manuels pour débutants et comparent leur choix de vocabulaire. Ils construisent un diagramme de Venn pour illustrer les volumes de lexique que ces cours partagent, comme le montre la figure 6-1.

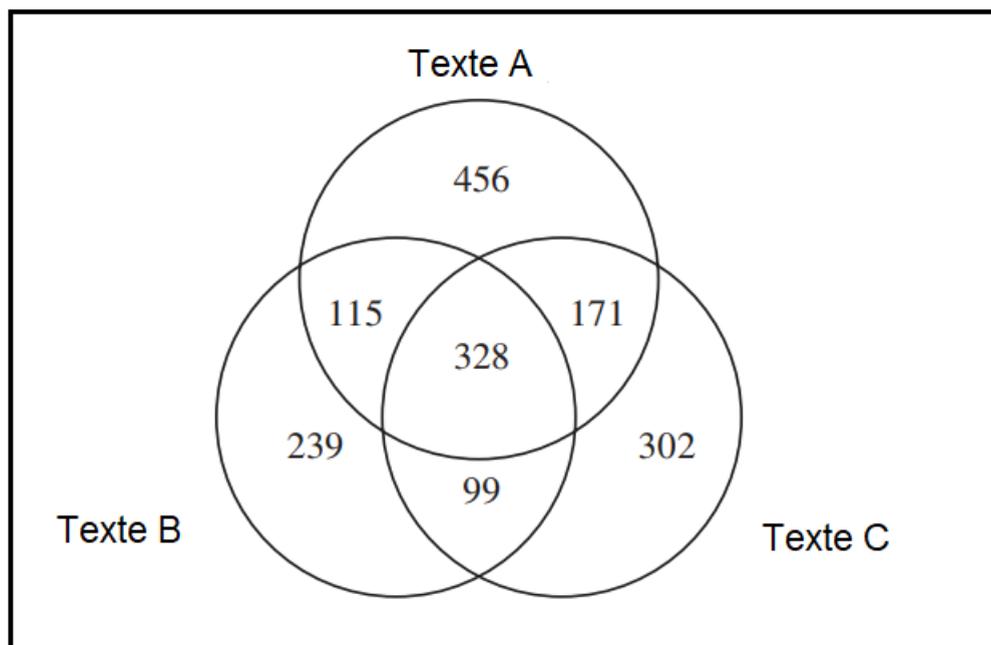


Fig. 6-1 : Vocabulaire commun aux trois manuels EFL pour débutants (Milton & Vassiliu, 2000, p. 448)

Le manuel A contient 909 items non lemmatisés, le manuel B 964 et le manuel C 1175 items. Ces chiffres confirment le type de chargement de vocabulaire que note Scholfield, et le soupçon que les apprenants EFL sont exposés à beaucoup plus de vocabulaire que nous ne le pensions est confirmé dans ce cas, car il est connu que ces textes combinés comprenaient un cours pour débutants de 100 heures. Milton et Vassiliu nettoient le corpus pour supprimer les noms de personnes et autres éléments non pertinents pour l'étude. Sur les 1710 items restants, seuls 328 sont communs aux trois textes.

Le nombre d'éléments de vocabulaire communs à tous les livres est très faible. Ce qui rend ce résultat surprenant, c'est l'accord général selon lequel un vocabulaire très fréquent est très important et devrait figurer en bonne place dans les manuels de cours. Cela devrait inclure la structure et les mots fonctionnels tels que les propositions et le verbe être, ce qui serait difficile

à éviter, pourrait-on penser. Si les rédacteurs de livres de cours essaient vraiment de choisir parmi ce matériel très fréquent, on peut s'attendre à des éléments plus courants. Mais si les écrivains sont si variés, quelle part de ce matériel très important et très fréquent y a-t-il dans les manuels ?

Tschichold (cité dans Milton, 2009) compare le corpus des manuels de cours de français langue étrangère aux niveaux 1 et 2 de français fondamental pour évaluer si les apprenants se voient refuser l'accès au vocabulaire structurel et fonctionnel très fréquent. Le corpus de quatre livres contient 3341 lemmes et a à peu près la même taille que les listes fondamentales en français, soit environ 3500 mots. La comparaison est résumée à la figure 6-2.

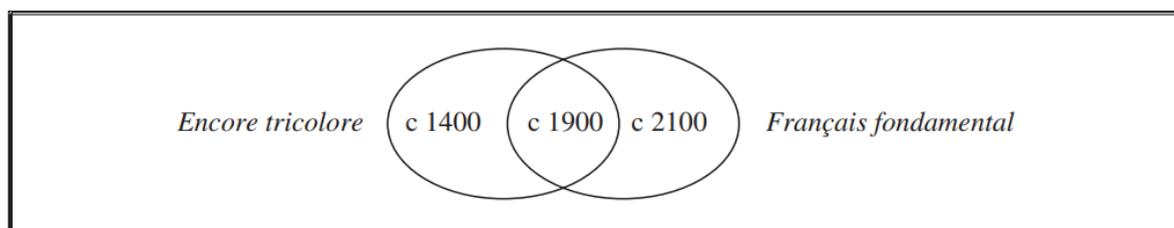


Fig. 6-2: Vocabulaire commun entre *Encore Tricolore* et français fondamental (Tschichold, cité dans Milton, 2009)

Plus de la moitié du contenu fondamental français est contenu dans le matériel didactique *Encore Tricolore*. Ce que serait un chiffre idéal n'est pas absolument clair, mais cela semble raisonnable. Français fondamental est une liste intelligemment construite contenant des éléments très fréquents qui sont essentiels pour la communication et le progrès en français.

Au Royaume-Uni, le problème de la sélection du vocabulaire dans les programmes de langues étrangères semble devenir de plus en plus aigu. *Encore Tricolore* est maintenant un manuel assez ancien et les manuels de cours de langues étrangères plus récents contiennent globalement moins de vocabulaire. (Häcker, 2008).

Milton et Vassiliu analysent les trois manuels de la figure 6-2, qui constituent un cours de première année pour les débutants en EFL et lemmatisent leur corpus afin de faire un calcul des mots dans les bandes de fréquences les plus élevées. Les 1710 items non lemmatisés ont été réduits à 1396 items lemmatisés par ce processus et les lemmes qui ont émergé ont été comparés aux listes de vocabulaire de Nation (1984). Les résultats sont présentés dans le tableau 6.1.

Tableau 6-1 : Contenu du vocabulaire du texte pour débutant par plage de fréquences

Niveau de Nation	Niveau 0	Niveau 1	Niveau 2	Niveau 3+
% Lemmes	8	40	6	46
Nombre de lemmes	105	558	84	649

Source: Milton and Vassiliu (2000: 449)

Selon Milton (2009), plusieurs choses ressortent de cette analyse. L'une est que, à l'instar du matériel français de Tschichold, des volumes à peu près égaux de vocabulaire très fréquent et structurel et un vocabulaire thématique moins fréquent sont introduits, malgré les pressions dues à la nécessité d'un contenu thématique et de couvrir d'autres éléments du programme, et malgré le fait que ce matériel est conçu pour seulement deux ans de travail au lieu de quatre. Au total, 663 des 1150 mots de niveau 0 et de niveau 1 sont introduits. Mais le besoin d'un contenu thématique est évident dans l'inclusion de 649 mots en dehors des niveaux les plus fréquents, les niveaux 0, 1 et 2. Environ 46% du programme de vocabulaire est donc composé de vocabulaire peu fréquent et cela peut paraître surprenant. La proportion de mots dans la seconde bande de 1000 vocabulaire apparaît faible. Seuls 84 de ces mots sont introduits, soit environ 6 % de la charge totale de vocabulaire. Dans une revue de deux textes différents, Vassiliu (cité dans Milton, 2009) a noté la même caractéristique. La préoccupation de Häcker selon laquelle les manuels de cours de langues étrangères au Royaume-Uni contiennent une quantité excessive d'éléments peu fréquents semble déplacée, bien que son point de vue selon lequel la qualité de ces mots, de longues listes dans des domaines limités, peut être correcte et le vocabulaire enseigné peut être moins largement utile pour la communication en conséquence. A priori, nous estimons de notre part que les manuels scolaires du cycle moyen en Algérie, contiennent excessivement d'éléments moins fréquents.

Un manuel efficace va probablement introduire un vocabulaire fréquent et peu fréquent en quantités à peu près égales. Il sera probablement thématiquement très diversifié. Il est difficile de dire, en l'absence d'informations sur l'utilisation des items, si ces proportions sont le bon équilibre pour permettre aux apprenants de faire de bons progrès linguistiques ou non.

6.3 Apprentissage du vocabulaire à partir de l'input en classe

Les résultats des expériences menées par Vassiliu suggèrent que présenter aux apprenants moins de vocabulaire conduit à une plus grande proportion du vocabulaire cible appris. Au fur et à mesure que le vocabulaire cible diminue, le nombre moyen de mots appris, exprimé en pourcentage du vocabulaire cible, augmente. Mais cela ne semble pas aider les apprenants en termes de quantité de vocabulaire appris. Les scores moyens de vocabulaire diminuent à mesure que les cibles de vocabulaire deviennent plus petites. Même si les apprenants apprennent proportionnellement plus de vocabulaire qui leur est présenté, ils finissent, en moyenne, par apprendre moins de vocabulaire.

La conclusion à tirer de ces résultats est que le choix du vocabulaire et le chargement des manuels sont loin d'être une préoccupation immatérielle dans l'apprentissage des langues ni ayant peu d'impact sur l'apprentissage. Elle semble avoir une incidence très directe sur le vocabulaire appris et sur la quantité de vocabulaire. Pour que cela se produise, les manuels devraient continuer à sélectionner et classer très soigneusement leur matériel et continuer à ajouter des quantités substantielles de vocabulaire peu fréquent. Cela devrait signifier que les manuels bien notés et séquencés incluent un vocabulaire de moins en moins fréquent ; qu'ils deviendront de plus en plus sophistiqués lexicalement.

Milton se demande si les très faibles taux d'apprentissage du vocabulaire parmi les apprenants de français au Royaume-Uni ne sont pas liés au vocabulaire qui leur est fourni dans le manuel, qui peut manquer d'une ampleur et d'une richesse suffisantes pour permettre aux apprenants de progresser. Milton pense que ces apprenants sont exposés à un volume de vocabulaire relativement modeste dans *Encore Tricolore* : environ 3 300 types lemmatisés au cours des quatre premières années d'étude. Le volume de vocabulaire très fréquent apparaît également faible avec seulement quelque 1900 des 3500 éléments de vocabulaire en français. L'étude de Milton (2006b) sur les apprenants qui utilisent ce livre a rapporté que les apprenants après quatre ans avaient une taille de vocabulaire moyenne de 592 mots lemmatisés

6.4 Recyclage et acquisition du vocabulaire dans les manuels

C'est un axiome d'un bon enseignement que le nouveau matériel doit être recyclé et répété s'il doit être appris de manière satisfaisante. L'enseignement du vocabulaire n'est pas différent, et les écrivains, pour que leurs livres de cours soient bons, devront envisager de recycler leur

vocabulaire lexical. Il est probablement impossible d'éviter de répéter beaucoup de vocabulaire fonctionnel et structurel, ce qui est essentiel pour un langage bien construit, c'est donc beaucoup moins un problème pour les écrivains. Nation suggère (1993) que peu de rédacteurs de manuels recyclent systématiquement le vocabulaire et, avec les grands volumes de vocabulaire présentés dans les manuels de cours, il peut même ne pas être possible de recycler chaque élément systématiquement. Néanmoins, les preuves suggèrent que cela peut être un élément important de la réussite de l'apprentissage.

Vassiliu (cité dans Milton, 2009) a étudié cela et réexaminé les données de ses trois manuels de cours et a testé l'assimilation du vocabulaire de 47 apprenants utilisant ce matériel. Il divise le vocabulaire contenu dans les manuels qu'il utilise en huit bandes selon la fréquence et teste 15 mots de chaque bande en utilisant le format Oui/Non précédemment décrit.

En gros, les résultats sont ceux que l'on peut attendre. Plus un élément est répété dans le manuel, et vraisemblablement plus les apprenants y sont exposés, plus il a de chances d'être appris. Mais même une répétition très élevée ne garantit pas l'apprentissage. Un apprenant de l'échantillon de Vassiliu n'a pas reconnu le mot *is* même s'il a été répété 1355 fois dans le texte. De même, une faible répétition dans le manuel de cours ne garantit pas qu'un mot ne sera pas appris. Une proportion étonnamment élevée de mots répétés une ou deux fois seulement a été apprise, malgré l'absence de recyclage. Il faut noter qu'une grande partie des types non lemmatisés de ce texte ne sont que rarement répétés. Soixante-quatre pour cent des types sont répétés cinq fois ou moins.

Plusieurs facteurs peuvent être en jeu ici. Il y a plus à apprendre que le contenu du livre de cours. Un facteur peut être l'effet de la présentation orale et du recyclage qui devrait, au moins en principe, compléter le contenu du manuel. Vassiliu (cité dans Milton, 2009) note que certains mots peu répétés dans le manuel, tels que *welcome* et *door*, étaient généralement connus des élèves, peut-être parce qu'ils faisaient partie d'une routine orale utilisée par l'enseignant. D'autres facteurs affectant l'apprentissage peuvent inclure les éléments de difficulté des mots discutés au chapitre 2. Ainsi, certains mots avec des taux d'apprentissage élevés mais une faible répétition des livres de cours, tels que *dolphin* et *actress*, sont très imagés, ce qui devrait aider à s'en souvenir. D'autres mots étaient apparentés au grec, comme *acrobat*, et les apprenants ont peut-être trouvé ces types de mots plus accessibles. Un autre facteur peut avoir été que tout le vocabulaire d'un manuel n'est pas disponible pour être lu de la même manière que ces mots. Certaines des expressions les plus fréquemment répétées dans le manuel figuraient dans le cadre

d'instructions écrites pour des exercices et celles-ci avaient des taux de reconnaissance relativement faibles. Peut-être que lorsque l'activité était évidente, les instructions n'avaient pas besoin d'être lues, ni même expliquées, et ces éléments de vocabulaire n'étaient pas consultés par les apprenants, même s'ils étaient présents dans le manuel. Orosz (2008) ajoute encore à la complexité en suggérant que même à des niveaux faibles, les apprenants sont suffisamment motivés dans les écoles hongroises pour suivre des cours supplémentaires et apprendre le vocabulaire de manière approfondie en dehors de la salle de classe.

Milton conclut que la répétition peut vraiment jouer un rôle pour aider à l'acquisition d'un large vocabulaire. Un manuel bien construit avec un bon choix de vocabulaire et qui inclut un peu de recyclage, peut améliorer l'apprentissage et aider à conduire à la croissance de grands vocabulaires, même si tout ne peut pas être recyclé.

Il apparaît également que l'enseignant ne se contente pas de répéter et de recycler la langue du manuel. Le vocabulaire utilisé par l'enseignant est non seulement plus important en quantité que celui contenu dans le cahier de cours, mais il est aussi plus varié. Un vocabulaire considérable est introduit qui ne se produit pas à l'écrit dans le manuel. Au cours de l'année, l'enseignant peut introduire oralement le double de ce que le manuel introduit. Cela ne se limite pas à un cycle particulier ou à une poignée de classes, cela semble être une tendance continue sur toute l'année. Un bon enseignant doit recycler le vocabulaire du manuel en classe et le développer pour donner un environnement lexical riche aux apprenants.

6.5 Conclusion

Beaucoup d'informations ont été présentées jusqu'ici, et pourtant il reste encore beaucoup de zones d'ombre. Nous avons très peu de bonnes informations sur la charge de vocabulaire des manuels de cours et sur les interventions des enseignants en classe, l'organisation de ce vocabulaire et la manière dont ces facteurs sont liés à l'apprentissage. C'est un domaine qui réclame une étude plus systématique. En l'absence de cela, nous nous retrouvons avec des affirmations largement non étayées sur la manière dont les choses doivent être faites : la répétition est bonne, voire essentielle, et introduire 8 à 12 mots par leçon est à peu près correct. Cependant, ces questions sont complexes et le calcul de l'effet du manuel de cours et de l'exposé de l'enseignant sur l'apprentissage du vocabulaire est obscurci par d'autres facteurs qui peuvent

affecter l'apprentissage du vocabulaire présenté aux apprenants et dont nous avons également très peu de connaissances.

Dans l'ensemble, on a l'impression qu'un environnement d'apprentissage en classe de qualité peut contribuer très directement à l'apprentissage du vocabulaire et à une très grande partie de celui-ci. Milton (2009) pense que Harris et Snow et R. Ellis (cité dans Milton, 2009) ont probablement tort de suggérer que la salle de classe joue peu ou pas de rôle dans l'apprentissage du vocabulaire. Pour de nombreux apprenants, il semble que la salle de classe et le manuel sont très importants dans leur apport et peuvent être une condition de réussite. Cela ne signifie pas nécessairement qu'il n'y a pas d'apprentissage en dehors de la salle de classe et à partir d'apports oraux et informels mais l'impact du milieu extrascolaire est altéré dans beaucoup de régions en Algérie par l'absence des formes orale et écrite du français dans le quotidien des apprenants.

Partie pratique

INTRODUCTION

Afin de répondre aux deux questions de recherche, deux études différentes ont été effectuées.

La première étude est une intervention expérimentale qui s'est réalisée dans le cadre d'une étude de cas et est intervenue en deux moments différents de l'année scolaire. Les contraintes des milieux de recherche ont rendu impossible la réalisation de la recherche avec un grand nombre de participants. Toutefois, nous cautionnons entièrement l'idée que «*l'intérêt d'un échantillon se juge pas de manière intrinsèque, en fonction de ses seules propriétés statistiques mais en fonction de son adéquation à une problématique ou une série d'interrogations précises.* ». (Martin, 2009, p. 26)

L'objectif de la première étude est d'évaluer l'étendue du vocabulaire de base chez les collégiens de 1^{ère} AM. Nous ne nous concentrerons cependant que sur le vocabulaire réceptif relevant de la première bande de fréquence. Pour connaître l'impact du milieu sociogéographique sur l'acquisition du vocabulaire, nous avons choisi deux types de participants. 31 participants effectuant leur scolarité dans un collège qui se situe dans un milieu semi-rural et 30 autres participants dont le collège se situe dans une zone urbaine.

La deuxième étude s'est appuyée sur la méthodologie usuelle de recherche linguistique en s'appliquant à un échantillon de corpus. C'est une étude descriptive analytique. Pour ce faire, nous nous avons utilisé une méthode d'échantillonnage raisonné en se basant sur un corpus constitué des textes supports du manuel de la 1^{ère} AM. Le principe de sélection dans l'échantillonnage raisonné est le jugement du chercheur quant à la typicité ou à l'intérêt du corpus constitué en fonction des besoins du chercheur dont le principal objectif de répondre à ses besoins spécifiques dans un projet de recherche. De ce fait, nous avons voulu à travers cette recherche explorer la couverture du vocabulaire dans les textes supports du manuel de la 1^{ère} AM.

CHAPITRE 1 :

VERIFICATION DE LA 1^{ERE} HYPOTHESE (ETUDE 1)

1.1 Méthodologie

1.1.1 Population

Nous avons pris pour notre expérimentation une classe évoluant dans un milieu semi-rural et regroupant 31 élèves, désormais Classe (A), et une autre classe évoluant dans un milieu urbain et regroupant 30 élèves, désormais Classe (B). L'âge moyen de la classe (A) est 11.8 ans et celui de la classe (B) est 11.4 ans. Nous avons voulu connaître à travers la variable (urbanité/ruralité) les proportions de la défiance lexicale dans les deux milieux différents. La Classe (A) est issue d'un milieu socio-culturel défavorisé et la Classe (B) d'un milieu moyen. Chaque élève aurait bénéficié d'environ 366 heures de cours de français tout au long des trois dernières années du cycle primaire. A la fin de l'année, il aurait bénéficié de 116 heures supplémentaires. D'après les professeurs de français de ces deux classes, les élèves seraient très faibles en français. Cette appréciation commune a été émise à l'issue d'une évaluation diagnostique intervenant en début de l'année scolaire. Notons que la compétence lexicale⁶ n'est pas objet d'une évaluation spécifique. Concernant le choix du niveau, en l'occurrence la première année moyenne, nous avons été motivé notamment par l'importance de ce palier qui se situe entre les deux cycles du fondamental et qui est considéré comme étant le palier d'adaptation et d'homogénéisation.

Le nombre d'heures d'enseignement du français au primaire est variable par an. Le temps pédagogique alloué au français en 3^{ème} année primaire est estimé à 96 heures, 144 heures en 4^{ème} année et 126 en 5^{ème} année. Ce temps inclut l'apprentissage et l'évaluation. Au terme du cycle primaire, l'élève aura à son actif 366 heures de cours de français. Le tableau 1 résume le

⁶ Nous entendons par « compétence lexicale » la taille de vocabulaire réceptif qu'un élève possède. Meara (2006) considère que la taille de vocabulaire est la dimension fondamentale de la compétence lexicale. Meara estime que les apprenants avec un grand vocabulaire maîtrisent mieux un large éventail de compétences linguistiques que les apprenants avec des vocabulaires plus petits.

nombre de participants, leurs milieux géographiques, leurs milieux socio-culturels ainsi que leurs âges, leurs sexes et le nombre approximatif d’heures de cours de français suivis.

Par le tableau 1-1, nous dressons le portrait de l’ensemble de la population.

Tableau 1-1 : Portrait de l’ensemble des sujets

Classes	Nombre d’élèves	Milieu géographique	Milieu socio-culturel	Age moyen	sexe		Nombre approximatif d’heures de cours de français suivis
					M	F	
Classe (A)	31	rural	défavorisé	11,8	17	14	366
Classe (B)	30	urbain	moyen	11,4	14	16	366

Nous avons tenu à préciser les heures de cours français dispensés en 3 ans pour avoir une idée sur le profil linguistique de nos sujets qui, précisons-le, ne connaissent le français que pendant les séances de français.

Rappelons que l’enseignement du français se décline durant les quatre années du cycle moyen en trois paliers :

- le 1^{er} palier, incluant la 1^{ère} année moyenne, vise à homogénéiser le niveau des connaissances moyennant les activités de compréhension et de production de textes oraux et écrits relevant de l’explicatif et du prescriptif ainsi qu’à adapter l’aspect comportemental des élèves à la nouvelle organisation collégienne consistant en la pluralité d’enseignants et de disciplines en développant des méthodes de travail efficaces.

- le 2^{ème} Palier, incluant les 2^{ème} et 3^{ème} années, vise à renforcer les compétences disciplinaires et méthodologiques près-acquises pour faire face à des situations de communication disparates moyennant des activités de compréhension et de production de textes oraux et écrits relevant du narratif

- le 3^{ème} palier, incluant la 4^{ème} année, vise à consolider les compétences précédemment installées lors des 1^{er} et 2^{ème} palier moyennant des textes de compréhension et de production de textes oraux et écrits relevant principalement de l’argumentatif ainsi qu’à exploiter la compétence globale du cycle moyen dans le but résoudre des situations-problèmes scolaires ou extrascolaires.

Donc, un élève de première année est censé disposer d'un profil d'entrée ou de connaissances antérieures susceptibles de supporter l'apport langagier du français langue étrangère au cours de l'année scolaire.

1.1.2 Le profil d'entrée « présumé » d'un élève de 1^{ère} AM

Le profil d'entrée d'un élève de 1^{ère} AM n'est que le profil de sortie d'un élève de 5^{ème} année primaire. Ce dernier est défini dans le guide du programme de la langue française de la 5^{ème} AP comme suit :

A la fin de la 5eAP, qui constitue l'année qui clôt le cycle primaire, le profil de sortie de l'élève est défini par l'Objectif Terminal d'Intégration suivant :

Objectif Terminal d'Intégration (désormais OTI) : Au terme de la 5^e AP, l'élève sera capable de produire, à partir d'un support oral ou visuel (texte, image), un énoncé oral ou écrit en mettant en œuvre les actes de parole exigés par la situation de communication.

Pour atteindre cet OTI, le programme de 5eAP a pour objectifs :

A l'oral – L'élève doit être capable de :

- adopter des stratégies adéquates de locuteur ;
- réagir à des sollicitations verbales par un énoncé intelligible et cohérent ;
- s'exprimer de manière compréhensible dans des séquences conversationnelles ;
- réagir à partir d'un support écrit ou sonore ;
- prendre la parole de façon autonome pour questionner, répondre, demander une information, donner une consigne, donner un avis ;
- produire un énoncé pour raconter, décrire, dialoguer ou informer ;
- dire des textes poétiques en s'appuyant sur des éléments prosodiques ;
- synthétiser l'essentiel d'un message oral dans un énoncé personnel ;
- marquer son propos à l'aide d'adverbes, d'interjections et de traits prosodiques.

1.1.3 Le profil de sortie « présumé » d'un élève de 1^{ère} A.M

Compétences globales de fin d'année : Au terme du 1er palier, dans une démarche de résolution de situations-problèmes, à partir de supports sonores et visuels, dans le respect des valeurs et en s'appuyant sur les compétences transversales, l'élève est capable de comprendre et de produire oralement et par écrit, en adéquation avec la situation de communication, des textes explicatifs et des textes prescriptifs.

Compétences terminales : Comprendre des textes oraux explicatifs, prescriptifs et ce, en adéquation avec la situation de communication. Chaque séquence correspond à un niveau de compétence.

1.1.4 Le cadre opérationnel

L'étude que nous avons menée s'inscrit dans une méthode expérimentale. Elle a été réalisée grâce à une expérimentation incluant un pré-test intervenant en début d'année et un post-test intervenant en fin d'année. L'objet de cette expérimentation est donc de rapporter une étude sur la taille de vocabulaire et son développement chez les élèves de 1^{ère} A.M et ce dans le but de collecter des données préliminaires susceptibles de fournir des réponses à certaines des questions soulevées ci-dessous.

1.1.5 Questions de recherche

Pour vérifier la 1^{ère} hypothèse, nous tenterons de répondre aux questions suivantes :

1. Quelles sont les proportions de connaissance de la première bande de fréquence chez un élève de 1^{ère} AM en début d'année ?
2. A quel niveau de CECR correspond le niveau d'un élève de 1^{ère} AM en début d'année?
3. Quels sont les mots les plus cochés lors des deux tests?
4. Combien de mots apprend un élève de 1^{ère} AM au cours d'une années scolaire ?

1.1.6 Choix et élaboration des instruments de mesure

Les différentes méthodes de l'évaluation du vocabulaire sont conçues en fonction de l'objet de l'évaluation lui-même. Lequel objet oscille entre deux visions. L'une considère le vocabulaire comme un ensemble de mots ayant des caractéristiques particulières en dehors de tout contexte, l'autre le considère comme indissociable au contexte. Il incombe par conséquent à l'évaluateur de trouver une vision consensuelle en prenant en compte le niveau des candidats ainsi que les objectifs assignés à l'évaluation du vocabulaire. Pourtant, il a été constaté depuis l'avènement de l'approche communicative une tendance à minoriser les éléments de faible niveau (mots, sons et structures) au profit des compétences langagières réceptives (compréhension de l'écrit, compréhension de l'oral) et productives (production de l'oral, production de l'écrit).

Il existe des preuves empiriques fiables qui confirment que les apprenants ont tendance à apprendre d'abord les mots à haute fréquence, puis à se concentrer sur les éléments à basse fréquence (Milton, 2007) . Cependant, les manuels de cours utilisés dans les salles de classe sont organisés thématiquement et pas nécessairement en fonction de la fréquence, fournissant ainsi un écart potentiel ou un «déficit» dans les profils lexicaux des apprenants (Milton, 2007, p. 51). Par exemple, un apprenant peut connaître un grand nombre de mots qui ne sont pas les plus fréquents dans la langue cible car le matériel de cours est basé sur des thèmes pertinents pour les apprenants de la langue (mais pas pour les locuteurs natifs). Cela peut être particulièrement vrai aux niveaux débutants où une connaissance limitée du vocabulaire est disponible. De notre part, nous examinerons à travers cette étude la taille du vocabulaire chez les élèves de première année moyen au début et en fin d'année.

De même, des recherches antérieures ont montré que la taille du vocabulaire est un bon indicateur de la capacité linguistique générale. Meara et Jones (cité dans Milton, 2009) affirment qu'il existe de bonnes corrélations entre la taille du vocabulaire et les compétences en écriture, la compréhension en lecture et les connaissances grammaticales. Les attentes de réussite sont fréquemment exprimées en termes de vocabulaire et certains programmes proposent des listes de mots. Par exemple, les vocabulaires de base dans les programmes français du GCSE comptent environ 1000 mots (cité dans Milton, 2009). L'une des bases du Cadre européen commun (CECR) est constituée des listes de mots (Van Ek et Trim 1990). La liste de mots de niveau B1 pour le français, équivalent au GCSE en Grande-Bretagne, contient environ 2000 mots. Cela suggérerait qu'un apprenant à ce stade devrait avoir un lexique de cette taille et il existe un écart évident entre les attentes concernant la taille du vocabulaire au niveau B1 en Grande-Bretagne et celle du CECR. De plus, Milton (2006) a constaté que les scores de vocabulaire des apprenants étaient en corrélation avec les notes du GCSE et du A-level. Ce n'est pas surprenant car il a été démontré que le développement du vocabulaire est lié à certains aspects du développement syntaxique dans l'acquisition de la L1 (Bates et Goodman, cité dans David, 2008). Cependant, ce domaine général nécessite plus d'attention. Par exemple, David, Rogers et Myles (cité dans David, 2008) n'ont trouvé que de très faibles corrélations entre des mesures spécifiques de complexité morphosyntaxique et les connaissances lexicales.

En Algérie et à l'instar de beaucoup de pays, l'évaluation de la compétence lexicale dans les épreuves à caractère formatif, sommatif ou certificatif est intégrée dans une évaluation globale de la compétence rédactionnelle qui englobe une variété de connaissances de type

déclaratif et procédural. A ce sujet et tant que praticien ayant participé plusieurs fois à la correction des copies d'élèves candidats à l'épreuve de français au BEM⁷, je me rallie à l'avis de Tréville qui estime que dans ce type d'évaluation « *la part du vocabulaire, celle de la syntaxe et des autres composantes de la compétence de communication, ne peuvent être délimitées qu'à l'aide de grilles de correction, de façon parfois assez approximative, d'ailleurs.* » (Tréville & Duquette, 1996, p. 132)

Il y a lieu de procéder donc à une évaluation de la compétence lexicale dans les programmes d'enseignement « *parce qu'elle fait partie intégrante de la pratique pédagogique et parce qu'elle peut livrer un grand nombre de données sur les différents éléments qui la composent.* » (Tréville & Duquette, 1996, p. 132)

De surcroît, cette évaluation pourrait avoir la capacité de mesurer efficacement la compétence générale. Il est même conféré au vocabulaire le pouvoir de pronostiquer la compréhension générale et les résultats scolaires.

A notre connaissance, aucune évaluation dédiée exclusivement à la mesure de la compétence lexicale est effectuée, prévue ou même envisagée lors du cursus scolaire de l'élève. Certes, les tests et instruments mesurant les connaissances lexicales ne courent pas les rues et encore moins pour le français langue étrangère mais nous avons essayé pour autant d'adapter les instruments qui existent ou du moins de nous en inspirer. Nous avons également été amené à nous plier à des contraintes d'ordre administratif vu que nous n'avons eu accès aux deux collèges que pendant deux matinées et durant une heure pour chaque classe. Il fallait par conséquent que l'épreuve soit rapide à passer mais aussi prolifique en matière de données.

L'épreuve que nous avons élaborée se décline en deux versions. L'une est intervenue en début d'année et l'autre en fin d'année, ce qui nous a permis de procéder à deux collectes de données. Les deux versions sont identiques en forme mais ne contiennent pas les mêmes items. Cette intervention en deux moments différents nous a servi d'établir un bilan diachronique sur l'évolution des connaissances lexicales de nos sujets.

Chaque version d'épreuve comptait une tâche visant l'évaluation du vocabulaire réceptif. Les consignes ont été expliquées aux sujets dans leur langue maternelle (l'arabe dialectal).

⁷ BEM : Brevet d'Enseignement Moyen

En élaborant l'épreuve, nous avons pris en considération le vocabulaire thématique enseigné en cycle primaire et nous avons également insisté à ce que les items proposés soient de grandes fréquences. Nous ne sommes donc pas venus à recourir à la plupart des mots du *Français Fondamental*. Avant le début de l'épreuve, nous avons tenté de reconforter les élèves en les rassurant que le test est utilisé uniquement pour des fins formatives et non pour attribuer une note et aussi en leur expliquant la tâche à accomplir dans leur langue maternelle.

Le test de vocabulaire français utilisé est une adaptation du test de niveaux de vocabulaire X_Lex (Meara et Milton, 2003), un test de taille de vocabulaire EFL qui est largement utilisé et s'est avéré donner des résultats fiables là où, bien sûr, les apprenants n'ont pas recours à taux très élevés de conjectures dans leurs réponses. X_Lex teste la connaissance de chacune des 5 premières bandes de fréquences de mots et estime la connaissance globale de ce vocabulaire. La version française est basée sur les listes de fréquences de Baudot (1992) qui sont tirées d'une variété de sources littéraires, médiatiques et autres principalement de France mais aussi du Canada. Sa portée et sa construction sont comparables à celles des listes bien établies dans d'autres langues également utilisées pour le comptage des fréquences (par exemple, Kucera et Francis, 1967), et le corpus est d'une taille suffisante (plus d'un million de mots courants) pour que les bandes les plus fréquentes soient utilisées comme indicateurs de l'usage standard du français. La liste est lemmatisée.

La version française de X-Lex est un test Oui/Non qui présente aux apprenants 60 mots, un par un. Les apprenants doivent indiquer s'ils connaissent chaque mot. Il y a 40 mots choisis au hasard dans chaque bande de fréquence de 1000 mots et 20 autres pseudo-mots qui sont conçus pour ressembler à des mots en français mais qui ne sont pas de vrais mots français. Le nombre de réponses Oui à ces pseudo-mots permet d'ajuster le score sur les vrais mots pour deviner et surestimer les connaissances. Le test a été livré en version papier et crayon (une version informatique est également disponible) afin de minimiser la perturbation des cours. Il n'y avait pas de limite de temps pour le test qui prend généralement 5 ou 10 minutes à compléter. De notre part, nous nous sommes contentés de la liste de mots appartenant à la première bande de fréquences vu que nous sommes partis du constat que nos sujets ont un vocabulaire très restreint.

Nous avons veillé à ce que les deux tests soient d'une grande congruence puisqu'ils sont conçus dans le but de démontrer l'absence des prérequis et notamment du vocabulaire basique et mesurer également la progression des apprenants quant à l'acquisition du vocabulaire après

une année d'apprentissage. Tous les items contenus dans les tests appartiennent à la première bande de mots de la liste de Baudot comme il a été déjà mentionné. Dans un premier temps, la liste contenait 1500 mais pour cette sélection, seuls les noms, les adjectifs et les adverbes et les verbes et quelques prépositions ont été retenus. Les déterminants qui sont d'une très grande fréquence ont été délibérément omis parce qu'ils présentent peu d'intérêt pour ce dossier. Pour ce qui concerne l'invention des leurres, nous avons fait recours au site www.lextutor.ca. Et pour fiabiliser le test, l'élève doit porter l'équivalent de chaque item reconnu dans sa langue maternelle. En effet, selon P. Nation (cité dans Bogaards, 1994) « *les tests qui demandent aux apprenants de donner un équivalent dans la langue maternelle sont les meilleurs tests de reconnaissance* ». Ce choix implique évidemment une maîtrise de la langue maternelle par les correcteurs. Par ailleurs, nous ne leur avons pas révélé l'existence de pseudo-mots.

Nous avons veillé également à ce que la liste réponde aux huit (08) critères énumérés par Nation :

- les mots sont inclus dans la liste de mots de base en vue d'aider les apprenants à exprimer autant d'idées que possible, et à exprimer les sentiments et les demandes qu'ils souhaitent faire ;
- les mots qui apparaissent fréquemment dans la langue sont inclus dans la liste, tandis que les mots moins fréquents sont exclus ;
- les mots qui apparaissent dans de nombreux types différents de la langue écrite sont préférés aux mots qui n'apparaissent que dans une gamme limitée de sources ;
- les mots dont le sens peut être expliqué par de simples paraphrases sont exclus de la liste de base, de même que les mots qui ont des synonymes plus simples ;
- les mots avec une morphologie régulière sont préférés aux mots avec des irrégularités marquées, et les mots qui apparaissent dans des cadres syntaxiques simples sont préférés aux mots avec des motifs syntaxiques plus complexes ;
- les mots particulièrement utiles pour définir d'autres mots sont inclus dans la liste de base ;
- les mots particulièrement fréquents dans les situations de classe sont inclus dans la liste de base ;
- la liste de base comprend un grand nombre de mots qui sont couramment prêtés dans d'autres langues.

Cependant, nous avons estimé que le huitième critère peut être omis vu que la langue maternelle des sujets et la langue française ne représentent aucun aspect de "congénarité".

Voici un exemple du test de Meara et de la consigne donnée aux participants :

- Cet exercice porte sur la **signification** des mots.

Lisez attentivement les mots de la liste.

Si vous connaissez la signification du mot, inscrivez O (pour oui) dans les parenthèses.

Si vous ne connaissez pas la signification du mot ou si vous n'êtes pas certain de sa signification, inscrivez N (pour non) dans les parenthèses.

Vous devez faire cet exercice rapidement (pas plus de 5 minutes).

Attention! Pour s'assurer que vous ne répondez pas au hasard, il y a aussi des mots imaginaires, qui n'existent pas en français, dans la liste.

1 [] choint	2 [] délicat	3 [] urgence
4 [] avenir	5 [] climat	6 [] plasant

Aussi, faut-il avouer que les connaissances mesurées au moyen de cet outil ne peuvent être considérées qu'assez superficielles car il mesure uniquement la taille du vocabulaire réceptif, c'est-à-dire le nombre de mots que les participants comprennent en reconnaissant leurs formes orthographiques.

1.1.7 Administration des épreuves

Nous avons expliqué les objectifs de notre recherche aux directeurs des deux collèges ainsi qu'aux deux enseignants auxquels sont attribuées les deux classes objets d'étude. Nous avons été autorisé sous réserve d'anonymat, à réaliser notre étude auprès des deux classes évoluant, rappelons-le, dans deux établissements et deux milieux distincts, en l'occurrence, rural et urbain. Toutefois, comme les enseignants sont tenus à mettre en œuvre le programme, nous avons été contraint à assigner une durée totale d'une heure pour les pré-tests en début de l'année, et la même durée pour les post-tests en fin d'année. Cette durée a compris la rassurance et le réconfortement des élèves, l'explication des consignes ainsi que l'exécution des tâches et le ramassage des copies. Cette contrainte de temps a donc été très déterminante quant au choix des instruments de mesure. Néanmoins, il est à préciser que la durée administrée à l'exécution de la tâche a été de 10 minutes. Meara (1992) recommande une durée de 5 minutes car il estime

que si les personnes testées prennent beaucoup plus de temps que cela, elles passent alors trop de temps à penser à des éléments individuels et par conséquent, on doit leur souligner que s'ils sont en cas de doute, ils doivent répondre NON. Vous obtiendrez des données plus fiables si le candidat remplit un test plus long ou un test de plusieurs niveaux.

1.1.8 Collectes de données

Pour calculer la proportion de mots connus, nous avons attribué des notes en % (de 0° au 100%) en nous référant à la formule de Richard C. Anderson et de Peter Freebody (1983, p. 9) où $p(C)$ représente les mots connus, $p(M)$ le nombre de mots réels et $p(PM)$ le nombre de pseudo-mots :

$$p(C) = \frac{p(M) - p(PM)}{1 - p(PM)}$$

Ainsi, un élève ayant cru reconnaître 70% de mots réels et 30 % de pseudo-mots aura un score de 57%. Un autre élève ayant cru reconnaître le même nombre de mots réels mais ayant coché seulement 5% de pseudo-mots aura un score de 65 %. Pour rendre l'opération plus facile, nous avons fait recours au programme Microsoft Excel. Certains élèves ont coché des mots vrais sans pour autant avoir pu donner une traduction juste ou approximative. Leurs réponses n'ont pas été prises en compte.

1.2 RESULTATS

1.2.1 Les proportions de connaissance de la première bande de fréquence chez un élève de 1^{ère} AM en début d'année

Quelles sont les proportions de connaissance de la première bande de fréquence chez un élève de 1^{ère} AM en début d'année ?

Afin de faciliter le calcul, nous avons retranscrit les résultats obtenus sur un classeur Excel. Le tableau 1-2 représente la répartition du nombre d'élèves en fonction des scores obtenus par les élèves de la classe (A) au prétest du vocabulaire.

Tableau 1-2 : Répartition du nombre d'élèves en fonction des score obtenus au prétest (Classe A)

Proportion de mots connus en %	<10	≥ 10 < 20	≥ 20 <30	≥ 30 <40	≥ 40
Nombre d'élèves	23	4	3	0	1
Pourcentage d'élèves %	74,19	12,9	9,67	0	3,22

Les scores de cette expérimentation semblent immédiatement évidents et peu surprenants. Les données présentées dans le tableau montrent nettement et ce, au premier abord visuel, l'hégémonie de la catégorie d'élèves n'ayant pas réussi à franchir le taux de 10 %. En effet, 23 élèves sur 31 de la classe (A), ce qui représente 74,19 % d'élèves, ont eu des scores inférieurs à 10 %, 4 élèves ont obtenu des scores compris entre 10 et 20%, 3 entre 20 et 30 % et un seul élève ayant franchi la barre de 30 %. Si nous considérons que le ratio de 60 % de mots pour 40 % de pseudo-mots, c'est-à-dire une proportion supérieure à 30%, est un score encourageant pour un candidat, nous aurons un seul candidat ayant obtenu un score acceptable parmi les apprenants de la classe (A). A vrai dire, ces résultats sont attendus vu qu'une grande partie d'élèves issues de milieu défavorable éprouvent d'énormes difficultés quant à l'apprentissage du français en général et l'acquisition du vocabulaire en particulier.

La figure 1-1 montre clairement que la grande partie d'élèves de la classe (A) n'ont pas réussi à obtenir un score supérieur à 30 %

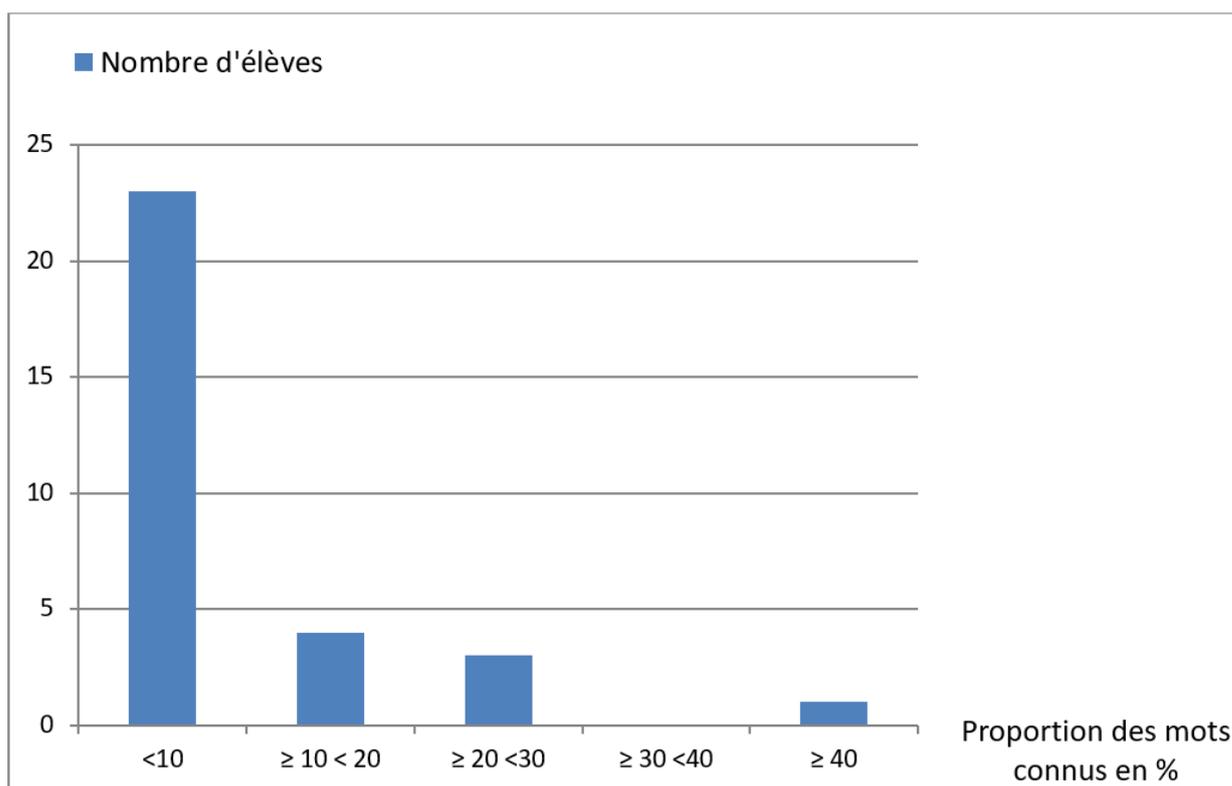


Fig. 1-1 : Répartition du nombre d'élèves en fonction des scores obtenus au prétest du vocabulaire (Classe A)

Le tableau 1-3 représente la répartition du nombre d'élèves en fonction des scores obtenus par les élèves de la classe (B) au prétest du vocabulaire.

Tableau 1-3 : Répartition du nombre d'élèves en fonction des scores obtenus au prétest (Classe B)

Proportion de mots connus en %	<10	≥ 10 < 20	≥ 20 < 30	≥ 30 < 40	≥ 40
Nombre d'élèves	15	5	5	3	2
Pourcentage d'élèves %	50	16,67	16,67	10	6,67

Les résultats présentés dans le tableau 1-3 ne contrastent pas beaucoup avec ceux représentant la classe (A). En effet, il est clair que pratiquement la quasi-totalité de la classe n'est pas parvenue à obtenir un score acceptable. Quinze (15) élèves, représentant la moitié de la classe ont eu des scores inférieurs à 10 %, 5 élèves ont eu des scores oscillant entre 10 et 20

%. Cinq (5) autres élèves ont obtenu des scores compris entre 20 et 30 %. Nous comptons seulement 5 élèves qui sont parvenus à obtenir des scores égaux ou supérieurs à 30 %. Le nombre de ces élèves représente seulement 16,67 % de la classe B.

La figure 1-2 illustre le grand nombre d'élèves de la classe B ayant eu de très faibles scores lors du prétest.

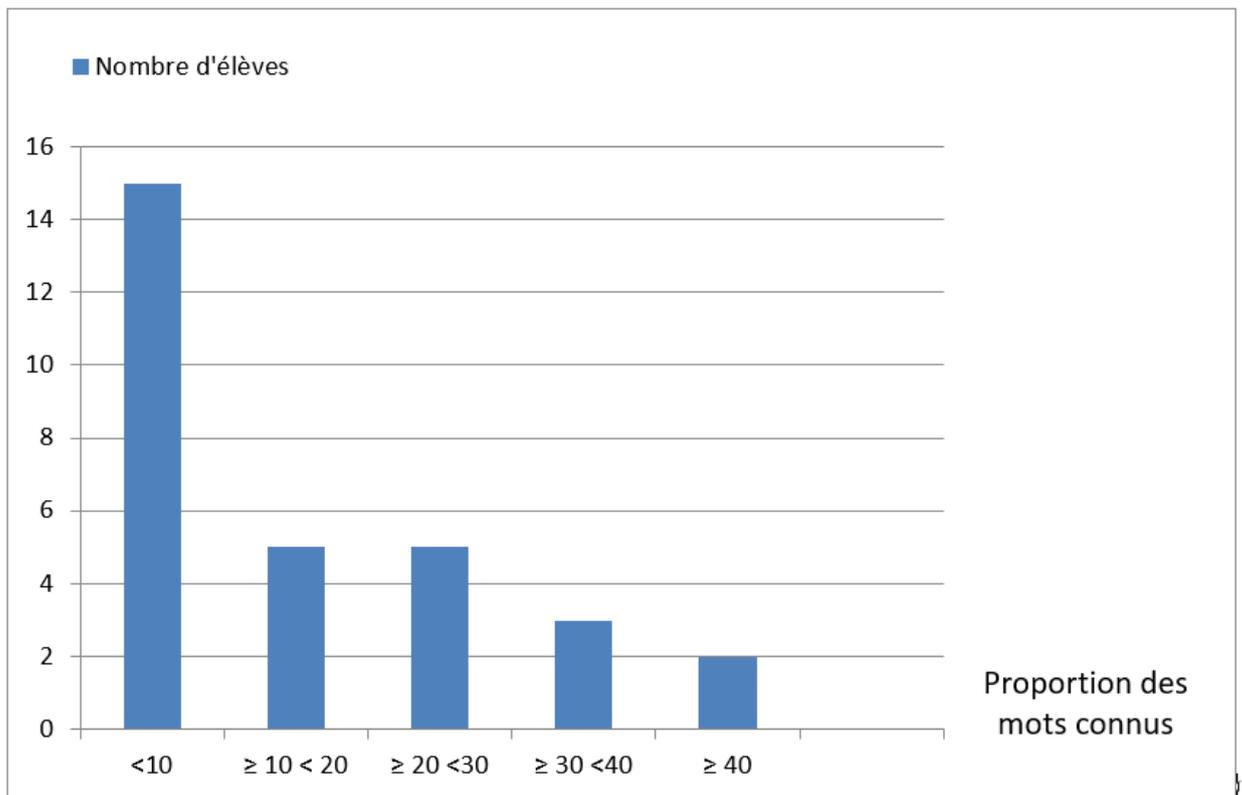


Fig. 1-2 Répartition du nombre d'élèves en fonction des score obtenus au prétest du vocabulaire (Classe B)

Nous allons ensuite calculer dans un classeur Excel les moyennes et les écart-types des deux classes. Les résultats sont indiqués dans le tableau 1-4

Tableau 1-4 : Résultats au prétest du vocabulaire

Classe	Moyenne %	Min %	Max %	Ecart-type %
A	6,27	-14,71	44,74	11,71
B	13,75	-11,11	76,32	17,88

Le tableau 1-4 montre les proportions de connaissance de la première bande de fréquence dans les classes respectives. Les variations, parfois incroyablement marquées, de connaissance du vocabulaire passif d'un apprenant à un autre est probablement la première remarque que l'on peut tirer de cette étude.

A la lumière de ces résultats, nous pourrions dessiner les deux profils lexicaux des classes respectives puis les comparer l'un avec l'autre. Pour ce faire, nous prendrions en compte les moyennes des proportions obtenus.

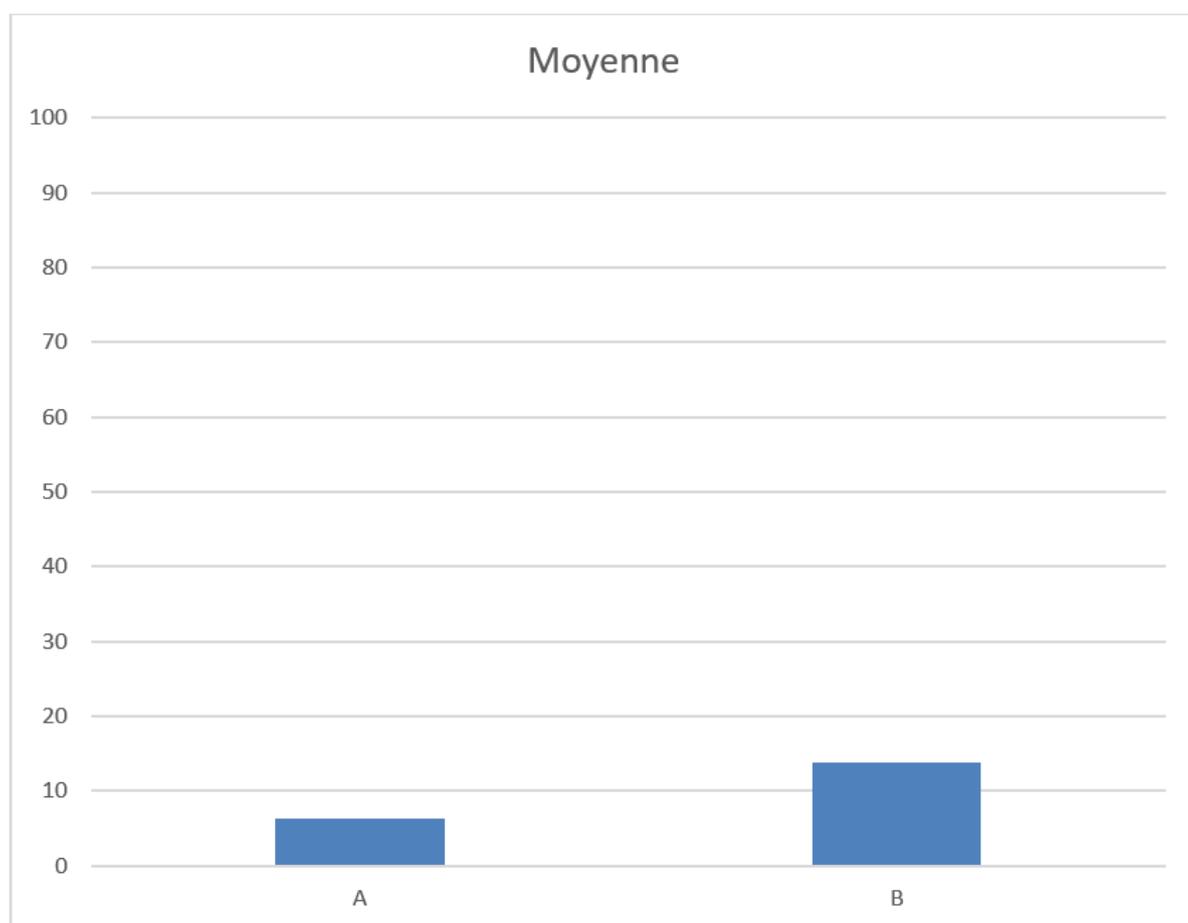


Fig. 1-3 : Profils lexicaux des deux classes

La figure 1-3 représente les profils produits par les apprenants des deux classes respectives en début d'année à l'issue d'un test X-Lex mesurant les proportions de connaissances de la première bande de fréquence. En comparant ces profils avec d'autres profils produits suite à d'autres expérimentations menées notamment par Meara et Nation, nous pouvons confirmer que les apprenants des deux classes, objet de notre étude, dessinent des

profils lexicaux irréguliers. Nous estimons que ce déficit affiché du niveau 1 de connaissance du vocabulaire constituera un défi majeur pour les enseignants si nous admettons que ces scores nous permettent de pronostiquer les résultats scolaires. Nous pouvons également constater le léger écart qui existe entre le profil lexical de la classe (A) évoluant dans un milieu semi-rural et celui de la classe (B) évoluant dans un milieu urbain.

1.2.2 Le niveau d'un élève de 1^{ère} année par rapport au niveau CECR

A quel niveau de CECR correspond le niveau d'un élève de 1^{ère} AM en début d'année?

Des recherches antérieures ont montré que la taille du vocabulaire est un bon indicateur de la capacité linguistique générale. Meara et Jones (cité dans David, 2008) affirment qu'il existe de bonnes corrélations entre la taille du vocabulaire et les compétences en écriture, la compréhension en lecture et les connaissances grammaticales. Par exemple, les vocabulaires de base dans les programmes français du GCSE comptent environ 1000 mots. L'une des bases du Cadre européen commun (CECR) est constituée des listes de mots. Pourtant, les estimations de la taille de vocabulaire chez les apprenants français dans l'Union Européenne ne sont pas les mêmes. Alors qu'en Espagne et en Grèce, les apprenants semblent avoir besoin de connaître, en moyenne, plus de 2000 mots français pour progresser au-delà des niveaux élémentaires A1 et A2, les données britanniques suggèrent que cela peut être réalisé avec moins de la moitié de ce nombre, moins de 1000 mots. Dans tous les cas, les chiffres suggèrent qu'il faut beaucoup moins de mots en français pour atteindre ce niveau de compétence qu'en EFL. Comme pour les données EFL, les scores moyens et la tendance générale des groupes masquent la grande variété de scores individuels qu'illustrent les chiffres de l'écart type.

Les scores moyens de taille de vocabulaire pour les apprenants de français langue étrangère en Grèce à tous les niveaux du CECR sont présentés dans le tableau 1-5. Les scores moyens de taille de vocabulaire pour les apprenants de français langue étrangère en Espagne à tous les niveaux du CECR sont présentés dans le tableau 1-6. Il n'y a pas de notes indicatives pour les tailles de vocabulaire pouvant être associées à chaque niveau.

Tableau 1-5 : Taille du vocabulaire du français langue étrangère et CECR chez les apprenants en Grèce (Milton & Alexiou, 2009, p. 202)

<i>CEF level</i>	<i>Mean</i>	<i>Max.</i>	<i>Min.</i>	<i>SD</i>	<i>n</i>
A1	1125.71	2550	0	620.40	35
A2	1756.25	2500	1500	398.60	8
B1	2422.72	3400	1800	517.37	11
B2	2630.00	2850	2250	251.49	5
C1	3212.50	3750	2600	473.24	4
C2	3525.00	4150	2900	883.88	2

Tableau 1-6 : Taille du vocabulaire du français langue étrangère et CECR chez les apprenants en Espagne (Milton & Alexiou, 2009, p. 202)

<i>CEF level</i>	<i>Mean</i>	<i>Max.</i>	<i>Min.</i>	<i>SD</i>	<i>n</i>
A1	894.44	2850	350	604.61	18
A2	1700.00	2750	500	841.50	9
B1	2194.44	3100	1100	717.39	9
B2	2450.00				1
C1	2675.00	3600	1900	643.23	6
C2	3721.42	4200	3200	416.19	7

Nous allons tenter d'associer un niveau CECR aux apprenants sujets de notre étude. Pour ce faire, il nous a fallu une reconversion des proportions de mots connus en nombre de mots connus. Nous avons effectué le calcul de la connaissance du vocabulaire des sujets en comptant le nombre de réponses Oui aux mots réels et en le multipliant par 50 pour donner un score brut sur 5000. Le nombre de réponses Oui aux pseudo-mots est ensuite calculé et multiplié par 250. Ce chiffre est déduit du score brut pour donner un score ajusté qui comprend ainsi une compensation pour les conjectures. Par exemple, si un élève a 30 bonnes réponses mais a toutefois répondu à 3 pseudo-mots, son score sera comme suit : $(30 \times 50) - (3 \times 250) = 1500 - 750 = 750$ mots connus. Les scores ajustés sont ceux rapportés dans cette étude. Ce sont des estimations de la taille du vocabulaire réceptif passif des apprenants.

Le tableau 1-7 montre les scores obtenus lors du prétest.

Tableau 1-7 : Résultats au prétest du vocabulaire

Classe	Moyenne	Min	Max	Ecart-type
A	27,24	-700	700	231,60
B	187	-450	1300	347,39

Les résultats indiquent qu'en début d'année, un élève de la classe (A) connaît en moyenne 27,24 mots. Tandis qu'un élève de la classe B connaît en moyenne 187 mots. Ces scores appuient une fois de plus notre hypothèse qui postule que les élèves de 1^{ère} AM souffrent d'un manque conséquent de vocabulaire élémentaire. Les moyennes de taille de vocabulaire les moins exigeantes sont ceux des apprenants espagnols de français langue étrangère. Ces moyennes avoisinent, comme indiqué au tableau ci-dessus, le chiffre de 900 mots pour le niveau élémentaire A1. Il ressort de tableau 1-7 qu'aucune des deux classes ne possède un profil lexical répondant aux normes CECR, ne serait-ce qu'à titre indicatif.

Encore une fois, ces résultats sont susceptibles de remettre en cause tout le programme de français de 1^{ère} AM qui, s'appuyant sur des profils d'élèves « irréalistes », continue de tracer ses objectifs en fonction de ce qui est censé être et non de ce qui en est réellement.

1.2.3 Les mots les plus cochés lors des deux tests

Quels sont les mots les plus cochés lors des deux tests ?

Le tableau 1-8 montre les mots qui ont été les plus cochés lors du prétest. Pour des raisons d'ordre méthodologique, nous avons fusionné les résultats des deux classes. Le mot « *médecin* » a été le plus coché avec 28 bonnes réponses, suivi de « *travailler* » avec 26 bonnes réponses puis « *jouer* » avec 20 bonnes réponses, « *chercher* » avec 18 réponses, « *fille* » avec 18 bonnes réponses, « *homme* » 14 bonnes réponses, « *arriver* » avec 15 bonnes réponses, « *combien* » avec 12 bonnes réponses et enfin « *fils* » avec seulement 7 bonnes réponses. Toutefois, nous remarquons l'absence notable des mots qui sont très fréquents tels que « *matin* », « *joie* », « *pays* » ou encore « *trouver* ». Nous pouvons comprendre la présence de quelques mots au sommet de cette liste mais nous nous montrons incapable de comprendre l'absence totale d'autres mots. En effet, le mot « *médecin* » est très fréquent en classe de langue et il peut être encore plus fréquent quand on parle de métiers en classe d'autant que c'est le métier le plus convoité par nos élèves. Le mot « *travailler* » est présent en tout heure puisqu'il

fait partie du métalangage. Toutefois, il n'est pas évident qu'il soit méconnu de plus de la moitié de notre population. L'absence du mot « *chercher* » dans le palmarès est encore plus intrigant : plus de deux tiers d'apprenants n'ont pas coché ce mot. Plus grave encore, des mots fonctionnels comme « *avec* » ou « *combien* » n'ont pas été cochés par une grande partie d'apprenants : 31 bonnes réponses pour « *avec* » et seulement 12 bonnes réponses pour « *combien* ». Vous pouvez imaginer la perplexité d'un élève ignorant le sens des mots comme « *avec* », « *combien* » ou « *trouver* » devant des mots comme « *féculent* », « *éco-citoyen* » ou « *dépister* ». Nous savons que la connaissance des formes phonologique et orthographique d'un mot peut ne pas être la même et nous ignorons dans quelle mesure les enseignants utilisent-ils le français en classe de langue pour expliquer les consignes. Autrement dit, nous ignorons les conditions d'exposition de nos apprenants aux formes phonologique et orthographique des mots en cycle primaire.

Tableau 1-8 : Quelques mots les plus cochés dans les deux versions du test

Item	Nombre de bonnes réponses
médecin	28
travailler	26
jouer	20
chercher	18
filles	18
homme	14
arriver	15
combien	12
fil	7
voiture	41
eau	28
titre	32
avec	31
soleil	33
expliquer	23
vivre	21
terre	14
parler	12

1.2.4 La moyenne du nombre de mots appris par un élève de 1^{ère} AM au cours d'une année scolaire

Combien de mots aurait appris un élève de 1^{ère} AM au cours de l'année scolaire ?

Selon Milton (2008, p. 334), les apprenants anglais gagnent en moyenne environ 170 mots par an jusqu'au GCSE et environ 535 mots par an entre le GCSE et le niveau « A ». Lorsque le temps passé en classe est pris en compte, il apparaît que les apprenants progressent d'environ 2,7 mots par heure de contact jusqu'au GCSE et d'environ 4,1 mots par heure au cours de l'étude de niveau « A ». Ces résultats sont globalement similaires à ceux de Milton et Meara (1998) qui suggèrent que les apprenants dans une variété de contextes linguistiques apprennent environ trois à quatre mots par heure de classe. Milton estime que ces scores semblent nous fournir une image convaincante de la façon dont la langue se développe. La taille du vocabulaire a généralement tendance à augmenter au cours de l'étude, comme on pouvait s'y attendre.

Pour connaître le développement de la taille de vocabulaire de nos deux classes au cours de l'année scolaire, nous avons recouru aux données collectées grâce au posttest, rappelons-le, intervenant en fin d'année scolaire.

La tableau 1-9 nous montre les scores obtenus lors du posttest.

Tableau 1-9 : Résultats au posttest interprétés en nombre de mots connus

Classe	Moyenne	Min	Max	Ecart-type
A	67,74	-800	950	293,00
B	260	0	1400	259,77

Le tableau nous indique qu'un élève de la classe (A) connaît en moyenne au terme de l'année scolaire 67,74 mots contre 27,24 mots en début d'année. Le gain en taille de vocabulaire de la classe (A) peut être facilement déduit en soustrayant la moyenne initiale de la moyenne finale. En effet, un élève de la classe (A) aurait appris en moyenne 40,5 mots du début de l'année scolaire à la fin de l'année. Si nous effectuons les mêmes calculs pour la classe (B), le gain moyen en taille de vocabulaire sera de 73 mots. Pour connaître la moyenne de nombre de mots appris par heure de contact, nous avons divisé la moyenne de mots appris par le nombre d'heures de contact estimé à 116 heures. Le tableau 1-10 résume les scores obtenus.

Tableau 1-10 : Moyennes de mots appris lors de l'année et par heure de contact

Classe	Moyenne de mots connus en début d'année	Moyenne de mots connus en fin d'année	Moyenne de mots appris	Moyenne de mots appris par heure
Classe A	27.24	67.74	40.5	0.35
Classe B	187	260	73	0.63

Bien que ces résultats soient encore une fois très prévisibles, ils dépassent de loin les estimations les plus pessimistes. Dans un premier temps, nous constatons le rendement plus considérable (avoisinant presque le double) de la classe (A) par rapport à la classe (B). Cette variation de scores peut être expliquée par le fait que les apprenants pauvres acquièrent beaucoup moins que la moyenne et il y a un très petit nombre de scores nuls tout au long des premières années d'apprentissage. Très probablement, il se peut aussi qu'il y ait des élèves testés qui sont si peu intéressés par le sujet qu'ils ne démontreront pas leurs connaissances dans un test de ce genre. Cependant, presque certainement aussi, ces faibles scores indiquent qu'il y a des apprenants qui gagnent très peu de leurs cours de langue étrangère et acquièrent en effet si peu de vocabulaire même sur plusieurs années d'apprentissage, qu'il est difficile de récupérer dans un test de ce genre. Le gain moyen de 2.4 mots par heure de contact considéré comme étant le moins flatteur par Milton est loin d'être réalisé par les apprenants de notre test. La figure 1-4 illustre le grand écart qui existe entre la consommation moyenne de vocabulaire considérée comme la plus faible par Milton et les scores réalisés dans notre expérimentation.

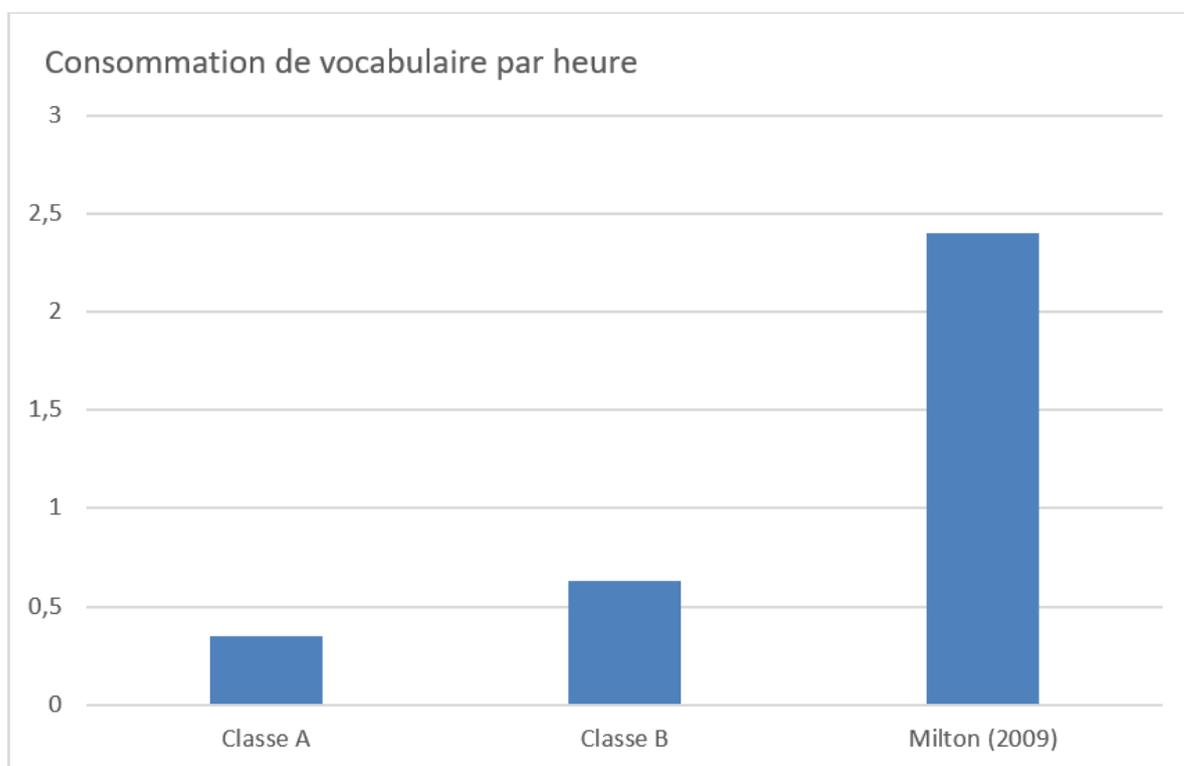


Fig. 1-4 : Comparaison entre les consommations moyennes de vocabulaire par heure

1.3 DISCUSSION

Comme le montre notre analyse de recherche, la pauvreté lexicale n'est pas seulement associée à notre population semi-rurale, mais affecte également notre population urbaine. Les problèmes liés au vocabulaire sont ressentis par les deux groupes, même à des degrés divers. Compte tenu de la petite taille de notre échantillon, nos résultats ne peuvent certainement pas être généralisés, mais ils renseignent sur l'ampleur de ce manque de mots. Le manque de vocabulaire qui, par son endémicité, s'est souvent révélé être la racine du mal, dont le plus saillant est la répulsion de la langue française. Cette étude étaye d'une manière claire et nette notre première hypothèse qui stipule que les difficultés d'appropriation de vocabulaire sont en premier lieu tributaires du manque de vocabulaire basique.

Nous voyons bien que pour les deux classes, une grande partie des apprenants (96,76 % de la classe (A) et 76,67 de la classe (B) n'ont pas réalisé un score satisfaisant. De surcroît, la variable (urbanité/ruralité) n'a pas été vraiment très significative. La différence entre les deux scores pourrait être imputable à la présence plus importante du français dans le milieu urbain comparé au milieu rural. Elle peut aussi s'expliquer par des méthodes d'enseignement ou

d'enseignants différents en primaire. Rappelons pour autant que nos apprenants apprennent le français en tant que langue étrangère et dans un contexte strictement formel. Nous ne savons pas dans quelle mesure les enseignants de primaire parlent français en classe. Ces apprenants ne sont pas entourés de français en dehors des cours, à travers des films, des chansons et des publicités, et il est beaucoup plus difficile pour eux d'accéder à la langue étrangère. Le français langue étrangère n'est pas considéré comme un atout éducatif ou professionnel par de nombreux élèves, voire la plupart, et leurs parents. Aujourd'hui, beaucoup d'indices indiquent que le français cède du terrain à l'anglais.

Les scores des deux versions du test ont révélé aussi qu'aucune des deux classes n'est éligible à un niveau CECR. La taille moyenne de vocabulaire estimée lors du prétest ne dépassant pas les 200 mots dessine deux profils lexicaux au-dessus de la moyenne exigée par le niveau de base A1. Les tailles moyennes du vocabulaire chez les apprenants de la classe (A) et ceux de la classe (B) sont respectivement estimées à 27,24 mots et 187 mots. Ces deux profils lexicaux contrastent avec les objectifs assignés au palier fondés sur des profils inexistantes. Ils constitueront également une entrave au processus d'installation ultérieure de compétences plus complexes.

Les apprenants de cette étude ont eu un total de 116 leçons de français au cours d'une année scolaire. L'assimilation du vocabulaire des apprenants a été calculée pour montrer combien de mots en moyenne sont appris par heure ; les chiffres ont montré une très faible assimilation de vocabulaire par heure de contact. Les apprenants de la classe (A) n'ajoutent qu'un seul mot en 3 heures approximativement. Quant aux apprenants de la classe (B), leur consommation en vocabulaire ne dépasse pas les deux mots en à peu près 3 heures. Cela semble largement inférieur aux résultats cités dans la partie théorique de notre travail qui concernent les études sur les taux de consommation du vocabulaire chez les apprenants.

1.4 CONCLUSION PARTIELLE

Que peut-on alors conclure de cette étude ? Une conclusion est que les apprenants ne semblent pas posséder des pré-acquis lexicaux et que par conséquent, ne font pas de progrès dans l'apprentissage du vocabulaire. La présente étude a montré que les élèves de 1^{ère} AM, qu'ils évoluent dans un milieu rural ou urbain, souffrent d'une carence lexicale aiguë après 3 années d'apprentissage de français au primaire. Cette carence semble persister, et ce, en dépit

des efforts fournis par les enseignants et par la tutelle. L'effet de Mathieu émerge en affirmant que la pauvreté en vocabulaire génère inéluctablement un ralentissement dans le processus d'acquisition et d'apprentissage du vocabulaire. Ce manque de vocabulaire s'explique par le fait que l'enseignement du vocabulaire n'est pas le centre d'intérêt des enseignants et des inspecteurs du primaire. De fait, l'objectif de construction de la compétence lexicale est paradoxalement déprécié au profit d'autres compétences qui y sont étroitement inhérentes. A ce sujet, Tréville (2001) s'interroge sur le fait paradoxal qui consiste en l'omission de la compétence lexicale dans une pédagogie qui se concentrerait sur la pratique de la communication. Il faut aussi remarquer qu'après l'avènement des approches communicatives, l'enseignement du vocabulaire se fait implicitement, c'est-à-dire que les apprenants, en s'exposant constamment à des supports authentiques écrits et oraux parviendraient à s'approprier du vocabulaire.

L'absence de vocabulaire de base endigue la formation de ce que J. Aitchison (cité dans Bogaards, 1994) appelle des « toiles verbales ». G. Doca (cité dans Bogaards, 1994) en arrive à la conclusion que « *les mots nouveaux sont dirigés systématiquement vers certaines « constellations » de significations, vers certains champs lexicaux où ils s'intègrent par un processus de sélection ou de « filtrage » des significations* ». Mais, faut-il encore que ces « constellations » existent d'abord chez l'apprenant.

Bien qu'elle ne nous permette pas de tirer des conclusions générales par rapport à la taille du vocabulaire réceptif d'un collégien de 1^{ère} année, du fait qu'elle est restreinte, notre étude contribue en revanche à inciter les chercheurs à se pencher sur ce sujet car c'est de leur concours qu'il devient possible d'éclairer dans le menu les causes de ce déficit et leur proposer dans le prolongement des pistes de solution. En outre, des recherches pourraient aussi porter sur les deux autres paliers du cycle moyen, à savoir les 2^{ème}, 3^{ème} et 4^{ème} années. Cette étude invite également les praticiens à consacrer plus de temps à des activités visant à enrichir le vocabulaire et à adopter une approche cohérente et systématique. Beaucoup de questions pourraient également être soulevées : quel contenu lexical enseigner au primaire ? Combien de mots doit connaître un élève à l'issue du cycle primaire ? Quels moyens faut-il mettre en œuvre pour développer la compétence lexicale chez les élèves. Ce genre de questions pourrait être abordé dans de futures recherches dédiées à l'enseignement/apprentissage du vocabulaire du français langue étrangère en Algérie.

1.5 LIMITES

A l'instar des études réalisées à ce jour dans ce domaine de recherche, notre étude comprend certaines limites. La limite la plus saillante et sans contredit la plus importante, est la taille réduite de notre échantillon. Ce dernier est constitué de seulement 61 participants. Ce petit nombre nous empêche de généraliser les résultats à l'ensemble des élèves de la première année moyenne.

« The results are likely to be more reliable both because a larger sample size is always likely to give better results than a smaller sample » (Milton, 2009, p. 72)

Nous justifions cet échantillon restreint par le fait que, depuis le début de ce projet, nous avons rencontré beaucoup de difficultés d'ordre administratif. Nous avons également eu beaucoup de mal à convaincre les deux directeurs de collège ainsi que les deux enseignants à adhérer à notre projet vu, qu'en général, les staffs éducatifs se montrent très réservés quant aux collaborations avec les chercheurs considérés souvent comme des personæ non gratae. De plus, en raison de la nature des tests, un plus grand échantillon aurait exigé beaucoup plus de temps et de volontaires et pourrait même avoir des répercussions négatives sur la faisabilité de l'étude elle-même.

La deuxième limite de la recherche concerne le test de vocabulaire et son lien avec la compréhension orale. En effet, le test de vocabulaire que nous effectuons mesure uniquement la reconnaissance de la forme écrite des mots, alors que la compréhension orale implique la reconnaissance orale des mots. Nous savons qu'une langue étrangère n'est pas seulement apprise par sa forme écrite. Il est possible d'apprendre une langue uniquement par le son et de ne jamais apprendre à la lire ou à l'écrire. Certaines méthodologies d'enseignement promeuvent en fait cette approche plus courantes, pour les personnes qui ont vécu à l'étranger. Apparemment, beaucoup d'entre elles peuvent devenir très compétents en expression orale mais, parce qu'elles ne l'ont jamais fait, ont du mal à lire le passage le plus simple dans leur langue étrangère. Il est donc possible de savoir comment les mots sonnent, sans reconnaître à quoi ils ressemblent à l'écrit. L'inverse n'est peut-être pas vrai. Les apprenants, lorsqu'ils sont confrontés à des mots écrits dans une langue étrangère, ont une forte tendance à sous-vocaliser et lorsqu'ils sont confrontés à des mots inconnus, ils peuvent en fait les « prononcer ». La fiabilité de la corrélation entre le test de vocabulaire et la compréhension pourrait être plus élevée si nous effectuions un test de vocabulaire supplémentaire, soit un test de reconnaissance

orale, Aural-Lex⁸ de Milton. Certes, nous avons rapporté une étude de Milton (2006) qui démontre une forte corrélation entre le X-Lex et l'Aural Lex. Il aurait tout de même pu être intéressant de l'observer nous-même auprès de notre population cible. Il se peut aussi que d'autres formes de mesure produisent des résultats différents des mesures de l'étendue du vocabulaire orthographique. Néanmoins, toutes ces mesures, du moins dans les populations d'apprenants, interagissent de manière prévisible et en relation, probablement, avec l'effet de fréquence.

Cependant, cela aurait demandé une somme de temps supplémentaire afin de créer une version française du test et, pour la passation du test, nous aurions eu besoin d'un accès à un laboratoire informatique. Personnellement, nous avons constaté beaucoup de confusion parfois insensées de la part des apprenants à l'égard des mots phonologiquement très différents comme « *cœur* », « *cour* » et « *corps* ». Nous assignons cela au système phonologique de la langue maternelle dépourvu de voyelles.

La troisième et dernière limite estimée se relève aussi du test de vocabulaire d'autant que ce dernier ne couvre que les 1000 mots les plus fréquents en français. Un profil lexical ordinaire suggère que la proportion de mots connus décroît au fur et à mesure qu'on monte à un niveau plus haut de fréquence. Néanmoins, certains apprenants pourraient bien présenter des profils atypiques en connaissant des mots aux niveaux 2000, 3000 ou 4000 tout en ignorant beaucoup de mot appartenant à la première bande de fréquence. Nous pensons pour autant qu'en dépit des résultats que pourront avoir les apprenants choisis comme échantillon, un déficit au niveau de la première bande de fréquence représente un sérieux avilissement à la qualité de la compétence lexicale.

⁸ Aural Lex est un autre test populaire qui évalue la taille du vocabulaire phonologique en présentant des mots hors contexte (Milton & Hopkins, 2006). C'est la version orale du test X-Lex. Il s'agit d'un test Oui/Non fourni par ordinateur, qui présente aux apprenants 120 mots, un par un. Les apprenants doivent indiquer s'ils connaissent chaque mot. Il y a 20 mots choisis au hasard dans chaque bande de fréquence de 1 000 mots, et 20 autres pseudo-mots qui sont conçus pour ressembler et sonner comme des mots en anglais mais qui ne sont pas de vrais mots anglais. Le nombre de réponses Oui à ces pseudo-mots permet d'ajuster le score sur les vrais mots pour deviner et surestimer les connaissances. Il n'y a pas de limite de temps pour le test, qui prend généralement 5 ou 10 minutes.

CHAPITRE 2 : **VERIFICATION DE LA 2^{IEME} HYPOTHESE (ETUDE 2)**

Notre 2^{ème} hypothèse, rappelons-le, stipule que les difficultés liées à l'appropriation du nouveau vocabulaire pour les élèves de 1^{ère} AM seraient également tributaires de la structure lexicale des textes supports proposés dans le manuel scolaire. Nous allons dans ce qui suit essayer de vérifier cette hypothèse.

2.1 Méthodologie :

D'une manière ou d'une autre, le vocabulaire a toujours été présent dans la méthodologie d'enseignement de la L2, de la grammaire-traduction traditionnelle à la méthode communicative. En fait, il a gagné du terrain depuis le milieu des années 70, si bien que l'on peut affirmer que le vocabulaire est aujourd'hui l'une des pierres angulaires de l'apprentissage d'une langue seconde.

Les concepteurs de manuels scolaires sont conscients de l'importance du vocabulaire dans l'apprentissage d'une langue étrangère. C'est pourquoi ils accordent une place considérable au vocabulaire dans leurs manuels. Normalement, le vocabulaire est présenté dans différents contextes tels que des dialogues, des narrations, des chansons ou d'autres types de texte. Néanmoins, il y a aussi des allusions explicites à certains groupes de mots qui constituent le vocabulaire cible des séquences didactiques. Ces mots sont traités tout au long de la séquence au moyen de différentes activités mais nous pensons aussi que les séances de compréhension de l'écrit et de lecture entraînent demeurent une source essentielle de l'input du vocabulaire en classe de langue.

La sélection d'un manuel de cours approprié est l'un des défis auxquels les enseignants et les institutions doivent faire face chaque année. Cependant, la sélection n'est pas facile du tout. Pourtant, il incombe aux enseignants et aux institutions de choisir les supports en fonction à la fois des besoins de l'apprenant et des conseils des experts. De plus, il appartient aux concepteurs

de répondre aux attentes des autorités, des enseignants et des apprenants. De notre côté, nous pensons qu'une grande partie d'enseignants n'est pas assez formée pour faire des choix judicieux de textes supports d'où la nécessité de porter un intérêt particulier quant à la sélection des textes supports vu que le manuel représente pour beaucoup d'enseignants l'unique source.

Malgré l'importance reconnue du vocabulaire et des manuels dans l'acquisition des langues étrangères, les raisons du choix de ces derniers ne sont pas toujours claires. Ainsi, les documents du Ministère de l'Education Nationale ne fournissent aucune information spécifique sur les mots que l'apprenant doit connaître à la fin de son enseignement primaire, moyen ou secondaire.

En règle générale, ces documents officiels font des allusions vagues et peu concluantes à un vocabulaire tel que « L'objectif est de permettre à l'élève d'élargir son stock lexical et d'en faire une véritable ressource linguistique en le mobilisant dans les activités de réception et de production. » (Document d'accompagnement du programme de français, cycle moyen)

Le présent exemple illustre l'incapacité du document à spécifier combien et quels mots sont exactement ceux que les apprenants doivent connaître ou apprendre. Et c'est ici que la fréquence et la fonctionnalité devraient jouer un rôle.

Pour cela et compte tenu du fait qu'il n'y a pas eu d'études qui ont examiné la couverture des éléments lexicaux des manuels scolaires destinés aux apprenants du collège et particulièrement du premier palier de l'enseignement moyen, nous avons eu l'idée de tenter de combler ce manque de recherche et d'enquêter sur les opportunités d'apprentissage du vocabulaire que les apprenants du FLE reçoivent dans les collèges algériens. Plus important encore, il est nécessaire de mener des recherches axées sur la réalisation d'une enquête approfondie sur les opportunités qui aideraient les apprenants algériens à approfondir leur connaissance des mots. En effet, des données plus qualitatives doivent être recueillies afin de mieux comprendre la manière dans laquelle le matériel lexical est présenté dans les manuels en se focalisant notamment sur l'aspect de la fréquence. À notre connaissance, aucune étude n'a été menée pour examiner les caractéristiques qualitatives des mots présentés dans les manuels utilisés dans le contexte algérien. Dans cet esprit, cette étude vise à effectuer la première analyse approfondie pour enquêter sur divers aspects des caractéristiques et de la distribution des mots couverts dans les textes supports utilisés dans les manuels de 1^{ère} A.M.

Le terme linguistique de corpus a été décrit (McEnery & Wilson, 1996) en termes simples comme l'étude du langage basée sur des exemples d'utilisation du langage dans la « vie réelle ». Il a une histoire relativement longue. La linguistique de corpus n'est pas une branche de la linguistique telle que la syntaxe, la sémantique et la pragmatique qui se concentrent sur la description ou l'explication de certains aspects de l'utilisation de la langue. C'est une méthodologie qui peut être appliquée à un large éventail d'études linguistiques.

Biber, Conrad et Reppen (1998, p. 4) énumèrent les caractéristiques essentielles de la linguistique de corpus :

- elle est empirique, analysant les schémas réels d'utilisation dans les textes naturels ;
- elle utilise une vaste collection de textes naturels, connue sous le nom de «corpus», comme base d'analyse;
- elle utilise largement les ordinateurs pour l'analyse, en utilisant à la fois des techniques automatiques et interactives;
- elle dépend à la fois de techniques analytiques quantitatives et qualitatives.

2.1.1 Matériaux

Étant donné que le but de cette étude est d'étudier la distribution et les caractéristiques des éléments lexicaux des textes du manuel scolaire de la 1^{ère} AM pour les séances de compréhension de l'écrit et de lecture, notre corpus sélectionné se compose de 8 textes. Le nouveau manuel de la première année est entré en vigueur en 2016/2017. Le guide du professeur prévoit un seul texte à analyser en séance de compréhension de l'écrit (lecture silencieuse) et que l'élève lira de façon expressive en séance de lecture-entraînement. Il justifie ce choix par le fait qu'un élève lit mieux un texte qu'il a étudié et dont il maîtrise le sens.

2.1.2 Modification des textes

Le manuel a été téléchargé électroniquement sous forme de fichier PDF. Nous avons ensuite procédé à la sélection des textes objet de notre étude un par un en les copiant et les collant dans un nouveau document Word. Une fois fini, et pour pouvoir extraire commodément les informations, nous avons enregistré les textes sous forme de document Word dans un seul document. Donc, la disponibilité de la version numérique du manuel nous a épargné beaucoup d'efforts et de temps puisque nous n'avons pas été obligé de scanner les textes et d'utiliser un logiciel OCR pour les convertir par la suite en documents exploitables.

Afin d'assurer l'exactitude du comptage des mots, nous avons examiné le fichier texte pour exclure toute information superflue telles que les sources. Le tableau 2-1 présente les types de textes proposés ainsi que leurs compositions lexicographiques.

Tableau 2-1 : Compositions lexicographiques des textes supports du manuel de la 1^{ère} AM

Texte	Type du texte	Tokens	Types	Lemmes
Des gestes simples pour une bonne hygiène corporelle	explicatif	134	84	78
Que devrions-nous manger et boire ?	explicatif	133	80	74
« Pour votre santé, bougez ! »	explicatif	139	83	82
Quel est le train le plus rapide ?	explicatif	138	92	85
Peux-tu m'expliquer la pollution de l'air ?	explicatif	135	92	89
Que se passerait-il si toutes les glaces des pôles fondaient ?	explicatif	156	106	94
Pourquoi devrions-nous utiliser des énergies renouvelables ?	prescriptif	140	100	95
Les bons réflexes pour devenir un éco-citoyen	prescriptif	139	88	82
	Total	1114	725	679

Comme l'illustre le tableau 1-1 ci-dessus, les textes proposés comme supports pour la 1^{ère} AM ont presque la même taille avec un nombre minimal de 133 mots et un nombre maximal de 156 mots. Au total, ces textes sont composés de 1114 tokens⁹, de 725 types¹⁰ et de 679 lemmes.

A posteriori, nous pouvons estimer que les textes sont relativement longs et difficilement exploitables en une heure.

2.1.3 Questions de recherche

Comme dans la première étude et pour vérifier la deuxième hypothèse, nous avons essayé de répondre aux questions de recherche suivantes.

1. Dans quelle mesure les mots de K1, K2, AWL et des listes de mots supplémentaires sont-ils représentés dans les textes supports du manuel de la 1^{ère} A.M. ?
2. Quel est le comportement de chaque liste de mots le long des huit textes ?
3. Quels sont les cent mots les plus fréquents dans les textes supports ?
4. Dans quelle mesure le vocabulaire thématique a-t-il été recyclé ?

2.1.4 Conception de la recherche

Étant donné que notre étude vise à mener une enquête approfondie sur la distribution et les caractéristiques des éléments lexicaux présentés dans les textes utilisés comme supports dans le manuel de la 1^{ère} A.M et afin de répondre aux sous-questions de recherche, nous avons opté pour une méthode mixte.

⁹ Une façon consiste simplement à compter chaque forme de mot dans un texte parlé ou écrit et si la même forme de mot apparaît plus d'une fois, chaque occurrence est comptée. Ainsi, la phrase, *Le bébé pleure dans le salon*, contiendrait six mots, même si « le » apparaît deux fois. Les mots qui sont comptés de cette manière sont appelés des tokens ou jetons,

¹⁰ Nous pouvons compter les mots dans la phrase "*Le bébé pleure dans le salon*" d'une autre manière. Quand on voit le même mot se reproduire, on ne le compte plus. Ainsi, la phrase de six tokens se compose de cinq types différents. Nous comptons les mots de cette manière si nous voulons répondre à des questions telles que « Quelle était la taille du vocabulaire de Shakespeare ? », ou « Combien de mots contient ce dictionnaire ? »

En termes généraux, les méthodes mixtes (MM) peuvent être définies comme une approche à la connaissance (théorie et pratique) qui tente d'envisager plusieurs points de vue et perspectives. Ivankova, Creswell et Stick (2006, p. 3) les définissent comme une procédure pour collecter, analyser et « mélanger » ou intégrer des données qualitatives et quantitatives (dérivées de méthodes) à un certain stade du processus de recherche dans une même étude dans le but de mieux comprendre le problème de la recherche. Briggs, Coleman et Morrison (2012) ont défini les *mixed methods research (MMR)* traduit en français par « *recherche par méthodes mixtes* » (RMM), comme des grands types de recherche dans lesquels des éléments ou des approches de la recherche quantitative et qualitative sont combinés ou mélangés dans une étude de recherche [traduction libre] (p. 122-123). Les données quantitatives et qualitatives générées par cette étude aident à mieux comprendre la relation entre la couverture lexicale des textes supports et la mesure dans laquelle ils offrent des opportunités d'apprentissage du vocabulaire aux apprenants.

Par conséquent, le type de conception de la méthode mixte que la présente étude a suivi était séquentiel explicatif, ce qui signifie que l'étude commence par la collecte et l'analyse des données quantitatives d'abord, puis passe à la collecte et à l'analyse des données qualitatives. Cela nous a permis d'obtenir des informations plus détaillées concernant la couverture des éléments lexicaux. La figure 2-1 illustre le processus que le chercheur a suivi pour recueillir les données.

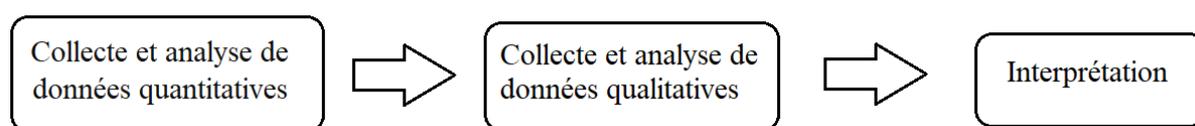


Fig. 2-1 : Séquence de collecte des données de la méthode mixte

En ce qui concerne les informations quantitatives, le type de données qui ont été recueillies dans la présente étude tournait autour de la couverture de K1¹¹, K2¹², AWL¹³ et des mots hors liste (ci-après MHL). Suite à l'analyse des données quantitatives, nous avons commencé à collecter des données qualitatives pour mieux déterminer si la couverture lexicale aiderait les apprenants à approfondir leur connaissance des mots. Pour ce faire, la conception de l'étude de cas a été choisie pour recueillir et analyser les données qualitatives car elle aide à parvenir à une compréhension aussi complète que possible du phénomène. Une étude de cas, telle que définie par Merriam (1998), étudie un phénomène dans un temps et un lieu délimités qui aidera à acquérir « *une compréhension approfondie de la situation et du sens pour les personnes impliquées* » [traduction libre] (p. 19).

L'étude de cas sélectionnée pour la présente étude tournait autour de l'analyse des listes de mots K1, K2 et AWL qui apparaissaient dans les matériaux sélectionnés. Selon Nation (2001) et Milton (2009), les chances d'apprendre un mot dans une langue seconde ou étrangère augmentent lorsqu'il a été recyclé cinq fois ou plus. Cette analyse contribue à une meilleure compréhension de la profondeur de la connaissance des mots que les textes fournissent aux apprenants.

Selon le Ministère de l'Éducation Nationale, les finalités de l'enseignement des langues étrangères définies dans la Loi d'Orientation sur l'Éducation Nationale (n°08-04 du 23 janvier 2008 – chap. I, art. 2 et chap. II, art. 4) sont citées dans le programme unique du cycle moyen. Elles visent à la mise en place de compétences de locuteur, de lecteur et de scripteur autonome dans des situations complexes, dans un monde en perpétuelle transformation de par l'explosion des connaissances et le développement des technologies de l'information et de la communication. Au-delà des spécificités de la discipline (domaines de connaissances et champ disciplinaire), l'enseignement du français, en complément avec les autres disciplines, contribue à la concrétisation des objectifs de transmission et d'intégration des valeurs républicaines et démocratiques, identitaires, sociales et universelles. Ainsi, l'objectif général de l'enseignement

¹¹ K1 représente les 1000 premiers mots les plus fréquents

¹² K2 représente le 2^{ième} millièm de mots les plus fréquents

¹³ AWL est l'abréviation de Academic Word List. En français, cette liste n'existe pas. Selon une étude de Cobb et Horst (2004), il semblerait qu'il n'y a pas de place pour un AWL en français, puisque dans la quasi-totalité des cas examinés, l'objectif d'atteindre 90 % de couverture textuelle peut être atteint en maîtrisant le vocabulaire commun du français. Nous entendons ici par AWL, la liste de mots commençant par le 3^{ième} millièm de mots les plus fréquents et éliminant les mots appartenant à la liste MHL (Mots Hors Liste).

du français dans les écoles publiques algériennes est d'aider les élèves à améliorer leurs compétences en français et à communiquer plus efficacement dans des situations réelles (quotidiennes). Étant donné que les mots des listes de mots K1 et K2 constituent généralement le pourcentage le plus élevé de mots dans tout discours écrit ou parlé, y compris les conversations quotidiennes et les programmes télévisés, nous avons analysé le contenu des textes supports en mettant en exergue les mots les plus fréquents présentés dans ces textes. En outre, nous avons mis l'accent sur le comptage des mots de l'AWL car la plupart des élèves ont besoin de développer leur vocabulaire académique pour pouvoir exploiter les différents types de textes proposés en cycle moyen, secondaire et même à l'université.

2.1.5 Analyse des données

Pour mener à bien l'investigation des items lexicaux, nous avons utilisé deux outils informatiques pour analyser les matériaux sélectionnés. Le premier est *Anatext*, un logiciel de traitement de corpus en ligne développé par Olivier Kraif de l'Université des Alpes à Grenoble, qui offre des fonctions d'exploration de corpus intéressantes et très utiles en particulier en ce qui concerne la lexicométrie, c'est-à-dire les statistiques lexicales. Par exemple, l'outil peut extraire la fréquence des mots, les mots les plus spécifiques ou des séquences répétées de mots (fragments répétés) du texte. Ce programme informatique vous permet également de classer le vocabulaire de n'importe quel texte en catégories de fréquence. Le second est *vocabprofil*, est une adaptation pour le français du logiciel *Lexical Frequency Profile* (Laufer et Nation, 1995) disponible sur www.lextutor.ca, qui mesure la richesse lexicale d'un texte en comparant ce dernier (mots, vocables et familles) à des listes de fréquence pré-établies. Ces listes sont basées sur l'analyse d'un corpus de plus de 50 millions de mots publiés en 1998 dans *Le Monde* et *Le Soir*, respectivement en France et en Belgique.

Hunston et Francis (2000) ont défini l'analyse basée sur la linguistique de corpus comme une façon d'étudier le langage en observant de grandes quantités de discours naturel stocké électroniquement, à l'aide d'un logiciel qui sélectionne, trie, associe, compte et calcule [traduction libre] (p. 15). Comme l'ont indiqué Hoffmann et al. (2008), la linguistique de corpus est une méthode essentiellement quantitative, ce qui signifie que la linguistique de corpus a tendance à compter les caractéristiques du langage (ou à faire en sorte que l'ordinateur les compte) dans le cadre de son analyse des caractéristiques linguistiques » (p. 18). McEnery et Wilson (1996) décrivent le terme linguistique de corpus comme l'étude du langage basée sur des exemples d'utilisation du langage dans la « vie réelle ». Il a une histoire relativement longue.

La linguistique de corpus n'est donc pas une branche de la linguistique telle que la syntaxe, la sémantique et la pragmatique qui se concentrent sur la description ou l'explication de certains aspects de l'utilisation de la langue. C'est une méthodologie qui peut être appliquée à un large éventail d'études linguistiques. Dans tous ces différents domaines d'étude linguistique, il y a cinq étapes principales lorsque nous examinons l'approche basée sur le corpus :

1. Question : Une question ou un modèle de recherche est conçu
2. Build : conception et compilation de corpus
3. Annoter : Analyse computationnelle du corpus
4. Récupérer : Analyses quantitatives et qualitatives du corpus
5. Interpréter : Interprétation manuelle des résultats ou confirmation de l'exactitude du modèle

Un modèle de processus simplifié de la méthodologie basée sur corpus généralement utilisée par les chercheurs est le suivant : il commencerait par l'identification d'une question de recherche, continuerait par la construction et l'annotation d'un corpus avec lequel étudier le sujet, et finirait par la récupération, l'extraction et l'interprétation des informations du corpus qui peuvent aider le chercheur à répondre à la question de recherche ou à confirmer les paramètres du modèle. Dans certains cas, le processus peut être itératif, où, suite à l'interprétation des résultats, un raffinement est nécessaire sur la question de recherche ou l'annotation du corpus.

Au fil des années, l'analyse linguistique a été au cœur des études de corpus. Cependant, avec la première moitié des années 1980, de nombreuses recherches ont été menées sur l'enseignement et l'apprentissage de corpus et de langues étrangères. John Sinclair, qui est professeur au département d'anglais, a fait la première étude importante dans ce domaine. Il a créé le projet COBUILD (Sinclair, 1988) qui consistait à exécuter les données du corpus dans des ordinateurs et à travailler avec ses implémentations dans l'enseignement des langues. Outre le projet COBUILD, il existe de nombreux autres projets visant à appliquer des études de corpus dans l'enseignement et l'apprentissage des langues étrangères. Toutefois, il semble que l'enseignement de langues est épargné par ce développement des études de corpus car peu d'enseignants ont la connaissance des outils de corpus, des logiciels informatiques sur corpus.

La figure 2-2 montre que les applications directes de l'utilisation de corpus dans l'enseignement et l'apprentissage des langues incluent l'apprenant et l'enseignant qui utilisent activement les corpus et la concordance au moyen d'activités pratiques. Les applications indirectes, quant à elles, concernent les chercheurs, les concepteurs de matériaux et les fabricants de programmes. Römer (2010) affirme que les applications directes de l'utilisation de corpus dans l'enseignement des langues cherchent à savoir comment enseigner, tandis que les applications indirectes cherchent à savoir quoi et quand enseigner.

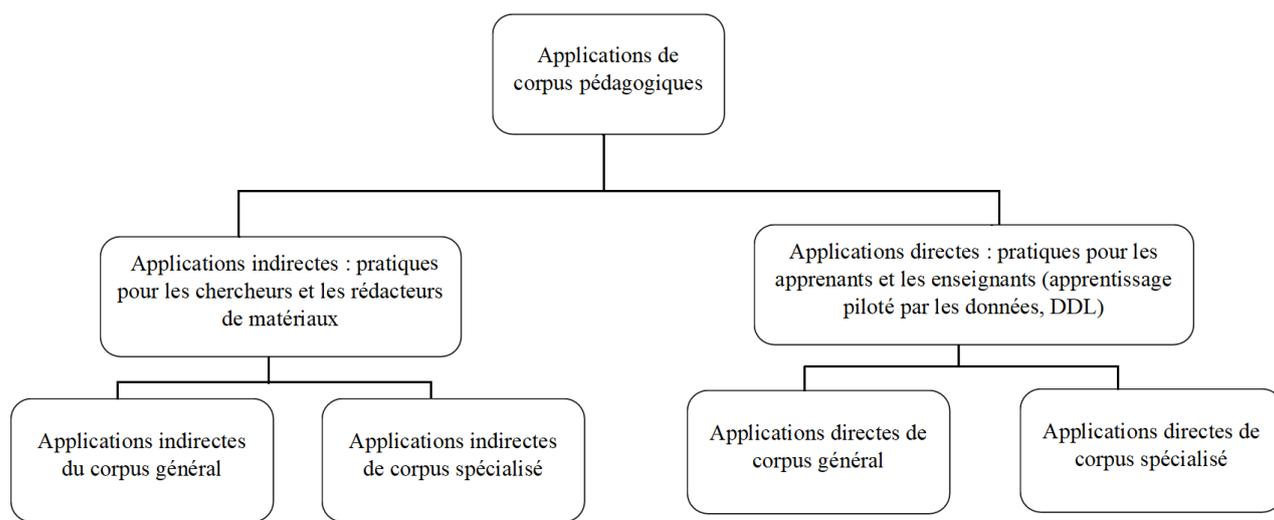


Fig. 2-2 : L'utilisation des corpus dans l'apprentissage et l'enseignement des langues (Römer, 2010, p. 9)

Un terme connexe et parfois chevauchant est la linguistique informatique, qui est également une méthodologie dans l'étude des langues. Grishman (1986) le décrit comme étant l'étude des systèmes informatiques pour comprendre et générer le langage naturel :

« *Computational linguistics is the study of computer systems for understanding and generating natural language.* » (Grishman, 1986, p. 4)

La linguistique computationnelle se concentre sur « *l'utilisation des ordinateurs comme un outil pour modéliser des fragments de théories linguistiques jugées d'un intérêt particulier* » (Boguraev et al, cité dans Rayson, 2002). La linguistique computationnelle s'est concentrée sur trois domaines d'application : la traduction automatique, la recherche d'informations et les interfaces homme-machine.

2.1.6 Profilage statistique

Les listes de mots triés en fréquence font depuis longtemps partie de la méthodologie standard d'exploitation des corpus. Sinclair (1991, p. 30) écrit : « *Quiconque étudie un texte a probablement besoin de savoir à quelle fréquence chaque forme de mot différente y apparaît* ». Tribble et Jones (1997, p. 36) décrivent une méthodologie pédagogique pour l'utilisation de textes en classe de langue, proposant que le point de départ le plus efficace pour comprendre un texte est une liste de mots triés par fréquence. La liste de fréquence enregistre le nombre de fois que chaque mot apparaît dans le texte ; il peut fournir des informations intéressantes sur les mots qui apparaissent (ou n'apparaissent pas) dans un texte. La liste peut être organisée par ordre de première occurrence, par ordre alphabétique ou par ordre de fréquence. L'ordre de première occurrence sert de guide rapide pour la distribution des mots dans un texte, une liste alphabétique est construite principalement à titre de référence, mais une liste par ordre de fréquence met en évidence les mots les plus courants dans le texte. Même les dictionnaires les plus traditionnels peuvent utiliser les informations de fréquence. Francis et Kučera (1967) poussent la simple liste de fréquence des mots un peu plus loin en signalant les fréquences grammaticales des mots. Cela donne des fréquences de mots avec leurs étiquettes de partie du discours. Le profil de fréquence d'un texte donné peut être comparé à celui d'autres textes similaires ou à celui de grands corps de texte. Étant donné que les éléments à haute fréquence ont généralement une distribution stable, des changements importants dans l'ordre des mots dans la liste de fréquence peuvent signaler des éléments d'intérêt pour le chercheur (Sinclair J. , 1991, p. 31). De telles techniques peuvent être réalisées manuellement pour un petit corpus mais sinon nous avons besoin de l'aide d'un programme informatique.

2.2 RÉSULTATS

2.2.1 Proportions des listes de fréquences dans les textes supports du manuel de 1^{ère} année.

Dans quelle mesure les mots de K1, K2, AWL et des listes de mots supplémentaires sont-ils représentés dans les textes supports du manuel de la 1^{ère} A.M. ?

Après avoir analysé les textes supports du manuel par le logiciel *Vocabprofil*, nous avons obtenu quatre listes de fréquences : K1 contenant les items des 1000 mots les plus fréquents en français, K2 avec ce vocabulaire appartenant au deuxième 1000 mots les plus fréquents, une

troisième liste (K3) avec des mots de moindre fréquence mais couramment présente dans l'enseignement secondaire et aussi dans les textes universitaires ; enfin un ensemble de mots n'appartenant à aucune de ces catégories, étiquetés les "Mots Hors Liste" (ci-après MHL). Il est important de se rappeler que notre objectif principal est d'établir un profil lexical des textes supports du manuel de la 1^{ère} année moyenne.

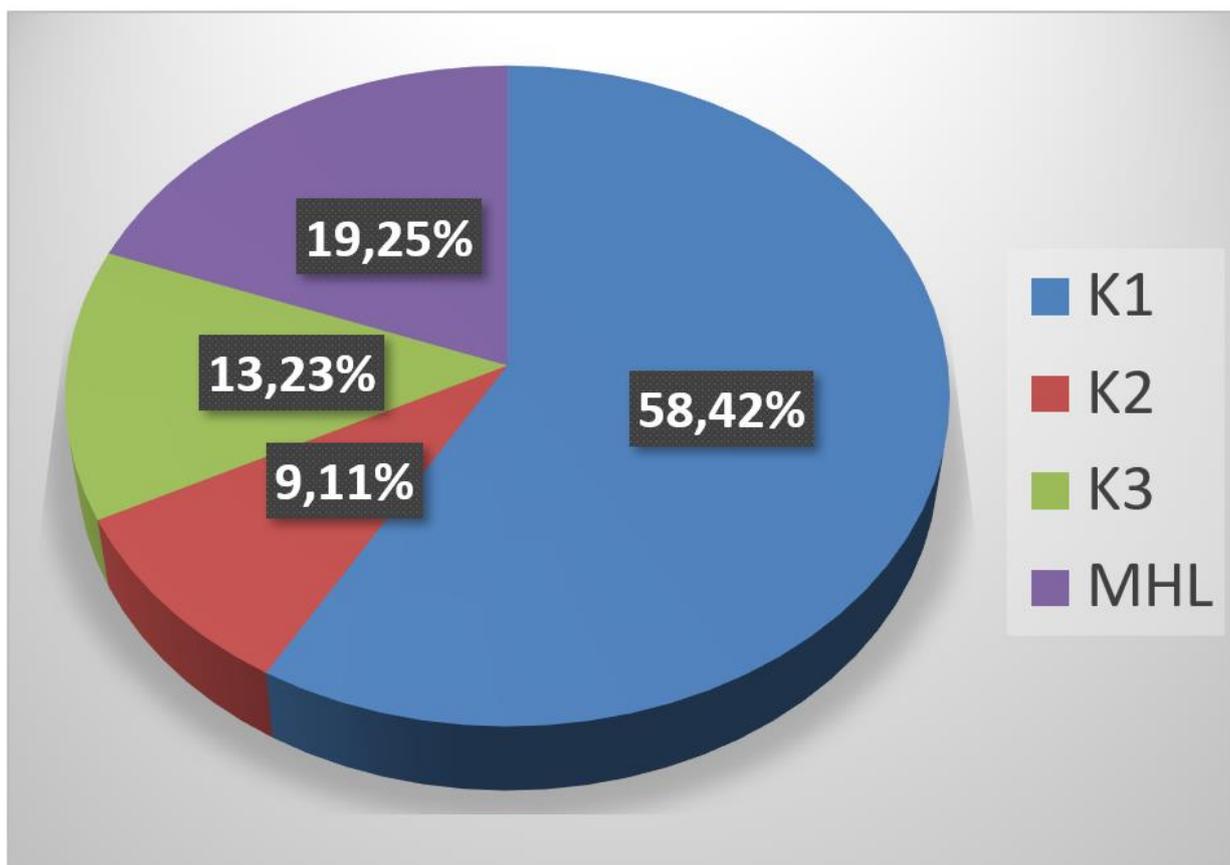


Fig. 2-3 : Pourcentages de représentation des listes de fréquences dans les textes support du manuel de la 1^{ère} année moyenne

La figure 2-3 montre comment les résultats indiquent une nette prédominance de K1 avec près de 58,42% des mots appartenant à ce rang de fréquence. Viennent ensuite les mots MHL avec 19,25% suivis des mots de K3 avec 13,23 % et enfin le moins représenté, K2 avec seulement 9,11 % de présence dans le corpus. En admettant que les listes de mots répertoriés au début de chaque séquence représentent le vocabulaire cible, 54 % de ces mots font partie de la catégorie K3+MHL, suivi de près par K1 avec 27,2%. Encore une fois, K2 est la moins remarquable, contenant le plus petit nombre de mots cibles, en d'autres termes, seulement 18,77

% du vocabulaire cible appartient à la K2. Ensemble, les listes K1 et K2 assurent une couverture de 67,53 %. Théoriquement, elles devraient couvrir plus de 80 % de l'ensemble de textes. Selon Cobb et Horst (2004), il apparaît que les apprenants de français, contrairement à ceux de l'anglais, peuvent atteindre le même niveau de couverture d'un texte académique (ou de tout autre type de texte) lorsqu'ils ont acquis la connaissance du sens des 2000 familles de mots français les plus courantes. Dans un premier temps, nous pouvons dire que les proportions des bandes de fréquences sont dans l'ensemble normales sauf pour la liste MHL qui est anormalement surreprésentée.

2.2.2 Le comportement de chaque liste de mots le long des huit textes supports

Quel est le comportement de chaque liste de mots le long des huit textes ?

En plus de l'analyse globale du manuel, nous devrions également donner un aperçu de chaque séquence didactique. Le tableau 2-2 montre les proportions de fréquence dans chaque séquence.

Tableau 2-2. Pourcentage des fréquences dans les textes supports

Séquence	Texte support	K1	K2	K1+K2	K3	MHL
S 1	Des gestes simples pour une bonne hygiène corporelle	60,21%	6,13%	66,34%	12,68%	20,98%
S 2	Que devrions-nous manger et boire?	56,08%	10,94%	67,02%	15,08%	17,90%
S 3	«Pour votre santé, bougez!»	61,98%	10,20%	72,18%	13,15%	14,67%
S 4	Quel est le train le plus rapide ?	68,98%	11,89%	80,87%	12,05%	7,08%
S 5	Peux-tu m'expliquer la pollution de l'air ?	57,32%	6,90%	64,22%	13,18%	22,60%
S 6	Que se passerait-il si toutes les glaces des pôles fondaient ?	56,56%	4,90%	61,46%	13,50%	25,04%
S 7	Pourquoi devrions-nous utiliser des énergies renouvelables ?	54,33%	13,69%	68,02%	12,10%	19,88%

S 8	Les bons réflexes pour devenir un éco-citoyen	51,87%	8,22%	60,09%	14,08%	25,83%
-----	---	--------	-------	--------	--------	--------

Comme pour les résultats des textes supports dans leur ensemble, la prédominance de K1 se manifeste dans chaque séquence. Sa présence oscille entre près de 51,87 % dans la séquence 8 et 68,98 % dans la séquence 4. En revanche, le poids de la liste de mots MHL est toujours supérieur aux listes K2 et K3 sauf pour la séquence 4. La liste des 2000 mots les plus fréquents (K1+K2) offre des couvertures allant de 60,09% dans la S8 à 80,87% dans la S4.

La figure 2-4 montre le comportement de chaque liste de mots le long des huit unités. Le chemin le plus régulier est représenté par K3, avec des mouvements ascendants et descendants doux. Nous constatons également que les listes se comportent de la même manière sauf dans la séquence 4 où les listes K2 et K3 s'égalisent presque et où la proportion de liste MHL est inférieure à celles des K2 et K3.

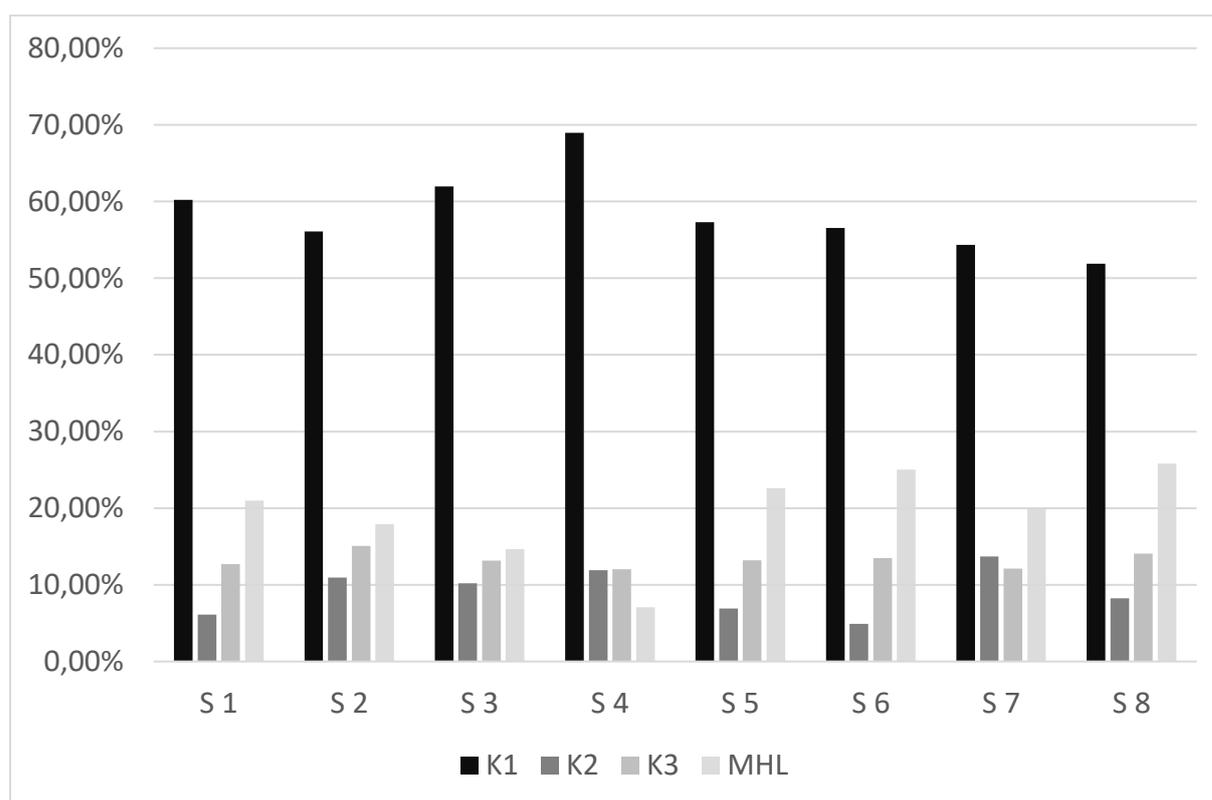


Fig. 2-4 Proportions des listes de fréquence dans les huit textes supports du manuel de la 1^{ère} AM

Donc, il est évident que la proportion des listes varie d'une séquence à une autre, selon les champs sémantiques qui prédominent dans chacune d'elles.

2.2.3 Les mots les plus fréquents dans les textes supports du manuel

Quels sont les mots les plus fréquents dans les textes supports du manuel ?

La liste des fréquences nous donne des informations sur le vocabulaire de base du corpus dans le manuel. Le tableau ci-dessous montre les 100 mots les plus fréquemment utilisés dans le manuel.

Tableau 2-3 : Les 100 mots les plus fréquents dans les textes supports du manuel de la 1^{ère} année moyenne

Rang	Lemme	Fréquence	Fréq. dans le corpus de réf. (par M)	Rang	Lemme	Fréquence	Fréq. dans le corpus de réf. (par M)
1	le	141	19118,475	51	lavage	3	4,63
2	de	78	31794,895	52	vitesse	3	56,36
3	du	40	5566,61	53	bon	3	1095,235
4	et	34	16224,435	54	pourquoi	3	1245,93
5	être	29	30713,08	55	OMS	3	
6	à	25	15668,695	56	mais	3	4523,025
7	un	24	13300,855	57	glace	3	77,01
8	en	19	10389,525	58	utilisation	3	4,12
9	pour	16	6530,765	59	aliment	3	5,055
10	que	14	3611,815	60	faire	3	
11	il	13	15306,775	61	passer	3	1312,22
12	se	11		62	Pour	3	6530,765
13	nous	10	4560,175	63	activité	3	32,835
14	plus	9	4416,555	64	environnement	3	7,17
15	qui	9	6496,51	65	sur	3	3937,755
16	ce	9	7186,04	66	personne	3	686,585
17	eau	9	361,26	67	année	3	332,69
18	tout	8	5635,48	68	humain	3	150,935
19	avoir	7	25283,525	69	rapide	3	67,23
20	santé	7	58,82	70	maison	2	513,85
21	par	7	2711,825	71	régulièrement	2	17,2
22	dans	6	6341,035	72	geste	2	155,245
23	ou	6	1931,83	73	gaz	2	34,835
24	énergie	6	47,6	74	ville	2	323,575
25	boire	5	295,175	75	légume	2	17,575
26	devoir	5	2397,445	76	réflexe	2	15,145
27	on	5	6849,315	77	aussi	2	1338,755
28	notre	5	838,01	78	autre	2	1375,865
29	manger	4	360,08	79	dire	2	5269,61
30	ne	4	17814,115	80	développement	2	9,64

31	utiliser	4	128,655
32	pouvoir	4	4173,48
33	train	4	260,265
34	mondial	4	15,435
35	quand	4	1582,985
36	consister	4	19,95
37	physique	4	39,17
38	au	4	4441,805
39	main	4	845,575
40	différent	4	109,21
41	maladie	4	55,01
42	air	4	593,485
43	jour	4	1135,055
44	comme	3	3167,47
45	certain	3	142,925
46	pollution	3	2,205
47	polluer	3	3,035
48	éviter	3	105,475
49	grand	3	1045,315
50	exercice	3	25,97

81	si	2	2657,95
82	aller	2	6328,615
83	site	2	11,69
84	mer	2	169,155
85	d'après	2	
86	Maglev	2	
87	votre	2	0,15
88	nombreux	2	53,485
89	entier	2	115,81
90	animal	2	90,04
91	agir	2	218,485
92	mètre	2	91,935
93	sans	2	1591,315
94	monde	2	782,915
95	éolien	2	1,515
96	élever	2	76,23
97	parc	2	38,73
98	polluant	2	0,13
99	perdre	2	471,675
100	son	2	3197,94

Comme le montre le tableau 2-3, les prépositions, les pronoms, les déterminants, les conjonctions, les verbes auxiliaires sont généralement en haut de la liste. On peut dire que leur place en tête de liste résulte de leurs fonctions ; en d'autres termes, ces mots ne sont pas répétés intentionnellement. C'est seulement qu'à partir du 20^{ème} rang qu'un mot lexical « santé » fait son apparition.

On voit que l'article défini « le » est le mot le plus fréquent dans le corpus avec une fréquence relative de 141 occurrences dans notre corpus. Il est suivi de la préposition « de » avec une fréquence relative de 78, de l'article partitif « du » avec une fréquence relative de 40 et de la conjonction de coordination « et » avec une fréquence relative de 34. Ces petits mots qui constituent la trame structurelle de la langue et qui sont les éléments récurrents des phrases se retrouvent dominer les listes de fréquence. Ils sont fonctionnels plutôt que lexicaux ; ils ont peu d'aide sur le contenu du corpus.

O'Keefe, McCarthy et Carter (2007) affirment que si les pronoms personnels « je » et « tu » et les marqueurs de discours tels que oui, bonjour, euh sont reflétés dans le rang élevé de la liste de fréquences de tout corpus, on peut dire que le corpus présente un environnement locuteur-auditeur, ce qui est bénéfique pour l'environnement communicatif des apprenants. O'Keefe, McCarthy et Carter (2007) ajoutent en outre que s'il y a une prédominance des

références à la troisième personne avec des prépositions et des conjonctions dans tout corpus, le corpus représente le "monde là-bas".

À cet égard, il ressort des 100 mots les plus fréquents des textes supports du manuel que les pronoms personnels « je » et « tu » sont absents de la liste ainsi que les éléments de marquage du discours tels que « bonjour » ne font pas partie des mots les plus fréquents dans les textes support du manuel objet d'étude. Cela montre que le vocabulaire du manuel n'aide pas l'environnement communicatif des élèves. Cette absence est due à la typologie textuelle choisie pour les élèves de 1^{ère} AM. En effet, 6 textes sur 8 sont des textes explicatifs et les deux restant sont des textes prescriptifs.

Concernant les parties du discours, on trouve 288 noms, 174 verbes, 90 adjectifs et 54 adverbes. Plus de la moitié des noms, 74% des verbes, près de 64% des adjectifs et l'ensemble des adverbes font partie de la Liste K1.

2.2.4 Le recyclage du vocabulaire thématique

Dans quelle mesure le vocabulaire thématique a-t-il été recyclé ?

La mémoire joue clairement un rôle clé dans l'apprentissage du vocabulaire et les avantages de la révision et de la répétition ont été démontrés dans les études sur l'apprentissage du vocabulaire. Kachroo (1961, cité dans Nation, 1990) a constaté que si les mots étaient répétés sept fois ou plus dans un manuel de cours, les élèves les apprendraient. A contrario, s'ils figuraient une ou deux fois seulement dans le livre, ils ne seraient pas appris par la plupart de la classe. Saragi et al. (1978) recommandent au moins 10 rencontres pour chaque nouveau mot pour faciliter l'apprentissage. Les manuels devraient donc recycler les mots de manière approfondie et, s'ils ne le font pas, le programme devrait prendre le relais. Comment cela peut-il être fait ? Le concept de « répétition élargie » s'est avéré important.

Cela suggère que l'apprenant devrait réviser le nouveau matériel peu de temps après sa première rencontre et devrait ensuite le rappeler à des intervalles progressivement croissants. Une proposition spécifique suggère des sessions de révision cinq à dix minutes après la période d'étude ; puis 2,4 heures plus tard, puis une semaine plus tard, un mois plus tard et, enfin, six mois plus tard (Russell, cité dans O'dell, 1997). Il s'agit clairement d'un point clé que les concepteurs de programmes doivent garder à l'esprit à deux niveaux ; les apprenants doivent être habilités en étant sensibilisés aux avantages d'apprentissage d'une telle méthode, mais le

programme d'enseignement lui-même doit également intégrer la répétition à des intervalles croissants.

Pour répondre à la question de recherche, nous avons recensé quelques mots appartenant au vocabulaire thématique traité en 1^{ère} AM afin d'examiner leurs occurrences dans chaque séquence didactique. La grande partie de ces mots font partie des listes thématiques de mots isolés au script¹⁴. Il est à noter qu'il n'y a aucune instruction ou recommandation qui explique l'utilisation de ces listes figurant au début de chaque séquence didactique. A notre connaissance, la méthode la plus répandue dans l'enseignement de ces listes consiste à faire apprendre le sens de chaque mot en langue maternelle, ce qui se classe parmi les techniques d'association de paires (*paired associate learning*). Van Parreren (cité dans, Bogaards, 1994) a farouchement combattu cette approche. Voici les arguments qu'il avance contre cette approche :

1. L'apprentissage par listes bilingues fait obstacle au contact direct avec la L2. Alors que la lecture en L2 suppose que l'apprenant reste autant que possible dans l'autre langue, les listes bilingues le ramènent continuellement dans sa langue maternelle.
2. Les listes bilingues suggèrent une correspondance parfaite entre les significations des mots dans les deux langues. Comme elles comportent surtout des mots isolés, il est très difficile de comprendre leur sens réel.
3. L'ordre fixe des mots sur la liste pousse les apprenants à tout mélanger : les mots qui présentent une ressemblance ou une relation quelconque « perdent leurs caractéristiques distinctives lorsqu'ils sont proches les uns des autres dans la mémoire, surtout lorsqu'ils sont présentés et appris dans la même situation d'apprentissage ».
4. Apprendre des listes bilingues est inefficace et même « parfaitement dénué de sens », parce que les mots collent les uns aux autres comme des grains de sable. Qu'ils soient présentés sur des listes alphabétiques ou regroupés en centres d'intérêt, comme « à la poste », « à l'hôpital », etc., les mots n'ont pas de prise dans la mémoire et sont pulvérisés par d'autres matériels à apprendre.

¹⁴ Un script est « un énoncé d'actions et d'événements habituels d'une situation stéréotypée ayant une structure qui conduit à un but, avec une spécification du rôle des participants. » (R. De Beaugrande, cité dans Vigner, 1993).

5. L'apprentissage par listes bilingues n'est pas motivant. C'est une activité qui ne fournit jamais de réponse aux questions que se posent les apprenants.
6. Beaucoup de mots qui sont connus dans le contexte de la liste sur laquelle ils ont été présentés, ne fonctionnent pas dans les situations de communication, c'est-à-dire aux niveaux concrets de l'emploi langagier.

Pourtant, ces arguments n'ont pas eu beaucoup d'écho, ni en théorie ni en pratique. L'utilisation de la langue maternelle est devenue une pratique systémique très répandue banalisée par la grande partie des enseignants qui cèdent facilement au « chantage linguistique » des apprenants brandissant la banderole de « Monsieur, on ne comprend rien » et invitant ainsi les enseignants à emprunter le chemin le plus court et le plus aisé qui consiste en l'utilisation de la langue maternelle dialectale.

Le tableau 2-4 montre le nombre total d'occurrences des mots appartenant aux champs thématiques au long de toutes les séquences ainsi que et la répétition de ces mots dans chaque séquence didactique.

Tableau 2-4 : Occurrences des mots appartenant au vocabulaire thématique dans les huit textes supports du manuel

Lemme	Fréquence dans le corpus	Fréquence dans la séquence pédagogique								Nombre de répétition au long des séquences
		S1	S2	S3	S4	S5	S6	S7	S8	
eau	8		2				3		3	3
santé	7		1	5		1				3
énergie	6							6		1
boire	5		4						1	2
maladie	4	1		2		1				3
air	4					4				1
pollution	3					3				1
polluer	3					3				1
lavage	3	2							1	2
vitesse	3				3					1
glace	3						3			1
aliment	3	3								1
environnement	3					1		1	1	3
humain	3	1	1	1						3
rapide	3				2				1	2
légume	2		2							1
repas	2	2								1
climat	1					5				2
obèse	0									0
corps	0									0
science	0									0
progrès	0									0
moderne	0									0
chimique	0									0
catastrophe	0									0
déforestation	0									0
destruction	0									0
respirer	1						1			1
renouvelable	2							2		1

Il ressort du tableau 2-4 que le mot le plus répété est le mot « *eau* » avec une fréquence de 8 occurrences. Ce même mot est également le mieux recyclé en 3 rappels. Le mot « *santé* » est cité 7 fois dans notre corpus mais à seulement 3 occasions. Le mot « *humain* » marque 3 apparitions dans 3 séquences consécutives. Nous pensons que nous sommes loin des recommandations de Gairns et Redman (1986, p. 94) qui suggèrent que les nouveaux mots doivent être répétés ; après un jour, une semaine, un mois et après six mois. Nation (2001) et Milton (2009) indiquent aussi que les chances d'apprendre un mot dans une langue seconde ou étrangère augmentent lorsqu'il est recyclé cinq fois ou plus.

Si le recyclage des mots n'est pas toujours systématique dans les manuels, cela peut expliquer l'effet que Vassiliu a relevé plus haut, que plusieurs centaines de mots, présents dans le manuel de cours, semblent ne jamais être appris. Ce sont peut-être les mots qui ne sont pas recyclés. Cet effet du recyclage sur la mémorisation est mieux illustré par la courbe d'Ebbinghaus.

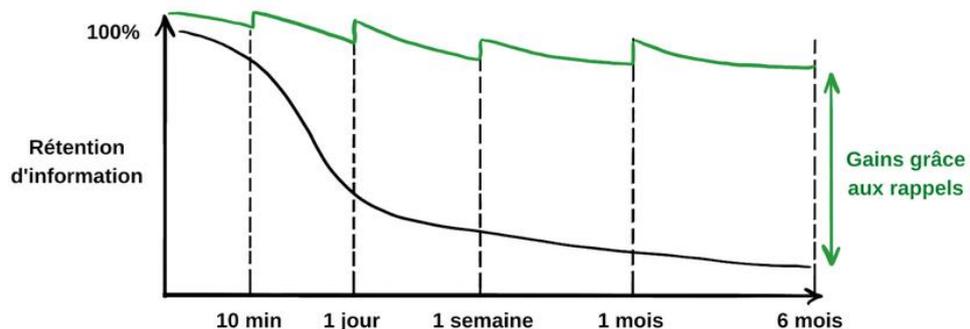


Fig. 2-5 : Courbe d'Ebbinghaus

Comme le dit Ebbinghaus (Cite dans Bogaards, 1994) : « *Si le réapprentissage est exécuté une deuxième fois, une troisième fois ou davantage, les listes sont gravées plus profondément dans la mémoire et disparaissent moins facilement, et elles finissent (...) par devenir des possessions de l'âme qui sont tout aussi disponibles que d'autres listes qui peuvent être significatives et utiles.* »

2.3 DISCUSSION

Les résultats de la première sous-question ont révélé que les mots appartenant aux listes K1 et K2 avaient le pourcentage le plus élevé de couverture lexicale dans tous les textes supports utilisés dans le manuel de 1^{ère} AM, soit une moyenne de 67,53 %. Cette hégémonie des deux premières listes aurait normalement favorisé une meilleure compréhension des textes et conséquemment une appropriation plus conséquente du nouveau vocabulaire. Cependant, nous remarquons une présence au-dessous de la moyenne de la liste K2 avec seulement 9,11 %. Le pourcentage des mots K2 est très faible bien qu'ils soient constitués de mots de contenu et soient plus essentiels pour le sens que les Mots K1. Nous voyons clairement qu'il devrait y avoir une augmentation du pourcentage des mots du deuxième millièm. Un autre résultat intéressant est que les mots hors liste (MHL) sont exceptionnellement trop utilisés avec le pourcentage de 19,25 %. Il devrait y avoir une diminution du pourcentage de mots hors liste car les mots hors liste peuvent être des noms propres, des mots peu courants, des abréviations, des acronymes, des mots appartenant à différents dialectes, un vocabulaire spécialisé, causant ainsi des problèmes pour la rétention du vocabulaire des apprenants. Bernard Toresse (1980, p. 20) s'intéresse sur l'utilité d'enseigner des mots rares à des jeunes enfants natifs : « ...*en revanche, certains vocables sont plus rares, voire exceptionnels; à l'école élémentaire, il s'agit d'étudier seulement les mots les plus courants, ceux qui sont utiles aux enfants : enseigner à des élèves du C.E.1, par exemple, que la queue de la cerise se nomme le pédoncule, ou que l'oie « cacarde », c'est littéralement perdre son temps, parce que les mots « pédoncule » et « cacarde » n'appartiennent même pas au vocabulaire d'un adulte cultivé.* »

Aussi, en analysant le comportement de chaque liste de mots le long des huit textes, nous avons constaté la même tendance globale dans presque tous les textes étudiés avec une présence plus marquée de la liste K1 dans le texte de la 4^{ème} séquence intitulé « Quel est le train le plus rapide ? ». Nous avons pu également remarquer le taux élevé des mots hors liste. Ce taux oscille entre 7,08% dans le texte de la 4^{ème} séquence et 25,83 % dans le texte de la huitième séquence.

La liste K2 est timidement représentée dans le texte de la séquence 6 avec seulement un taux de 4,90 %. Sa proportion atteint son apogée dans le texte de la S7 avec un pourcentage maximal de 13,69 %. A première vue, une augmentation des mots K2 au détriment des mots K3 et des MHL pourrait être souhaitable. Premièrement, les élèves en 1^{ère} AM sont censés avoir appris cette langue pendant 3 ans. Après 3 années d'enseignement, une part considérable du vocabulaire de la K1 est censée avoir été apprise, ou du moins effleurée. Ainsi, une réduction

des mots K1 à la recherche d'une plus grande présence de mots K2 ne serait pas excessivement absurde. Deuxièmement, une réduction de la liste MHL serait également plus que justifiée si l'objectif principal était d'apprendre les mots les plus fréquents. À cet égard, il n'est pas nécessaire dans l'immédiat d'inclure autant de mots AWL et, en plus, de les étiqueter comme cible.

L'analyse lexicométrique nous a permis de connaître les 100 mots les plus fréquents dans les textes supports. Il ressort de cette analyse que la liste des 100 mots les plus fréquents est dominée par des mots fonctionnels. Il est également à noter que paradoxalement des mots apparaissant dans les listes du vocabulaire cible n'ont pas fait partie de cette liste.

La quatrième sous-question de recherche traite d'un autre principe important de l'introduction de nouveaux apports de vocabulaire, qui est la répétition. Des études sur l'acquisition d'une L2 ont souligné l'importance du recyclage des mots nouvellement appris (Nation, 1990 ; Web, 2007). Offrir aux apprenants les expositions répétées nécessaires contribuera à améliorer leur apprentissage des mots. Le nombre suggéré pour recycler les mots nouvellement introduits varie de 5 à 20 fois. Pourtant, les résultats de la quatrième sous-question de recherche ont révélé que les nouveaux mots sont insuffisamment répétés. Cette insuffisance concerne surtout les mots relevant du vocabulaire thématique. Cette lacune se retrouve également dans l'étude menée par Matsuoka et Hirsh (2010), dans laquelle ils ont procédé à une analyse lexicale d'un livre de cours, intitulé *New Headway Student's Book Upper-Intermediate*. Les résultats ont montré qu'un grand nombre de mots n'étaient présentés qu'une seule fois dans le manuel de cours. Cette exposition limitée des mots diminue les chances d'apprendre, augmentant le besoin d'interventions pédagogiques. Le développement de matériel supplémentaire ciblant l'utilisation de ces mots rarement répétés pourrait potentiellement aider les apprenants à reconnaître les mots et à se familiariser avec leurs caractéristiques, y compris leur forme, leur partie du discours et leur signification.

Ces résultats nous amènent à penser que notre étude n'est pas un cas isolé et qu'il semble y avoir une présence significative de mots hors liste dans les matériels didactiques. Ainsi, nous sommes d'accord avec Bell et Gower (1998) que le langage présenté dans de nombreux manuels a peu de rapport avec l'utilisation réelle et plus avec la convention des manuels. Lorsque Bell et Gower parlent de convention de manuel de cours, ils ne font référence à rien de moins qu'à la tentative des manuels de cours pour répondre aux besoins et fonctions de communication des enfants qui, selon Bell et Gower, ne s'accommodent pas toujours de la fréquence.

Notre étude s'est intéressée à la structure lexicale des textes supports du manuel. Il en ressort de premier abord, que les textes supports n'introduisent pas adéquatement un vocabulaire accessible pour des apprenants souffrant de carence lexicale. En vérité, un livre de cours destiné à des apprenants de très bas niveau s'intéresserait d'abord à combler la déficience lexicale des apprenants en se focalisant sur les 2000 mots les plus fréquents. Il est possible que les textes considérés comme bons à certains égards par le concepteur du manuel de 1^{ère} AM ne le soient pas pour beaucoup. Day (cité dans Milton, 2009), par exemple, rapporte que le matériel authentique, destiné à intéresser et à motiver les apprenants et qui introduit un grand nombre de nouveaux éléments lexicaux, peut dépasser la compréhension des apprenants et manquer du type de répétition significative qui est nécessaire à l'apprentissage. Peut-être aussi, le matériel peut-il être conçu pour développer un autre aspect de la performance linguistique, comme la fluidité ou la profondeur de la connaissance du vocabulaire, plutôt que d'élargir l'étendue de la connaissance du vocabulaire

2.4 CONCLUSION PARTIELLE

Il est apparu de notre deuxième étude que les conclusions que nous pouvons tirer de notre analyse lexicographique des textes supports du manuel de la 1^{ère} AM corroborent notre 2^{ème} hypothèse qui stipule que l'une des causes des difficultés liées à l'appropriation du vocabulaire peut être attribuée à la disconvenance des textes supports proposés aux élèves de la 1^{ère} AM. Ces textes, rappelons-le, contiennent beaucoup de vocabulaire inhabituel et peu fréquent, ce qui ne convient pas à des apprenants déficitaires de niveau 1. Devenant frustrés par le nombre de mots qu'ils ignorent et/ou qu'ils doivent rechercher dans un dictionnaire, les apprenants sont dépassés par la difficulté des textes auxquels ils sont confrontés. La densité de mots inconnus est bien supérieure au niveau recommandé de 95 % à 98 % de mots connus nécessaires à la compréhension et à l'apprentissage (voir Schmitt, 2008). Entant qu'enseignant, j'estime que ce n'est pas convenable, c'est même insensé d'enseigner à un petit collégien des mots tels que « réchauffement », « tri » alors qu'il ignore le sens des mots comme « ciel », « fils » ou « avec ».

Nous sommes conscient que nous avons très peu de bonnes informations sur la charge de vocabulaire des manuels de cours et des discussions en classe, l'organisation de ce vocabulaire et la manière dont ces facteurs sont liés à l'apprentissage. C'est un domaine qui demande une étude plus systématique. En l'absence de cela, nous nous retrouvons avec des affirmations en

grande partie non étayées sur la façon dont les choses devraient être faites : la répétition est bonne voire essentielle, et introduire 8 à 12 mots par leçon est à peu près correct. Cependant, ces problèmes sont complexes et le calcul de l'effet du manuel de cours et du discours de l'enseignant sur l'apprentissage du vocabulaire est assombri par d'autres facteurs qui peuvent affecter l'apprentissage du vocabulaire présenté aux apprenants et dont nous avons également très peu de connaissances. Certaines informations potentiellement utiles ont cependant émergé.

Milton (2009) estime que dans les vrais manuels de cours, il y a beaucoup de variété. Ces livres peuvent présenter une quantité étonnamment importante de vocabulaire, des milliers de mots en peu de temps, même pour les débutants absolus. Et ils peuvent également choisir parmi une très grande variété de domaines thématiques, de sorte que les mots les plus fréquents sont souvent enseignés avec des quantités substantielles de vocabulaire peu fréquent. Si cela semble loin d'être idéal, alors le produit de ces livres, l'apprentissage du vocabulaire par les élèves, suggère le contraire. Au moins au collège, il semble que les apprenants se montrent dans la plupart du temps incapables de s'imprégner de ce vocabulaire. Il semble également qu'en présence d'une grande densité de mots inconnus, l'apprenant peut être submergé et inapte à utiliser, à son avantage, les indices disponibles.

Il est peut-être vrai que dans d'autres milieux plus favorables, ce genre de volume et de variété d'exposition du vocabulaire pourrait permettre aux bons élèves d'être mis au défi et de bien progresser, sans écraser absolument les moins capables. Les enseignants réussis semblent trouver un équilibre entre le vocabulaire fréquent et non fréquent qui permet aux apprenants de bien progresser. Milton (2009) pense que cet équilibre semble être d'environ 50 % de vocabulaire fréquent et 50 % de matériel thématique peu fréquent dans les premiers manuels de cours. Il ajoute que cela permet au texte de conserver les caractéristiques du langage normal au moins en ce qui concerne la couverture par les mots les plus fréquents. Et comme nous l'avons bien constaté, ces proportions préconisées par Milton n'ont pas été retrouvées parmi les textes supports du manuel de 1^{ère} AM.

Les cours réussis ne recyclent pas non plus systématiquement tout leur vocabulaire comme on le pense généralement essentiel pour l'apprentissage. Bien qu'il existe une relation entre la répétition de mots et l'apprentissage, Milton (2009) estime qu'il semble tout simplement peu pratique de recycler chaque élément de vocabulaire le nombre de fois requis tout en conservant un manuel de cours utile. Malgré la relation entre la répétition et l'apprentissage, il semble qu'un grand nombre de mots qui ne sont présentés qu'une seule fois soient encore

appris. Cela peut être lié aux caractéristiques des mots eux-mêmes et certains peuvent être particulièrement mémorables, mais peuvent également être liés à d'autres choses que font les apprenants qui réussissent, souvent en dehors de la classe, comme apprendre des listes de mots. Les manuels de cours réussies gèrent également l'ajout de vocabulaire et introduisent progressivement des charges de vocabulaire peu fréquentes de plus en plus importantes, de sorte que les textes que les apprenants utilisent deviennent de plus en plus lexicalement sophistiqués. Les textes élémentaires peuvent être relativement peu sophistiqués tandis que les textes plus avancés peuvent être assez difficiles, même pour les locuteurs natifs. Cette dernière observation peut sembler une évidence de bon sens, vous ne voulez pas submerger les débutants, mais elle remet en cause une adhésion trop dogmatique à l'authenticité des matériaux en classe. Des matériaux authentiques peuvent, bien sûr, être utilisés en classe, mais ce serait un avantage s'ils étaient sélectionnés avec soin pour les bonnes caractéristiques de vocabulaire.

Les bons manuels peuvent varier énormément afin de refléter les intérêts des apprenants. Ce sont des outils qui n'ont de vie et de sens qu'en présence d'un enseignant. Ils ne sont jamais destinés à être le carcan d'un programme d'enseignement dans lequel l'enseignant ne prend aucune décision d'ajouter, d'animer ou de supprimer. Le fait que les supports de cours soient parfois traités trop étroitement - par exemple, en raison du manque de temps de préparation des enseignants, des excès de pouvoir du ministère ou de l'institution, des exigences des examens ou du manque de formation professionnelle - ne doit pas être utilisé comme une raison pour radier les manuels de cours mondiaux.

Il semble probable que le vocabulaire fonctionnel et structurel doive être enseigné tôt, mais la plupart des éléments de vocabulaire peuvent être ajoutés à tout moment répondant aux intérêts des élèves. Peut-être qu'une bonne croissance du vocabulaire est un produit de cette variété.

L'aspect de la contribution orale des enseignants suggère que cela aussi peut être beaucoup plus riche que nous ne le pensions, mais il n'est toujours pas clair comment la présentation lexicale orale et orthographique peut être combinée le plus avantageusement pour accélérer l'apprentissage. Il est intéressant de noter que l'enquête de Donzelli (2007) a été en mesure de refléter l'importance de l'apport du discours de l'enseignant et l'apport du manuel de cours quant à l'apprentissage lexical en classe. Milton (2009) estime que l'enseignante qui a fait l'objet d'étude de Donzelli a réussi à fournir un environnement lexical étonnamment riche à ses apprenants en instruisant oralement une quantité de vocabulaire équivalente ou presque à

celle contenue dans le manuel scolaire. Ses apprenants ont réussi et peut-être que cette pratique en était, en partie, responsable.

Dans l'ensemble, l'impression est donnée qu'un environnement d'apprentissage en classe de qualité peut contribuer très directement à l'apprentissage du vocabulaire et à de très grandes quantités de celui-ci. Nous pensons que Harris et Snow (2004) et Ellis (1994) ont probablement tort de suggérer que la salle de classe joue peu ou pas de rôle dans l'apprentissage du vocabulaire. Pour de nombreux apprenants, il semble que la classe et le manuel soient très importants dans leur apport et puissent être une condition de réussite. Cela ne signifie pas nécessairement qu'il n'y a pas d'apprentissage en dehors de la salle de classe et à partir d'apports oraux et informels mais cela représenterait des cas très particuliers comme celui d'un apprenant évoluant dans un milieu où la présence du français est assez banale.

2.5 LIMITES DE RECHERCHE

La collecte de données quantitatives et qualitatives dans la présente étude nous a aidé à obtenir des informations plus détaillées, permettant une meilleure compréhension de la nature des mots inclus dans les textes. Malgré le fait que l'utilisation d'une conception de méthode mixte dans la recherche a aidé à surmonter les faiblesses généralement trouvées dans une méthode, un certain nombre de limitations importantes ont été trouvées dans l'étude. Pour commencer, il y a la problématique de la représentativité des textes supports destinés aux séances de compréhension de l'écrit et de lecture entraînement. Nous ignorons dans quelle mesure ces textes sont réemployés et réinvestis dans d'autres activités. Autrement dit, nous n'avons pas une idée sur les proportions de la présence de ces textes dans la panoplie des sources d'input dans une séquence didactique intégrale bien que nous considérons toujours que ces textes soient une source principale d'input.

Il y a aussi les différents aspects sur lesquels nous aurions pu nous pencher pour déterminer dans quelle mesure la répétition des mots inclus dans les textes facilite l'apprentissage des élèves. Au lieu de cela, nous avons seulement tenté d'analyser la fréquence des occurrences de mots, que les mots soient présentés cinq fois ou plus dans les textes. Aucune attention n'a été accordée à la mesure de la diffusion des mots dans les textes. De nombreux chercheurs en acquisition de la mémoire et du vocabulaire ont convenu que l'introduction de nouveaux mots sur une plus longue période de temps aidera les apprenants à construire leur vocabulaire ainsi qu'à apprendre des mots plus facilement.

Des limites existent également dans la présente étude pour la généralisabilité de l'analyse qualitative. Seuls quelques mots apparaissant au moins cinq fois dans les textes ont été sélectionnés au hasard pour l'étude de cas. Les résultats obtenus à partir de cette analyse ont révélé que les opportunités d'apprentissage ne sont pas adéquatement offertes aux collégiens pour approfondir leur connaissance de ces mots arbitrairement sélectionnés. Cependant, même avec ces résultats, il n'est pas pratique de les généraliser à d'autres mots apparaissant très fréquemment dans les textes, car certains peuvent avoir une forme d'inflexion ou une collocation, par exemple.

CHAPITRE 3 :

IMPLICATIONS ET RETOMBEES PEDAGOGIQUES

3.1 Introduction

Les mesures de la connaissance du vocabulaire s'appuient sur des théories et des modèles du lexique mental, et sur la façon dont l'apprentissage se produit, pour leur fournir une validité (Milton, 2009, p. 238) [Traduction libre]. Milton (2009) cite que les modèles, comme le modèle de fréquence de l'apprentissage, sont utilisés pour justifier la sélection du vocabulaire que nous utilisons pour les tests, et soutenir par conséquent la validité et l'utilité des informations qu'ils fournissent. Il estime aussi qu'un test qui fournit systématiquement des résultats incompréhensibles peut s'appuyer sur un modèle d'apprentissage inapproprié et tester les mauvaises choses. Nous cautionnons entièrement l'idée de Milton qui stipule que les mesures de vocabulaire fournissent également des résultats qui peuvent nous renseigner sur la validité de ces théories et des théories plus générales de l'acquisition du langage. Si, par exemple, les mesures d'acquisition du vocabulaire nous disaient systématiquement que le lexique le plus fréquent n'a pas tendance à être appris tôt au cours de l'apprentissage, alors nous devrions remettre en question le modèle de fréquence. L'intention de ce chapitre sera donc de revenir au point de départ de notre étude et de reconsidérer ce que les résultats des mesures de vocabulaire peuvent nous dire sur l'exactitude et la validité des théories sur lesquelles nous aimerions nous appuyer pour mieux rentabiliser l'enseignement/apprentissage du vocabulaire basé sur l'aspect de la fréquence.

Il y a un certain nombre d'idées à considérer ici. L'une est la question du stockage lexical : comment organiser les mots dans le cerveau, y accéder et les utiliser ? Utilisons-nous vraiment le bon modèle pour tester les mots sous la bonne forme afin d'arriver à des conclusions utiles sur l'apprentissage ? Une seconde est la façon dont nous nous appuyons sur les données de fréquence. Les chercheurs dans le domaine du vocabulaire s'appuient sur ces informations à un degré inhabituel et bien plus que les chercheurs dans d'autres domaines de l'acquisition du langage. La construction des tests de vocabulaire a souvent tendance à ignorer d'autres caractéristiques des mots et de l'apprentissage des langues en général, ce qui pourrait affecter

l'apprentissage. Les tests ainsi construits caractérisent-ils réellement les apprentissages de manière satisfaisante ? Une troisième question réside dans l'hypothèse que toutes les théories tendent à supposer que tous les apprenants sont susceptibles de se comporter de la même manière. Que peuvent nous dire les mesures de vocabulaire sur le degré d'uniformité dont font preuve les apprenants dans l'acquisition d'une langue ? Que peut-il nous dire sur les niveaux de connaissances et de performances que nous devrions attendre des apprenants ?

Nous essayerons à travers le présent chapitre d'examiner ces questions à tour de rôle, puis d'aborder d'autres points et questions qui ont également été soulevés.

3.2 Le vocabulaire, élément incontournable dans l'apprentissage des langues

Les tests de vocabulaire les plus largement utilisés ont tendance à s'appuyer sur des listes de mots isolés qui ont été lemmatisés ou réduits à des familles de mots. Ces mots peuvent être testés isolément ou en contexte, de manière réceptive ou productive. Mais l'hypothèse est que si nous testons une forme d'un mot, alors les apprenants sont susceptibles de connaître d'autres formes d'un mot dans le même lemme ou la même famille de mots et cela reflète la façon dont les apprenants d'une langue étrangère (L2) traitent et stockent réellement les mots : une forme de base avec des inflexions et des dérivations régulières. Est-ce correct ? Les apprenants stockent-ils vraiment leurs mots et les traitent-ils de cette façon ?

Milton (2009) pense que les informations dont nous disposons suggèrent globalement que les flexions et les dérivations les plus fréquentes et régulières, qui formeraient généralement un lemme, ont tendance à être apprises tôt, tandis que les dérivations moins fréquentes sont ajoutées plus tard en cours d'apprentissage. Ce type de découverte donne du crédit à des théories plus larges de l'apprentissage des langues, telles que la théorie de la *processabilité* de Pienemann. Le modèle de Pienemann, qui est basé sur la grammaire fonctionnelle lexicale (Kaplan & Bresnan, cité dans Milton, 2009) et sur la description procédurale de Kempen et Hoenkamp (cité dans Milton, 2009) de la génération du langage, part également du principe que l'acquisition de la grammaire est motivée par l'apprentissage des mots. Dans son modèle, l'accès aux lemmes est la première étape de la hiérarchie de *processabilité*, sans laquelle aucun autre apprentissage ne peut avoir lieu. Milton cite que dans une version récente de la grammaire générative, *The Minimalist Program*, les différences entre les langues sont principalement de nature lexicale. Ainsi, Cook (cité dans Milton, 2009) suggère que le programme minimaliste est conduit par le lexique. Cela a des implications pour l'acquisition du langage, car cela

suggère que la tâche à laquelle l'apprenant est confronté est principalement l'apprentissage du lexique. Bogaards va jusqu'à dire que « *C'est qu'il est souvent tout à fait possible d'utiliser à bon escient un mot sans se servir de sa syntaxe.* » (1994, p. 39)

C'est ce que font souvent les touristes, ils apprennent quelques éléments de la langue d'accueil mais ignorent les règles syntaxiques.

Il est donc évident que l'acquisition de ces éléments lexicaux en quantité suffisante déclenche l'établissement de paramètres grammaticaux. Cette approche se reflète dans l'hypothèse d'apprentissage lexical (Ellis, 1997), selon laquelle la connaissance du vocabulaire est indispensable pour acquérir la grammaire.

« *Learning grammar involves abstracting regularities from the stock of known lexical sequences.* » (Ellis, 1997, p. 6)

Ces théories placent l'apprentissage des mots au cœur même de l'apprentissage des langues ; quelque chose sans lequel l'acquisition d'autres aspects du langage ne peut pas se produire. De ce fait, le vocabulaire semble être une partie tout à fait indispensable de l'apprentissage des langues et où le nombre d'éléments de vocabulaire enseignés pourrait être limité à ceux nécessaires pour illustrer les structures linguistiques.

Milton constate que dans les mesures de l'étendue du vocabulaire décrites dans son livre *Measuring Second Language Vocabulary*, un chiffre a surgi à maintes reprises, ce qui donne une idée de l'échelle des exigences d'apprentissage du vocabulaire de la langue. En anglais, un seuil de 2000 à 3000 mots est récurrent. Il couvre environ 80% de la plupart des textes, c'est le minimum requis pour permettre aux apprenants d'acquérir une sorte de compréhension générale et pour permettre les débuts d'une communication indépendante dans des situations linguistiques authentiques. Ce chiffre peut varier d'une langue à l'autre, mais l'implication que les programmes d'études devraient enseigner et exiger dans les tests de gros volumes de vocabulaire est celle qui est générale à toutes les langues. Milton (2009, p. 240) précise que, pour atteindre une réelle fluidité, les apprenants ont besoin de milliers de mots supplémentaires et, en anglais encore une fois, les chiffres avoisinant les 10 000 mots qui sont courants. Dans cette optique, certaines finalités fréquemment adoptées de l'enseignement des langues peuvent être remises en cause.

En Algérie, que ce soit au collège ou au lycée, les enseignants enseignent souvent les stratégies de compensation pour tenter de contourner la déficience en vocabulaire. Ces stratégies ne sont pas et ne peuvent pas être un substitut adéquat à la connaissance du vocabulaire puisqu'en présence d'un grand nombre de mots inconnus, les apprenants peuvent être dépassés et incapables d'utiliser les indices disponibles à leur avantage.

Si cela semble une leçon évidente à souligner, il faut se rappeler qu'il existe certains programmes de langues, et le programme algérien de langues étrangères dans nos écoles, collèges et lycées en est un exemple, qui semblent l'avoir oublié.

Les apprenants sont très dépendants de la langue à laquelle ils sont exposés dans les manuels et en classe. Nous pensons que l'adhésion à l'idée que le matériel à enseigner doit être authentique et que par conséquent les textes proposés dans les manuels sont relativement longs et substantiels, a généré une forme de langage qui affiche en termes de fréquences de mots, quelque chose de très différent du langage normal. Les enseignants algériens, produit de cette tendance, et qui partent en France pour un stage ou pour un voyage touristique sont vite "fichés" comme des gens parlant « comme des dictionnaires ».

Il semble probable qu'un vocabulaire très fréquent et de nombreuses formes de mots fréquents manquent. Les apprenants ne peuvent pas apprendre ce qu'ils ne rencontrent pas. La langue normale que les apprenants devront maîtriser pour devenir fluide ne leur est pas illustrée en quantité et par conséquent, elle n'est pas apprise. L'authenticité des matériels peut couvrir un large éventail de choses, et dans l'enseignement des langues, rien ne remplace des matériels bien écrits pour les apprenants qui illustrent la langue dans les limites des connaissances des apprenants et les mettent suffisamment au défi de se développer. Il se peut que les matériels d'enseignement des langues étrangères en Algérie ne soient pas aussi performants, d'où les résultats pas très flatteurs dans les épreuves de français tous niveaux confondus.

3.4 La fréquence du vocabulaire comme modèle d'apprentissage

Une deuxième hypothèse qui sous-tend les mesures que nous avons de l'apprentissage du vocabulaire, est que les apprenants sont particulièrement sensibles à la fréquence des mots et que cela se reflètera dans la manière dont les mots sont ajoutés au lexique L2. Les tests s'appuient sur des listes de fréquences car les mots fréquents sont plus susceptibles d'être rencontrés par les apprenants. Pour apprendre un mot, vous devez le rencontrer, et vous êtes

plus susceptible de rencontrer ces mots fréquents, et de les rencontrer à plusieurs reprises, plus que des mots moins fréquents. Cela semble très simple et pourtant l'idée pose des problèmes à de nombreux linguistes universitaires. Tschichold (cité dans Milton, 2009) est un exemple récent qui remet en question la validité des informations de fréquence remises par rapport à l'apprentissage et aux tests de langue. Tschichold inclut d'autres facteurs qui affecteraient l'estimation de la fréquence. Un de ces facteurs est l'âge du corpus ; on s'appuie encore souvent sur la liste de mots *The General Service List* de West (1953) pour l'anglais et sur la liste du linguiste Jean Baudot (1992) pour le français. Des questions sont également soulevées sur la façon de gérer les unités multi-mots, la polysémie et l'unité de comptage. À cette liste pourrait être ajoutée l'influence des facteurs de difficulté des mots tels que la longueur et la parenté, qui devraient tous atténuer l'impact de la fréquence des mots qu'un apprenant acquiert.

Malgré toutes ces préoccupations, Milton (2009) soutient que les tests qui utilisent des informations de fréquence se révèlent remarquablement robustes et réalisables. Il semble que les tests basés sur le lemme comme unité de comptage donnent des résultats satisfaisants, encourageants qui démontrent clairement un effet de fréquence : les lemmes fréquents sont réellement appris en plus grand nombre que les lemmes moins fréquents. Ce fait n'est pas influencé de manière significative par l'âge du corpus ou sa taille, une fois que vous dépassez environ un million de mots dans un corpus raisonnablement bien construit. Les mots les plus fréquents dans un corpus ne sont pas susceptibles de changer radicalement au fil du temps, même avec l'ajout de matériel supplémentaire à un corpus. McGavigan (cité dans Milton, 2009) suggère que l'effet de fréquence s'étend même à des unités à plusieurs mots très peu fréquentes. Tout comptage cohérent basé sur une définition raisonnablement intelligente de ce qui est compté semble révéler cet effet. Cet effet de fréquence est si fiable que les caractéristiques de difficulté des mots, communément acceptées comme influentes pour déterminer si un mot sera appris ou non, n'ont pas d'impact significatif sur lui (Milton & Daller, 2007). Malgré les doutes, Milton persiste sur l'importance de la fréquence dans la détermination des matériels d'apprentissage des langues.

« ... *the importance of frequency in vocabulary learning is as near to a fact as it is possible to get in L2 acquisition.* » (Milton, 2009, p. 242)

Une conclusion qui en est tirée est l'importance de la répétition à l'apprentissage ou au moins à l'apprentissage du type d'automatisation de l'utilisation qui est important pour une utilisation fluide de la langue (Hilton, 2008). La répétition porte encore beaucoup de

connotations négatives. Hilton (2008, p. 162) la décrit même comme « une hérésie pédagogique » pour de nombreux enseignants. Cela est dû à l'exercice plutôt stupide et douloureux qui caractérisait la méthode audio-linguale et que de nombreux apprenants trouvaient si insatisfaisant. Pourtant, comme le démontre Vassiliu (cité par Milton, 2009), un bon matériel d'enseignement recycle régulièrement beaucoup de matériel lexical très fréquent et ce matériel est ensuite appris par la plupart des apprenants. Des études récentes sur l'acquisition et le traitement de la langue seconde (par exemple, Segalowitz 2003) ont inévitablement réhabilité ces principes cognitifs fondamentaux : il n'y a pas de mémorisation sans répétition (dans aucun domaine, y compris le développement de la L1), et beaucoup du langage (le lexique mental L1, par exemple) est une information déclarative efficacement stockée dans la mémoire à long terme. En Algérie, et sous l'influence de la mouvance de l'approche communicative qui tourne le dos aux préceptes béhavioristes, les exercices répétitifs visant la systématisation sont plus ou moins absents de nos classes de langue.

Devrions-nous peut-être regarder de plus près, et à un niveau théorique, l'importance du redoublement dans l'enseignement. Il y a un paradoxe à affirmer l'importance de la répétition dans l'apprentissage du vocabulaire car il est clairement impossible de répéter et de recycler systématiquement chaque élément de vocabulaire. Il n'y a pas assez de temps, même dans le cours de langue étrangère le plus généreusement programmé, pour que cela se produise. Et pourtant, les gens apprennent de grandes quantités de vocabulaire en langues étrangères et utilisent couramment des mots relativement peu fréquents qu'ils semblent n'avoir rencontrés qu'une ou deux fois. La preuve des manuels réussis est qu'un large éventail du lexique le plus fréquent est recyclé, tandis que le vocabulaire spécialisé, thématiquement lié et moins fréquent a tendance à ne pas l'être. C'est peut-être le meilleur compromis qui puisse être obtenu, et permet de rencontrer et d'automatiser des formes assez normales de mots fréquents et leurs combinaisons. Le lexique moins fréquent nécessite une contribution plus individuelle de la part des apprenants eux-mêmes. Les apprenants qui réussissent, semble-t-il, emploient des stratégies, comme les systèmes de mots-liens pour faciliter la mémoire, mais créent également leurs propres opportunités d'exposition et de répétition ; même s'il ne s'agit que de répéter et d'apprendre des listes de mots. Les preuves suggèrent que cela peut être un moyen très efficace de développer un lexique L2 important.

3.5 L'apprentissage du vocabulaire et les différences individuelles

Milton (2009) émet l'hypothèse qui stipule que tous les apprenants sont à peu près les mêmes et suivront les mêmes types de processus dans l'apprentissage des langues. Nous supposons que tous les apprenants seront mieux en mesure d'apprendre des matériaux répétés, par exemple. Bien sûr, cela ne doit pas être le cas. L'une des caractéristiques des mesures de l'apprentissage du vocabulaire présentées ici est la grande variété d'apprentissages et de progrès réalisés par les apprenants. Les apprenants peuvent s'asseoir dans la même classe, être exposés au même matériel, faire les mêmes exercices et encore arriver à des résultats d'apprentissage de vocabulaire très différents. Certains apprenants font d'énormes progrès et d'autres non. Il peut y avoir plusieurs raisons à cela. Les apprenants sont motivés différemment, par exemple, et vous ne pouvez pas faire apprendre une langue étrangère à quelqu'un qui ne le souhaite pas. Néanmoins, alors que nous comprenons quelque chose sur la façon dont les groupes d'apprenants développent le vocabulaire dans une langue étrangère, nous ne comprenons vraiment pas et ne pouvons pas encore prédire comment les individus se comporteront. Nous n'avons pas lié de manière satisfaisante l'apprentissage du vocabulaire avec les explications des différences individuelles dans l'apprentissage pour produire un modèle qui puisse rendre compte de la variation dans l'apprentissage du vocabulaire. Tout ce que nous savons, c'est que la variation dans l'apprentissage des langues est normale et que le type de variation que nous voyons dans l'apprentissage du vocabulaire, qui peut être quantifié, est énorme. Il est évident que différentes aptitudes peuvent prédisposer les apprenants à acquérir des types de vocabulaire subtilement différents.

Les études d'aptitudes à l'apprentissage des langues ont une longue histoire, mais ont rarement été spécifiquement appliquées à l'apprentissage d'un lexique. Dans le très influent test d'aptitude au langage moderne (MLAT) de Carroll et Sapon (cité dans Milton, 2009), l'aptitude est divisée en quatre sous-éléments : capacité de codage phonémique, sensibilité grammaticale, capacité d'apprentissage inductif des langues et capacité d'apprentissage par cœur. Le dernier, l'apprentissage par cœur, est ici intéressant car il est testé via un test de mémorisation par cœur d'associés de vocabulaire : 40 paires de mots sont présentées, l'une dans la langue maternelle et l'autre dans une langue inconnue, et celles-ci doivent être mémorisées en peu de temps. Les personnes testées sont ensuite invitées à se rappeler autant de ces paires de mots que possible. Plus les apprenants peuvent en retenir, plus leur mémoire est grande et plus leur aptitude est grande. On suppose que la mémoire des gens, en particulier, variera et cela devrait faire des

différences dans une tâche d'apprentissage telle que l'apprentissage de milliers de mots de langues étrangères. Mais la capacité de mémoire par cœur à elle seule ne peut probablement pas expliquer la nature de la variation dans l'apprentissage du vocabulaire. Les volumes de vocabulaire appris pourraient être les mêmes, mais la nature du vocabulaire connu pourrait varier considérablement d'un individu à l'autre. Certains apprenants sont décrits comme des apprenants déficitaires de niveau 2, où la connaissance de la deuxième bande de fréquence de 1000 mots est plus faible que prévu, tandis que d'autres ont un déficit structurel et ne connaissent pas les 1000 mots les plus fréquents. Les profils produits par ces apprenants sont illustrés à la Figure 3-1 et à la Figure 3-2.

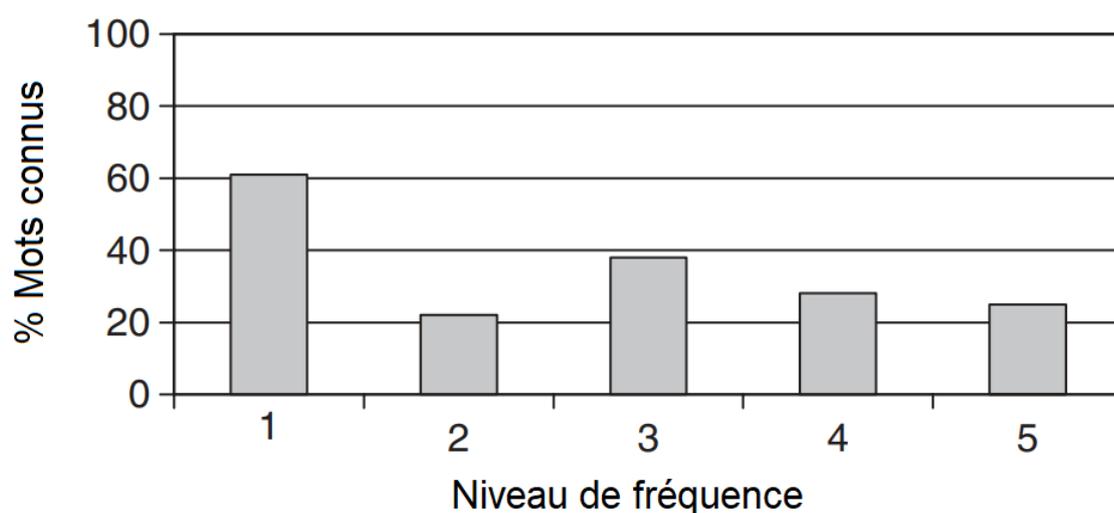


Fig. 3-1 : Profil de fréquence illustrant un déficit de niveau 2 selon Milton (2007a)

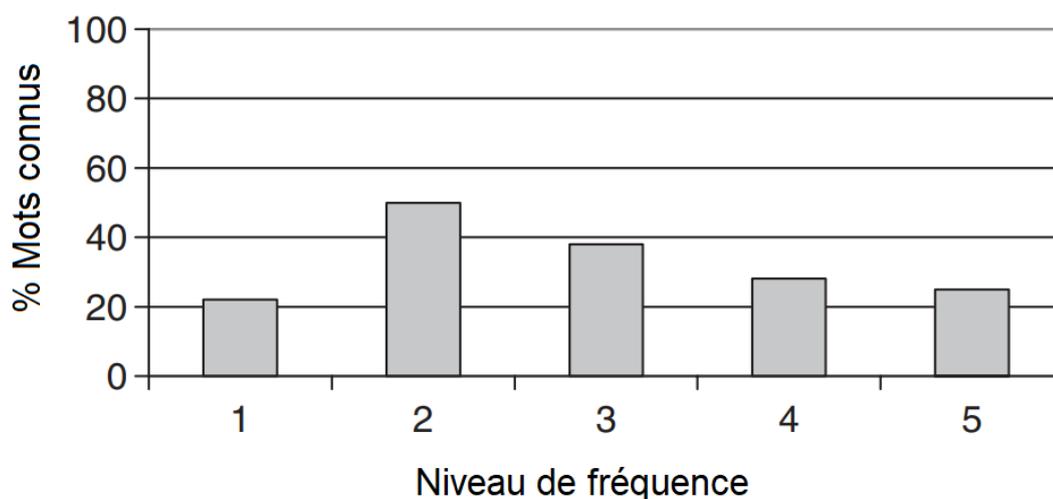


Fig. 3-2 Profil de fréquence illustrant un déficit structurel Milton (2007a)

Milton (2007a) a enquêté sur 21 apprenants grecs qui ont passé deux tests du Meara et al. (2001) gamme de tests d'aptitude. Il s'agissait de LAT_B, une tâche d'apprentissage d'associés appariés conçue pour tester la mémoire dans l'apprentissage des langues, et LAT_C, une tâche de reconnaissance de règles linguistiques conçue pour tester les compétences d'apprentissage inductif et analytique des langues. Les apprenants ont été regroupés selon leurs profils, 10 profils normaux et 11 déficitaires de niveau 2, et leurs scores à ces tests d'aptitude sont calculés. Les scores moyens sont présentés à la figure 3-3.

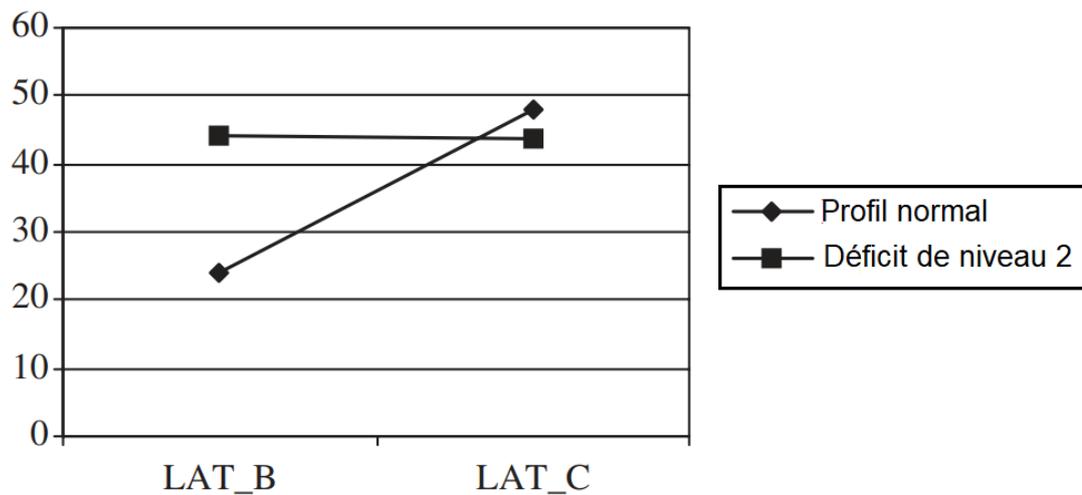


Fig. 3-3 : Scores moyens aux tests d'aptitude pour deux groupes d'apprenants avec des profils de vocabulaire différents (Milton, 2007a, p. 55)

Il apparaît que les différences d'aptitudes peuvent influencer la nature du vocabulaire appris. La différence de score moyen au test de mémoire est particulièrement marquée et les apprenants ayant obtenu un score plus élevé au test de mémoire, LAT_B, ont tendance à afficher un déficit de niveau 2. Les apprenants qui affichent des profils normaux ont comparativement moins bien réussi au test de mémoire, bien que leurs scores au test analytique soient légèrement plus élevés. Les résultats suggèrent qu'il y avait à la fois des effets de groupe et des effets de test qui étaient statistiquement significatifs. Ces résultats suggèrent que différentes forces et styles d'apprentissage peuvent réellement influencer le vocabulaire des langues étrangères que les apprenants acquièrent en classe, et que cet effet est particulièrement visible dans le degré d'apprentissage du vocabulaire dans les deux premières bandes de fréquence de 1000 mots. Ceci étant dit, il apparaît que l'importance de la fréquence se réaffirme dans les profils globaux que présentent ces deux types d'apprenants.

Les apprenants peuvent varier d'autres manières. L'une est la manière dont ils développent l'automatisme dans l'utilisation de la langue étrangère. Par exemple, Kroll et al. (2006) soulignent que même les bilingues fluides seront plus lents dans les tâches de dénomination d'images par rapport aux locuteurs unilingues. Il s'agit d'un domaine d'étude intéressant, car le travail ici suggère un moyen de gérer les variations individuelles en donnant un sens au type de mesures que cette étude produit. Segalowitz et F. Freed (2004) notent que les individus peuvent varier dans la vitesse et l'automatisme avec lesquelles ils lisent et reconnaissent les mots dans leur langue maternelle (L1), et cela a pour effet de saper les tentatives de bon usage des mesures d'automatisme dans l'apprentissage de la L2. La variation individuelle de l'automatisme du traitement de texte a pour effet de masquer le type de progrès systématique que les apprenants feront à mesure qu'ils améliorent leurs connaissances et leur maîtrise de leur langue étrangère. Les apprenants de langues étrangères deviennent plus automatiques dans leur utilisation de la langue à mesure qu'ils s'améliorent, mais il semble qu'il n'y ait pas de niveau de performance défini qui caractérise les utilisateurs avancés de langues par rapport aux utilisateurs moins compétents, car même les apprenants les plus compétents varieront énormément dans cette facette de la performance. La façon dont Segalowitz (2004) gère cela est de prendre une mesure de l'automatisme des apprenants dans leur L1 pour fournir une ligne de base par rapport à laquelle les performances et les progrès dans la L2 peuvent être mesurés. Les progrès des apprenants dans leur L2 peuvent alors être mesurés par rapport à leur propre performance en L1 plutôt que par rapport à une seule norme de performance externe qui peut ne pas être appropriée ou utile. Les résultats que cette approche produit semblent donner beaucoup plus de sens à l'apprentissage.

Ce type d'approche pourrait nous aider à comprendre d'autres aspects de l'apprentissage du vocabulaire où la variabilité est une caractéristique importante des tests. Même quelque chose comme les tailles de vocabulaire suggérées par Milton sont des exigences pour certains niveaux de performance et pourraient être reconsidérées dans cette optique. Nous avons tendance à supposer que tous les locuteurs natifs ont un vocabulaire très étendu et que les apprenants se compareront mal à ce niveau de connaissance jusqu'à ce qu'ils puissent acquérir un vocabulaire étendu. Le chiffre de 17 000 familles de mots de Goulden et al. en moyenne pour les locuteurs natifs a tendance à être considéré comme une norme qui caractérisera tous les locuteurs natifs anglais, par exemple. Cependant, Milton (2009) souligne que sa propre utilisation du test de reconnaissance suggère qu'il existe une variation importante et assez systématique autour de ce chiffre. Selon Milton, les étudiants de premier cycle de langue maternelle qui entrent à

l'université ont tendance à marquer environ 9 000 mots à ce test, les étudiants de troisième cycle environ 13 000 à 14 000 et les membres du corps professoral plus de 20 000. Même parmi les étudiants britanniques de premier cycle, il existe des variations, car les anglophones unilingues obtiennent en moyenne 9500 tandis que les bilingues gallois-anglais obtiennent environ 8000 avec une déficience marquée dans le domaine de la liste de mots académiques de Coxhead. Mais ce sont tous des locuteurs natifs qui fonctionnent à un niveau élevé grâce à l'anglais. Et tandis que les chiffres pour le vocabulaire nécessaire à la couverture et à la compréhension des textes ne peuvent pas être écartés, peut-être, comme pour l'automatisme, nous devrions nous attendre à ce que les apprenants développent un vocabulaire dans une langue étrangère équivalente en échelle à leur L1 plutôt que d'aspirer à une taille unique qui même les locuteurs natifs peuvent ne pas atteindre dans de nombreux cas. Cela serait certainement conforme à la réflexion actuelle de Fitzpatrick sur les tâches d'association de mots. Pour être plus clair, l'élargissement du stock lexical en langue maternelle inciterait à un élargissement parallèle du stock lexical de la langue cible.

3.6 La variation individuelle et la dimension de la profondeur du vocabulaire

Un autre aspect de la variation individuelle est la manière dont les individus gèrent les associations de mots. Cela a des implications sur la façon dont nous percevons la connaissance des mots chez l'apprenant L2 et sur l'importance de développer la profondeur du vocabulaire en plus de l'étendue du vocabulaire. L'hypothèse sous-jacente à l'organisation du lexique chez les apprenants de L2 est qu'il est quantitativement et qualitativement différent de celui des locuteurs natifs, et qu'à mesure que l'apprenant grandit en compétences, connaissances et aptitudes, le lexique deviendra plus natif. La preuve de l'étude de Fitzpatrick (2006) est qu'il n'y a probablement pas une seule structure ou taille de type natif à laquelle les apprenants peuvent se référer ; les locuteurs natifs varieront dans leurs préférences pour les associations et leur organisation. L'analyse de la nature des associations entre les items lexicaux (Henriksen, 2008) suggère en outre que bon nombre de ces liens ne sont pas particuliers à une première ou à une deuxième langue, mais sont généraux à la langue, de sorte que la tâche dans la L2 est de développer une étendue suffisante de vocabulaire que ces liens peuvent être démontrés. Bon nombre de liens soigneusement notés et comptés dans les tests de profondeur de vocabulaire, tels qu'ils sont démontrés dans les travaux de Wolter (cité dans Miton, 2009), sont peu susceptibles de nous en dire beaucoup sur le développement de la dimension de profondeur L2.

Ils peuvent nous informer sur le nombre de mots qu'un apprenant possède, afin qu'ils puissent produire les associations générales de langue décrites par Henriksen, ou ils peuvent nous parler de la façon dont un apprenant progresse vers la reconstruction de ses préférences d'association L1 qui, en l'absence d'informations détaillées sur les réponses d'une association L1 d'apprenants, nous en dira peu sur leur développement.

Milton (2009) pense que cela nous dit quelque chose sur la façon dont nous devrions voir cette dimension de la profondeur du vocabulaire qui peut s'avérer être une idée moins utile et moins informative que nous ne l'avons pensé jusqu'à présent. Le très large éventail de liens qui sont généralement inclus dans cette dimension peut nous en dire très peu sur le lexique L2, car ceux-ci ne peuvent imiter que le lexique L1. Tester l'existence de ces liens dans la L2 peut seulement nous dire si ces mots existent dans le lexique L2 et rien de nouveau sur l'état du lexique. Il y a d'autres problèmes, tels que la gestion des unités de plusieurs mots qui, suggère McGavigan, peuvent être apprises et gérées de telle manière qu'elles soient traitées comme des lexèmes uniques et devraient de toute façon être réellement incluses dans l'étendue du vocabulaire. Milton (2009) pense que, du point de vue de l'apprentissage de la L2, l'intérêt est vraiment beaucoup plus étroit que la plupart des définitions de la profondeur ne le suggèrent, et il ajoute que cet intérêt se focalise uniquement sur les aspects où les liens sont différents de la L1. Cela pourrait signifier que la profondeur est utilement limitée aux collocations telles que celles que Gyllstad teste, où votre L1 est peu susceptible de vous dire quel verbe (de *make*, *take* ou *do*) vous auriez besoin de lier avec *bath*, ou des connotations où les significations secondaires des mots peuvent être différents. Ceci, à son tour, rend très difficile la création d'un test général d'approfondissement du vocabulaire pour les apprenants de toutes les langues.

3.7 Variation individuelle, âge et difficulté des mots

Traditionnellement, les apprenants étaient soit des adultes matures, soit des adolescents à l'école, mais de plus en plus, l'apprentissage des langues est systématiquement présenté à des apprenants beaucoup plus jeunes. Certaines des hypothèses que nous faisons sur les apprenants eux-mêmes et la nature du processus d'apprentissage devraient changer en conséquence.

En effet, Milton estime qu'il existe une différence très importante entre les très jeunes apprenants et les apprenants adultes. Cette différence consiste en le degré d'alphabétisation qu'ils possèdent. Les apprenants adultes d'une langue étrangère seront pour la plupart des

lecteurs compétents et expérimentés, capables d'utiliser des stratégies telles que la prédiction d'idées et de mots qui apparaîtront dans un texte et la reconnaissance des mots par leur forme plutôt que par leur orthographe. Les jeunes apprenants seront très différents et seront beaucoup moins à l'aise à la fois dans la reconnaissance des mots et le passage du mot au sens. Les jeunes apprenants, semble-t-il, sont susceptibles d'être beaucoup plus dépendants de la langue orale que les adultes. Il semble inévitable que ces deux types d'apprenants traitent, apprennent et stockent les mots de manière très différente. Milton (2009) note que les débutants ont tendance à privilégier l'apprentissage des mots phonologiques par rapport à l'orthographe, suggérant qu'ils favorisent le stockage phonologique. Une partie du problème sous-jacent ici est peut-être que les débutants ont tendance à être beaucoup plus jeunes et qu'ils doivent stocker de nouveaux mots dans une langue étrangère de cette façon, car ils n'ont pas l'arsenal de compétences et d'expérience pour les stocker orthographiquement.

Pour les jeunes apprenants algériens, l'absence d'unités phonétiques en arabe ou en berbère par rapport au système phonétique du français génèrent souvent des situations de confusion quant aux concepts stimulés par les images acoustiques. En effet, pour beaucoup d'apprenants, les mots "*sœur*", "*sourd*" ou "*sûr*" par exemple, peuvent avoir un seul concept.

Néanmoins, et au fur et à mesure que les apprenants améliorent leurs compétences en langues étrangères, ils deviennent inévitablement plus âgés et gèreront mieux la forme orthographique des mots dans les deux ou dans toutes leurs langues. La tendance observée chez les apprenants à passer d'un lexique phonologiquement chargé à un lexique fortement chargé orthographiquement peut donc être une caractéristique de l'âge et du développement cognitif autant que du niveau de langue.

La même idée devrait nous obliger à remettre en question certaines des notions longtemps chéries de ce qui rend les mots faciles ou difficiles. Milton (2009) note que les caractéristiques des mots telles que la parenté et la longueur, supposées affecter la difficulté et l'apprentissage, se sont révélées insaisissables lorsque leur impact sur l'apprentissage d'un lexique entier est examiné. La cognition exige que les apprenants aient un vocabulaire étendu en L1 et qu'ils aient la perspicacité de reconnaître qu'un mot de la L1 peut avoir une traduction similaire dans la L2 s'ils choisissent de traiter la langue de cette manière. Milton se montre sceptique à l'idée que les très jeunes apprenants gèrent la tâche d'apprentissage du vocabulaire avec ce degré de perspicacité ou de sophistication. Milton (2009) conclut que toutes les théories que nous

utilisons dans l'apprentissage des langues devront être repensées si elles doivent s'appliquer aux jeunes apprenants avec la même pertinence qu'aux adultes.

3.8 Implications pour la pratique de l'enseignement du vocabulaire

L'un des objectifs de la mesure de l'assimilation du vocabulaire et des progrès doit être de réinjecter des informations dans le processus d'apprentissage et d'enseignement afin d'optimiser le temps consacré à ces activités.

La principale leçon qui se dégage de l'étude de l'apprentissage du vocabulaire dans les langues étrangères est que les apprenants doivent apprendre beaucoup de vocabulaire s'ils veulent atteindre un niveau de communicabilité indépendant. Les volumes peuvent varier légèrement d'une langue à l'autre, mais des milliers de mots sont nécessaires et il faudra des centaines d'heures pour réaliser ce type d'apprentissage. Il serait probablement avantageux pour toutes les personnes impliquées dans le processus d'apprentissage des langues que certains de ces objectifs soient explicités dans les cours et les programmes, afin que les apprenants puissent commencer à comprendre où ils se situent par rapport à ces objectifs, plus tangibles, de l'apprentissage des langues. Cela semble être un domaine où il n'y a pas de raccourcis à l'apprentissage.

Les apprenants qui réussissent, et par conséquent les bons cours, parviennent à organiser l'acquisition du vocabulaire de manière à ce qu'il soit appris en quantités régulières sur de longues périodes de temps. Il semble que ce soit une caractéristique des manuels de cours réussis qu'ils incluent, sans surprise à la lumière de ce qui précède, beaucoup, beaucoup de vocabulaire, et un nouveau vocabulaire est introduit à intervalles réguliers. Une caractéristique des cours réussis en EFL est qu'ils présentent une grande variation thématique et peut-être que cela permet aux apprenants d'être suffisamment engagés et motivés pour permettre l'apprentissage, tout en offrant l'occasion d'introduire un nouveau vocabulaire. Il ne semble pas y avoir de limite supérieure (dans des limites raisonnables) aux quantités qui peuvent être présentées et aux volumes que les bons apprenants apprendront, à condition que le matériel permette de revoir de temps à autre certains des sujets et le vocabulaire qu'ils contiennent. L'une des caractéristiques des manuels en langue française est qu'ils manquent de cette grande variété de couverture thématique et qu'ils présentent ainsi un peu moins de vocabulaire. Pour beaucoup d'élèves, les thèmes abordés en première année moyenne, par exemple, semblent beaucoup plus abstraits et ne suscitent par conséquent que peu d'intérêt. Ce n'est peut-être pas

un hasard si l'apprentissage du vocabulaire des langues étrangères est si problématique dans nos écoles.

C'est une norme de bonne pratique d'enseignement que les formes orales et écrites des nouveaux mots soient présentées aux apprenants et les témoignages de vrais apprenants montrent à quel point la connaissance de la forme phonologique et orthographique des mots est importante pour la réussite des examens ultérieurs. Mais l'évidence montre aussi que ces examens nécessitent, à la longue, une connaissance du vocabulaire orthographique plutôt supérieure à celle phonologique pour réussir. Il semble tout à fait satisfaisant pour les apprenants d'apprendre simplement la reconnaissance visuelle d'un vocabulaire très peu fréquent.

Pourtant, Il semblerait être une croyance répandue que les apprenants ne retiennent pas beaucoup de vocabulaire de ce qui est enseigné explicitement, du manuel, en classe. Cette croyance est dangereusement trompeuse, car les apprenants semblent apprendre une grande partie du vocabulaire le plus fréquent et beaucoup de vocabulaire peu fréquent, explicitement à partir du contenu des manuels de cours. Mais les livres ne se limitent pas aux mots les plus fréquents. Une caractéristique des manuels de cours réussis est qu'il semble y avoir un mélange égal des matériaux thématiques les plus fréquents (les 2000 mots les plus fréquents dans les corpus) et les moins fréquents. Il semble que les volumes et la répétition du vocabulaire fréquent s'accumulent au fil des ans, de sorte que les apprenants parviennent à en couvrir et à maîtriser la majeure partie. La recherche suggère également que les apprenants n'acquièrent pas de mots implicitement sans effort conscient, comme cela a été suggéré, mais ils peuvent, en utilisant des tâches informelles qu'ils aiment et où l'apprentissage du vocabulaire est une cible de l'activité, acquérir d'énormes quantités de mots à partir d'activités comme écouter des chansons, lire des bandes dessinées et regarder des films en langue étrangère. C'est une bonne pratique de préconiser ces activités depuis de nombreuses années, mais la recherche suggère qu'elles peuvent être beaucoup plus efficaces dans l'apprentissage du vocabulaire qu'aucun d'entre nous ne l'avait réellement imaginé. Une combinaison de bonnes pratiques en classe et d'efforts bien dirigés en dehors de la classe peut commencer à expliquer comment les apprenants acquièrent un très grand vocabulaire de milliers de mots, afin qu'ils puissent atteindre une fluidité fonctionnelle. L'idée que les enseignants n'ont pas à enseigner et que les apprenants n'ont pas à essayer d'apprendre du vocabulaire est cependant absurde. Cela demande un effort délibéré et un temps considérable de la part de l'apprenant.

Il y a très peu de preuves sur l'impact de la contribution orale de l'enseignant à l'apprentissage, mais cela doit être important lorsqu'il s'agit d'une source principale d'entrée pour la réalisation phonologique des mots que les apprenants acquièrent. L'idée reçue est qu'un environnement lexicalement riche est bon pour l'apprentissage. Et l'enseignant de Donzelli¹⁵, qui a réussi à illustrer les mots du manuel et à aller au-delà pour fournir un environnement très riche en est l'exemple.

La recherche liant la taille du vocabulaire aux examens et le Cadre européen commun de référence pour les langues (CECR) commence à s'adapter aux chiffres indicatifs pour ces niveaux et normes, aidant à fournir aux apprenants et aux enseignants des objectifs plus clairs dans la gestion du processus d'apprentissage. Passer de 2000 à 2500 mots (sur 5000) en anglais semble être un seuil pour passer du niveau débutant au niveau intermédiaire, où l'usage de la langue peut commencer à devenir autonome. De 6000 à 7000 mots sont nécessaires pour la maîtrise orale et de 8000 à 9000 pour la maîtrise écrite et pour atteindre le type de maîtrise requis pour les examens au niveau C2 du CECR. Dans ce contexte, Milton (2009) pense qu'il y a beaucoup à dire pour étendre l'utilisation des tests informels et à faible enjeu. L'avantage des tests qui sont rapides et faciles à administrer et qui sont faciles à noter est que le stress, les dépenses et les perturbations causés par les tests formels à enjeux élevés peuvent être évités, tout en permettant d'accéder à des commentaires utiles sur le niveau et les progrès.

¹⁵ Dans une étude publiée en 2007 ; Donzelli a analysé l'apport oral en langue étrangère, reçu en classe par un groupe d'élèves de CM1. Cette analyse a été effectuée en fonction des listes de fréquences élaborées par Nation (1986). L'étude a comparé également l'apport du discours de l'enseignant avec l'apport du manuel de cours et a offert quelques indications sur l'assimilation des apprenants. Cette étude a suggéré que si le discours de l'enseignant ainsi que l'apport écrit disponible à partir du manuel représentaient des environnements lexicaux tout aussi riches, la quantité de vocabulaire produit par le premier serait nettement supérieure à celle disponible à partir du second. Les données collectées par Donzelli ont montré qu'une grande partie du vocabulaire peut être acquise par les apprenants en classe.

CONCLUSION GENERALE

Nous avons tenté à travers cette thèse de cerner quelques difficultés inhérentes à l'enseignement/apprentissage du vocabulaire au collège, particulièrement pour la classe de 1^{ère} AM. A l'instar d'autres études sur l'enseignement/apprentissage du vocabulaire, notre étude suggère que plus nous comprenons le processus d'apprentissage du vocabulaire, plus il semble être important pour l'ensemble du processus d'apprentissage des langues. L'apprentissage du vocabulaire n'est pas quelque chose qui peut être mis de côté, et le volume de vocabulaire ne peut pas non plus être réduit à des niveaux infimes pour s'adapter à la commodité d'un cours, si la capacité de communiquer dans la langue étrangère est le but de l'apprentissage. La présentation systématique du vocabulaire, et en grande quantité, devrait fournir un environnement pour la création de programmes de langue très différent de ce qui a été le cas au cours des 50 ou 60 dernières années, bien que ces idées mettent un certain temps à se frayer un chemin dans les pratiques enseignantes en classes de langue. Les mesures que nous sommes en mesure de faire de l'étendue du vocabulaire en particulier suggèrent que le vocabulaire pourrait également avoir un rôle important à jouer dans l'évaluation de la langue, car la croissance du vocabulaire est un facteur principal dans le développement des compétences de communication. Il devrait également jouer un rôle beaucoup plus important dans l'évaluation, car il peut ajouter un degré d'objectivité au processus d'évaluation linguistique qui est uniquement subjectif et impressionniste.

Les mesures du vocabulaire ont également suggéré l'importance des études de corpus et de fréquence à la fois dans l'apprentissage et dans la conception de tests pour mesurer l'apprentissage. En utilisant des tests basés sur la fréquence, il est possible de suggérer avec une certaine confiance ce que les groupes de connaissances de vocabulaire d'apprenants posséderont après certaines quantités d'entrée de langue ou à des niveaux de performance spécifiés. Ceci est rassurant, mais ne doit pas nous faire oublier que la fréquence ne peut pas prédire ou expliquer tout l'apprentissage du vocabulaire et, en particulier, nous avons très peu à expliquer comment les individus peuvent varier même lorsque les groupes se comportent de manière relativement prévisible. Peut-être que le centre d'attention dans la mesure de

l'apprentissage du vocabulaire devrait passer des études de groupe à celle de l'individu pour aider à combler cette lacune.

Les mesures que nous avons du vocabulaire ont remis en cause certaines des dimensions que nous avons utilisées pour essayer de caractériser commodément la connaissance du vocabulaire. En particulier, la profondeur du vocabulaire peut ne pas être une idée aussi utile ou perspicace qu'elle est devenue populaire dans les études de vocabulaire au cours des dernières années. Milton (2009, p. 75) précise que la prise en compte d'autres aspects de la connaissance du vocabulaire et des performances mériterait une enquête plus systématique.

Milton (2009) suggère aussi que des objectifs de vocabulaire clairs et à grande échelle devraient être une caractéristique principale de tout bon programme d'enseignement. Le processus de mesure de l'apprentissage du vocabulaire nous permet d'être assez sûrs qu'il s'agit d'une exigence de réussite de l'apprentissage des langues. Nous avons également une idée de la manière dont les mots pourraient être présentés de manière optimale aux apprenants et de la manière dont ils doivent être impliqués afin d'accélérer l'apprentissage. Cependant, l'étude du contenu des manuels de cours et des discours des enseignants en classe n'en est qu'à ses balbutiements et nous avons encore du chemin à parcourir avant de vraiment comprendre comment et pourquoi les apprenants acquièrent exactement les mots qu'ils font, tout en n'apprenant pas les autres. Nous sommes conscient que les enseignants ressentent souvent le besoin d'enseigner explicitement les connaissances et, à la lumière des doutes exprimés sur la validité conceptuelle de la dimension, nous nous demandons si le temps consacré explicitement à ces tâches, au détriment de l'élargissement du lexique, serait du temps bien dépensé. Une condition pour améliorer la profondeur du vocabulaire, quelle que soit sa définition, est d'augmenter la largeur du vocabulaire pour fournir suffisamment de mots pour qu'un réseau complexe s'établisse et se développe. Pour nous, le problème de l'enseignement/apprentissage du vocabulaire est une combinaison d'un contenu lexical inapproprié avec une présentation peu attractive qui concurrencerait mal la tentation des nouvelles technologies .

Bien qu'il ne nous permette pas de tirer des conclusions générales par rapport à la taille du vocabulaire réceptif d'un collégiens de 1^{ère} année et des retombées pédagogiques qu'il en découle, du fait qu'il est restreint, notre travail contribuerait en revanche à inciter les chercheurs à se pencher davantage sur ce sujet car plus de recherches sont effectivement nécessaires pour savoir plus sur les causes de cette déficience lexicale observée chez les élèves au collège et au lycée et parfois même à l'université. En outre, et pour que la boucle soit bouclée, des recherches

pourraient aussi porter sur les deux autres paliers du cycle moyen, à savoir les 2^{ème}, 3^{ème} et 4^{ème} années moyennes. Cette étude invite également les praticiens à consacrer plus de temps à des activités visant à enrichir le vocabulaire et d'adopter une approche cohérente et systématique. Elle invite également les concepteurs du matériel didactique à prendre en considération les « axiomes » linguistiques actuelles pas trop flatteuses pour l'apprentissage du français en général, car nous savons pertinemment que ce qui est un bon manuel pour un concepteur n'est pas nécessairement un bon manuel pour un enseignant et encore moins pour un élève. Malgré nos décisions imprévisibles en tant que praticiens, le matériel qui nous est proposé devrait toujours être utilisable par des enseignants moins formés, non formés ou formés différemment. Les choses doivent être précisées à l'enseignant inexpérimenté sans condescendance avec l'enseignant expérimenté car, comme le soulève Häcker (2008), il existe actuellement un cercle vicieux d'enseignants qui produit des apprenants ayant une connaissance insuffisante des langues étrangères, dont certains deviennent plus tard des enseignants, de sorte que chaque nouvelle génération d'enseignants est, en moyenne, moins compétente que la précédente.

Il y a aussi un aspect que nous devons « tacler » dans cette conclusion. C'est le choix des thèmes. Notre expérience nous avait montré que trop de textes sur des sujets tels que l'environnement, l'alimentation et le changement climatique ne seraient pas attrayants pour un adolescent de 11 ou 12 ans. Alors qu'un grand nombre d'élèves semblaient s'intéresser à des sujets quotidiens tels que l'argent, les relations, les vêtements et la nourriture, beaucoup moins d'élèves dans les cours de langue générale s'intéressaient aux sujets comme l'activité physique ou le tri sélectif des ordures ménagères. Bell et Gower (2011, p. 143) préconisent même un équilibre entre les articles sérieux et « amusants ». Certes, des objectifs pragmatiques obligent mais il faut toujours savoir faire des concessions surtout quand c'est au profit des apprentissages.

Pour conclure, beaucoup de questions pourraient également être soulevées : quel contenu lexical enseigner au primaire ? Combien de mots doit connaître un élève à l'issue du cycle primaire ? Quels moyens faut-il mettre en œuvre pour développer la compétence lexicale chez les élèves. Quel apport pourrait avoir le discours de l'enseignant sur l'apprentissage lexical ? Ce genre de questions pourrait être abordé dans de futures recherches dédiées à l'enseignement/apprentissage du vocabulaire du français langue étrangère en Algérie.

BIBLIOGRAPHIE

- Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund, Y. (2008). *Corpus linguistics with BNCweb-a practical guide*. Peterlang.
- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425–438.
- Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.
- Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (ed.). *Comprehension and Teaching: Research Reviews*, 77-117.
- Anderson, R., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge dans Hutson. B. A. *Advances in reading language research. A research annual*, 2, 319-337.
- Anscombe, J.-C. (2001). Le rôle du lexique dans la théorie des stéréotypes. *Langages*(142), 57-76.
- Bell, J., & Gower, R. (2011). Writing course materials for the world: a great compromise. In B. Tomlinson, *Materials Development in Language Teaching* (pp. 135-150). Cambridge: Cambridge University Press.
- BENTOLILA, A. (2007). *Rapport de mission sur l'acquisition du vocabulaire à l'école élémentaire*. Retrieved from <https://www.vie-publique.fr/sites/default/files/rapport/pdf/074000221.pdf>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Bogaards, P. (1994). *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris: Hatier.
- Calaque, E. (2000). *Enseignement et apprentissage du vocabulaire, hypothèses de travail et propositions didactiques*. LIDIL 21.
- Calaque, E. (2000). *Enseignement et apprentissage du vocabulaire, hypothèses de travail et propositions didactiques*. LIDIL 21.

- Calaque, E., & DAVID, J. (2004). *Présentation*. Dans : Elizabeth Calaque éd., *Didactique du lexique: Contextes, démarches, supports* (pp. 7-15). Louvain-la-Neuve, Belgique: De Boeck Supérieur.
- Carver, R. (1994). Percentage of Unknown Vocabulary Words in Text as a Function of the Relative Difficulty of the text: Implications for Instruction. *Journal of Literacy Research*, 26, 413-437.
- Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In B. Laufer, & P. Bogaards (Eds.), *Vocabulary in a Second Language: Selection, Acquisition, and Testing* (pp. 15-38). Amsterdam: John Benjamins.
- Corbin, D. (1980). Compétence lexicale et compétence syntaxique. *Modèles linguistiques*, II(2), 52-138.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Cuq, J.-P. (2003). *Dictionnaire de didactique du français langue étrangère et seconde*. Paris: Paris, Cle internationa.
- Cuq, J.-P. (2004). Chapitre 4 : Le lexique en situation d'apprentissage guidé : pour une méthodologie d'enseignement interventionniste dans l'enseignement du français langue étrangère. In E. C. éd, *Didactique du lexique: Contextes, démarches, supports* (pp. 61-71). Louvain-la-Neuve, Belgique: De Boeck Supérieur. doi:<https://doi.org/10.3917/dbu.didle.2004.01.0061>
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.
- David, A. (2008). Vocabulary breadth in French L2 learners. *Language Learning Journal*, 36(2), 167-180. Retrieved from <http://dx.doi.org/10.1080/09571730802389991>
- Donzelli, G. (2007). Foreign language learners: Words they hear and words they learn, a case study. *Estudios de Lingüística Inglesa Aplicada (ELIA)* 7, 103-126.
- DUBOIS Jean, e. c. (1999). *Dictionnaire de linguistique et des sciences du langage*. Paris: Larousse. « Collection Expression ».

- Ellis, N. (1997). *Vocabulary acquisition: Word structure, collocation, word-class, and meaning*. (M. McCarthy, & N. Schmidt, Eds.) Cambridge: Cambridge University Press.
- Fitzpatrick, T., Al-Qarni, I., & Meara, P. (n.d.). Intensive vocabulary learning: a case study. *The Language Learning Journal*, 36(2), 239-248.
- Gairns, R., & Redman, S. (1986). *Working with Words: A Guide to Teaching and Learning Vocabulary*. Cambridge: Cambridge University Press.
- GOUGENHEIM, G., MICHEA, R., & RIVENC, P. S. (1959). *L'élaboration du Français fondamental, 1er degré*. Paris: Didier.
- Goulden, R., Nation, I., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics 11*, 341-363.
- Grishman, R. (1986). *Computational linguistics: an introduction*. New York: Cambridge University Press.
- Grossmann, F. (2011). Didactique du lexique : état des lieux et nouvelles orientations. *Pratiques n°149-150*(149/10), 163-183. doi:<https://doi.org/10.4000/pratiques.1732>
- GROSSMANN, F., & BOCH, F. (2003). Production de textes et apprentissage lexical : l'exemple du lexique des émotions et des sentiments. *Repères*(28), 117-135.
- Häcker, M. (2008). Eleven pets and twenty ways to express one's opinion: The vocabulary learners of German acquire at English secondary schools. *Language Learning Journal*, 36(2), 215-226.
- Harris, V., & Snow, D. (2004). *Doing it for Themselves : Focus on Learning Strategies and Vocabulary Building*. London: CILT.
- Hatch, E. (1983). *Psycholinguistics: A Second Language Perspective*. Rowley, Mass: Newbury House.
- Hatch, E. (1983). *Psycholinguistics: A Second Language Perspective*. Rowley, Mass: Newbury House.
- Hatzigeorgiu, N., Mikro, G., & Carayannis, G. (2001). Word length, word frequencies and Zipf's law in the Greek language. *Journal of Quantitative Linguistics 8* (3), 175-185.

- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36(2), pp. 153-166.
- Hindmarsh, R. (1980). *Cambridge English Lexicon*. Cambridge: Cambridge University Press.
- Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403-430.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods*, 18(1), 3-20.
- Kilgarriff, A. (2006). *BNC database and word frequency lists*. Retrieved from <https://www.kilgarriff.co.uk/bnc-readme.html>
- Kučera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Brown University Press, Providence.
- Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307-322.
- Laufer, B. (1990). Words you know: how they affect the words you learn. In J. Fisiak (ed.). *Further Insights into Contrastive Linguistics*, 573-593.
- Laufer, B. (1997). What's in a word that makes it hard or easy: some intralexical factors that affect the learning of words. (N. S. McCarthy, Ed.) *Vocabulary: description, acquisition and pedagogy*, pp. 140-155.
- Leeman, D. (2005). Le vertige de l'infini ou la difficulté de didactiser le lexique. *Le français aujourd'hui*, 148(1), 89-99. doi: <https://doi.org/10.3917/lfa.148.0089>
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation*. Guérin : Montréal.
- Lewis, M. (1993). *The Lexical approach*. London: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the Lexical Approach :Putting Theory into Practice*. Language Teaching Publications.
- Martin, O. (2009). *L'analyse des données quantitatives*. Paris: Armand Colin.

- Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language*, 22(1), 56-70.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh University Press, Edinburgh.
- McEnery, T., Wilson, A., Sanchez-Leon, F., & Nieto-Serano, A. (1997). Multilingual Resources for European Languages: Contributions of the Crater Project. In *Literary and Linguistic Computing* (Vol. 12, pp. 219-226). Oxford. doi:<https://doi.org/10.1093/lc/12.4.219>
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. Abingdon: Linguistic, Routledge Applied.
- Meara, P. (1980). Vocabulary Acquisition: A Neglected Aspect of Language Learning. *Language Teaching*, 13, 221-246. doi:10.1017/S0261444800008879
- Meara, P. (1992). EFL vocabulary tests. *Wales University: Swansea Centre for Applied Language Studies*.
- Meara, P. (1996). The Dimensions of Lexical Competence. (G. Brown, K. Malmkjaer, & J. Williams, Eds.) *Competence and Performance in Language Learning*, pp. 35-53.
- Meara, P., & Milton, J. (2003). *X-Lex, The Swansea Levels Test*. Newbury: Express.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Publishers, Jossey Bass.
- Milton, J. (2006a). X-Lex: The Swansea Vocabulary Levels Test. In P. D. C. Coombe (Ed.), *Proceedings of the 7th and 8th Current Trends in English Language testing (CTELT) Conference* (Vol. 2, pp. 29-39). UAE: TESOL.
- Milton, J. (2006b). Language Lite? Learning French Vocabulary in School. *Journal of French Language Studies*(16), 187-205. doi:10.1017/S0959269506002420
- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.). Cambridge: Cambridge University Press.

- Milton, J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. In *Modelling and Assessing Vocabulary Knowledge* (pp. 47–58). Cambridge, UK: Cambridge University Press.
- Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *Journal of French Language Studies*, 18, 333-348. doi:10.1017/S0959269508003487
- Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition*. Multilingual Matters.
- Milton, J., & Alexiou, T. (2009). Vocabulary size and the common European framework of reference for languages. In B. Richards, H.M. Daller, D. *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Application*, 194-211.
- Milton, J., & Meara, P. (1998). Are the British really bad at learning foreign languages? *Language Learning Journal* 18, 68-76.
- Milton, J., & Vassiliu, P. (2000). Frequency and the lexis of low-level EFL texts. In K. Nicolaidis, & M. Mattheoudakis (Eds.), *Proceedings of the 13th Symposium in Theoretical and Applied Linguistics* (pp. 444-455). Thessaloniki: Aristotle University of Thessaloniki.
- Nation, P. (1984). *Vocabulary Lists: Words, Affixes and Stems*. English University of Wellington: New Zealand: English Language Institute.
- Nation, P. (1990). *Teaching & Learning Vocabulary*. New York: Newbury House.
- Nation, P. (2001). *Learning Vocabulary in Another Language, First Edition*.
- Nation, P. (2004). *A study of the most frequent word families in the British National Corpus*. (P. Bogaards, & B. Laufer, Eds.) Amsterdam: John Benjamins.
- Nation, P. (2013). *Learning Vocabulary in Another Language, Second Edition*. Cambridge.
- Neveu, F. (2004). *Dictionnaire des sciences du langage*. Paris: Armand-Colin.
- Niklas-Saminen, A. (1997). *La lexicologie*. Paris: Arman colin.

- O'dell, F. (1997). Incorporating vocabulary into the syllabus. In N. Schmitt and M. McCarthy (eds). *Vocabulary: Description, Acquisition and Pedagogy*, pp. 258-278.
- O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- Orosz, A. (2009). The growth of Young learners' English vocabulary size. In M. Nikolov (Ed.), *Languages, Early learning of Modern Foreign* (pp. 181-194). Bristol: Multilingual Matters.
- Palmer, H. (1917). *The Scientific Study and Teaching of Languages*. London: Harrap. Retrieved from <https://archive.org/details/in.ernet.dli.2015.90560/page/n9/mode/2up>
- Parkinson, D. (2005). *Oxford Business English Dictionary for Learners of English*. Oxford: Oxford University Press.
- PICOCHÉ, J. (1992). *Précis de lexicologie française, L'étude et l'enseignement du vocabulaire*. Paris Nathan-Université.
- PICOCHÉ, J. (n.d.). Enseigner le vocabulaire : La théorie et la pratique. (E. VIGDOR, Ed.)
- Rayson, P. (2002). Matrix: a Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison [Thèse de doctorat]. Retrieved from <https://ucrel.lancs.ac.uk/people/paul/publications/phd2003.pdf>
- Richards, B., & Malvern, D. (2007). Validity and threats to the validity of vocabulary measurement. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 79-92). Cambridge: Cambridge University Press.
- Richards, B., & Malvern, D. (2008). Word frequency and trends in the development of French vocabulary in lower intermediate students during Year 12 in English schools. *Language Learning Journal*, 36(2), 199-214.
- Römer, U. (2010). Using general and specialized corpora in English language Teaching: Past, present and future. In M. C. Campoy-Cubillo, B. Bellés-Fortuño, & M. Gea-Valor, *based approaches to English Language Teaching* (pp. 18-38). Londres: Continuum.

- SARDIER, A. (2006). Construire la compétence lexicale : quelle place en didactique pour le cotexte? UNIVERSITÉ GRENOBLE ALPES.
- Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Studying Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173-199.
- Simard, C. (1994). Pour un enseignement plus systématique du lexique. *Québec français*(92), 28-33.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter, & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-160). London: Longman.
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36(2), 139-152.
- Tharp, J. (1934). The Basic French Vocabulary and Its Use. *The Modern Language Journal*, 19(2), 23-131.
- Tréville, M.-C. (2001). *Vocabulaire et apprentissage d'une langue seconde. Recherche et théories*. Logiques (Editions) .
- Tréville, M.-C., & Duquette, L. (1996). *Enseigner le vocabulaire dans une classe de langue*. Paris: Hachette Livre.
- Tribble, C., & Jones, G. (1997). *Concordances in the Classroom*. Athelstan, Houston, Texas.
- University of Cambridge Local Examinations Syndicate. (2001). *Cambridge Examinations, Certificates & Diploma: Regulations 2001*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Vancomelbeke, P. (2004). *Enseigner le vocabulaire*. Paris: Nathan.
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*(2), 217-234.

- Vigner, G. (1993). Le monde, les mots et l'école : Eléments d'une didactique du vocabulaire à l'école élémentaire. *Repères* 8, 191-209.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65.
- Weinreich, U. (1968). *Languages in contact: findings and problems*. La Haye: Mouton Publishers.
- Wesche, M., & Paribaht, T. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review* 53, 13-40.
- Williams, E. (1974). *Reading In The Language Classroom*. London: Macmillan Publishers Ltd.

ANNEXES

Prétest de vocabulaire X-lex (Niveau 1K)

Lisez attentivement la liste ci-dessous. Pour chacun des mots, indiquez si vous en connaissez les sens ou non en inscrivant O ou N en regard de chacun. Traduisez en langue maternelle chaque mot dont vous connaissez le sens.

- | | | |
|----------------------------------|-------------------------------------|----------------------------------|
| 1 [O] jamais ^{مستحيل} | 2 [N] gélard | 3 [N] inconnu |
| 4 [O] entendre ^{يسمعه} | 5 [O] cacher ^{يخفي} | 6 [N] tartelle |
| 7 [N] contre | 8 [N] offrir | 9 [N] s'empâtrer de |
| 10 [N] préjate | 11 [N] rencontrer | 12 [N] frais |
| 13 [N] voix | 14 [N] plusieurs | 15 [N] haine |
| 16 [N] tenir | 17 [N] agrutable | 18 [N] souver |
| 19 [N] proine | 20 [O] cheval ^{حصان} | 21 [N] théâtre |
| 22 [N] pied | 23 [N] plassart | 24 [N] diroir |
| ✓ 25 [O] chambre ^{غرفة} | 26 [O] occuper ^{قطع} | 27 [N] étranger |
| 28 [N] dernier | 29 [N] désespoir | 30 [N] prêter |
| 31 [N] culon | 32 [N] coûter | 33 [N] mallenier |
| 34 [N] commencer | 35 [O] compagnon ^{جبار} | 36 [N] riquer |
| 37 [N] front | 38 [N] plaine | 39 [N] accorder |
| 40 [N] besoin | 41 [N] violence | 42 [N] dîter |
| 43 [N] jupaire | 44 [N] inventer | ✓ 45 [O] désert ^{صحراء} |
| 46 [N] jît | 47 [N] souhaiter | 48 [N] abattir |
| 49 [N] apercevoir | 50 [N] prétenser | 51 [N] hésiter |
| ✓ 52 [O] garder ^{يحرس} | 53 [N] distance | 54 [N] soirée |
| 55 [N] ralive | 56 [N] course | 57 [N] cordre |
| ✓ 58 [O] oublier ^{ينسى} | ✓ 59 [O] supporter ^{يشتدح} | 60 [N] guirum |

6V
0F

Posttest de vocabulaire X-lex (Niveau 1K)

Lisez attentivement la liste ci-dessous. Pour chacun des mots, indiquez si vous en connaissez les sens ou non en inscrivant O ou N en regard de chacun. Traduisez dans la langue maternelle chaque mot dont vous connaissez le sens.

- | | | |
|----------------------|------------------|-----------------|
| 1 [O] soleil الشمس | 2 [N] visage | 3 [N] qui |
| 4 [O] expliquer تفسر | 5 [N] vaste | 6 [N] touceul |
| 7 [N] carbie | 8 [N] hiver | 9 [N] outrir |
| 10 [N] aige | 11 [N] blailant | 12 [N] produire |
| 13 [N] voiture | 14 [N] nouveau | 15 [N] terre |
| 16 [N] vamaille | 17 [N] corps | 18 [N] recevoir |
| 19 [N] comprendre | 20 [N] garmente | 21 [N] bonfe |
| 22 [N] perdre | 23 [N] air | 24 [N] donc |
| 25 [N] ãle | 26 [N] roude | 27 [N] salmer |
| 28 [N] faible | 29 [N] pendant | 30 [N] tuer |
| 31 [N] fule | 32 [N] changer | 33 [N] soment |
| 34 [N] savoir | 35 [N] ciel | 36 [N] diner |
| 37 [N] vivre | 38 [N] mouvement | 39 [N] puissant |
| 40 [N] picher | 41 [N] r risse | 42 [N] titre |
| 43 [N] exister | 44 [N] science | 45 [N] couge |
| 46 [N] ville | 47 [N] courir | 48 [N] vent |
| 49 [N] bacter | 50 [N] diff rent | 51 [N] parler |
| 52 [N] d fauter | 53 [N] eau | 54 [N]  norme |
| 55 [N] cher | 56 [N] octe | 57 [N] do te |
| 58 [N] vouloir | 59 [N] force | 60 [N] fatide |

1 V

0 F

Post_test

Texte support de la 1^{ère} séquence (Projet I)

PROJET

1

Séquence 1

Je comprends mon texte

Des gestes simples pour une bonne hygiène corporelle

texte

À la maison et à l'école, les adultes te parlent régulièrement de l'importance d'avoir les mains propres. Pourquoi le lavage des mains est-il si important ?

Le lavage des mains consiste à se débarrasser des saletés, des matières grasses et des microbes qui passent facilement d'une personne à une autre. De la même manière, ils passent d'un objet à une personne. Se laver les mains, se couvrir la bouche et le nez quand on tousse ou quand on éternue nous permet donc d'éviter d'être malades et de rendre les autres malades. En effet, le toussotement et l'éternuement peuvent transmettre des maladies graves.

C'est le chercheur français Louis Pasteur qui nous a le plus appris à nous protéger des microbes.

*D'après « L'hygiène des enfants en question »
Brochure sur la santé (2012)*

Je vérifie ma compréhension du texte

1. Quelle est la phrase qui se rapporte à la définition du lavage des mains ?
2. Que risquent les personnes qui n'ont pas une bonne hygiène des mains ?
3. Que faut-il faire pour éviter de rendre les autres malades ?
4. Dans le tableau suivant, coche la case qui convient.

	La phase de questionnement	La phase d'explication	La phase de conclusion
1 ^{ère} partie			
2 ^{ème} partie			
3 ^{ème} partie			



Le sais-tu ?

La définition est un des procédés explicatifs que tu as appris. En voici un exemple.

Un microbe : nom masculin, être vivant très petit qu'on ne peut voir qu'au microscope et qui provoque des maladies. (Dictionnaire Le Robert)



Texte support de la 2^{ème} séquence (Projet I)

PROJET

1

Séquence

Je comprends mon texte

Que devrions-nous manger et boire ?

texte

Comme tous les êtres humains nous sommes obligés de manger et de boire quotidiennement, mais que doit-on manger et boire au juste ?

La consommation de différentes sortes d'aliments assure une bonne santé. Un repas équilibré comprend des féculents (pâtes, riz ou pain) et cinq fruits et légumes par jour. Une petite quantité de protéines (fromage ou viande par exemple) complète le repas. Consommés en faible quantité, les graisses et les sucres sont des aliments utiles.

L'eau nous est bénéfique car notre organisme en a besoin pour fonctionner correctement. Tous les jours, nous perdons de l'eau en transpirant et en urinant.

C'est pour cela que nous buvons et mangeons des aliments contenant de l'eau, tels que des fruits, des légumes et de la soupe.

*Louise Spilsbury, «Tout savoir sur tout»,
Ed. Parragon (2008)*

Je vérifie ma compréhension du texte

1. Quelles sont les catégories d'aliments citées dans le 1er paragraphe ?
2. Quels sont les aliments que l'on consomme en principe à chaque repas ?
3. Dans le 2ème paragraphe, l'eau est présentée comme bénéfique à notre organisme. Explique pourquoi.
4. Quels sont les aliments qui contiennent de l'eau ?

Je retiens

Une bonne hygiène de vie, c'est :

- bien se laver,
- avoir une alimentation riche et variée,
- boire régulièrement de l'eau.



Le sals-tu ?

Les boissons sucrées tels que les sodas et jus sont à déconseiller à cause de leur forte teneur en sucre. Il vaut mieux boire de l'eau. >

Je comprends mon texte

« Pour votre santé, bougez ! »

texte

Pourquoi l'Organisation Mondiale de la Santé (OMS) propose aux états membres de marquer la Journée mondiale de l'exercice physique le 10 mai de chaque année ?

En encourageant l'organisation de marches, de courses de vélo, d'activités sportives de loisir et d'aménagement de parcs et d'espaces publics, l'OMS veut pousser les gens à se dépenser physiquement. Comme le déclare le Dr Pekka Puska, Directeur à l'OMS : La Journée mondiale qui a pour thème « Pour votre santé, bougez ! » insiste sur les bienfaits de l'exercice physique pour la santé et le bien-être dans le monde entier. L'objectif consiste à associer la prévention des principales maladies chroniques, la promotion de la santé et le développement social.

Pour tout être humain, l'exercice physique est un puissant moyen de prévention des maladies graves.



« Organisation Mondiale de la Santé »
Activités physiques et santé

Je vérifie ma compréhension du texte

1. Quel est le sujet du texte ?
2. Qui est l'auteur de cette initiative ?
3. Que marque la journée du 10 mai de chaque année ?
4. Explique l'expression : « Pour votre santé, bougez ! »
5. Qui est le Dr Pekka Puska ?

Je retiens

Avoir une bonne hygiène de vie, c'est :

- se laver,
- se brosser les dents,
- avoir une alimentation saine,
- boire régulièrement de l'eau,
- avoir fréquemment une ou plusieurs activités physiques.

Texte support de la 1^{ère} séquence (Projet II)

PROJET

2

Séquence 1

Je comprends mon texte

Quel est le train le plus rapide ?

texte

Les trains à grande vitesse sont de plus en plus nombreux à circuler à travers les différents continents. Mais alors pourquoi le Maglev chinois est présenté comme le train en service commercial le plus rapide du monde ?

Tout simplement parce qu'il peut atteindre une vitesse de 431 km/h. Le Maglev relie l'aéroport de Pudong à la station de Longyang Road en 7 minutes 20 secondes pour un parcours de 30,5 kilomètres. Il se déplace au-dessus de son rail sans le toucher. C'est grâce à cette technologie qu'il peut atteindre des vitesses élevées.

De nombreux trains sont utilisés par les habitants des banlieues, autrement dit les gens qui habitent en dehors de la ville. Ils font la navette pour aller chaque jour travailler dans les grandes agglomérations.

« Tout savoir sur tout, »
Éd. Parragon



Je vérifie ma compréhension du texte

1. Comment se nomme le train présenté dans le texte ?
2. Dans quel pays circule-t-il ?
3. Retrouve dans la partie explicative du texte la raison pour laquelle ce train est aussi rapide.
4. D'après le texte, qui utilise ce moyen de locomotion ?

Je retiens

Le Maglev est un train différent des autres trains. Il n'utilise pas de roues. Grâce à la technologie développée par les Japonais depuis 1964, ce moyen de transport circule sur un seul rail sans être en contact avec le sol.



Le sals-tu ?

«La définition» et «d'énumération» sont des procédés explicatifs que tu connais déjà. À présent, tu vas utiliser « la reformulation » qui consiste à reprendre une explication en employant d'autres mots. >

Texte support de la 2^{ème} séquence (Projet II)

PROJET

2

Séquence

Je comprends mon texte

Peux-tu m'expliquer la pollution de l'air ?

texte

L'air que nous respirons est de plus en plus pollué. Quelle en est la raison ?

On dit que l'air est pollué quand il est modifié par des éléments qui sont nuisibles à notre santé et à notre environnement. Ce sont des gaz, souvent invisibles, qui polluent l'air. Mais les polluants atmosphériques peuvent être d'origine naturelle (la fumée des volcans, certaines plantes, la foudre...). Malheureusement, ils sont dus pour la plupart aux activités humaines :

- transports ;
- industrie ;
- chauffage et climatisation ;
- agriculture avec l'utilisation d'engrais, de pesticides sans oublier les émissions animales ;
- incinération des déchets ;
- utilisation de produits d'entretien, de différentes colles ;
- meubles en bois aggloméré...

La conséquence de cette pollution est la multiplication des maladies respiratoires et le changement du climat.

*Clive Gifford (2013)
« Le grand livre qui explique tout »*

Je vérifie ma compréhension du texte

1. Retrouve dans le texte la définition de la pollution de l'air.
2. Quels sont les polluants de l'air d'origine naturelle ?
3. Cite les polluants de l'air dus aux activités humaines.

Je retiens

L'air devient de plus en plus chargé de polluants. Cela provoque des maladies respiratoires et bouleverse le climat.

Texte support de la 3^{ème} séquence (Projet II)

PROJET

2

Séquence

Je comprends mon texte

Que se passerait-il si toutes les glaces des pôles fondaient ?



La surface de la banquise n'a jamais été aussi réduite. C'est l'annonce inquiétante que vient de faire le National Snow and Ice Data Center.

Plusieurs études le confirment : les banquises perdent du terrain d'année en année à cause du réchauffement climatique provoqué par les différentes pollutions. 10% des terres de notre planète sont recouvertes de glace. Il s'agit des calottes glaciaires : l'Antarctique et le Groenland. Ces glaciers font par endroits jusqu'à 4000 mètres d'épaisseur. A une certaine température, ils fondent légèrement, se cassent et forment des icebergs qui partent à la dérive. L'eau des glaces fondues rejoint celle des océans, par conséquent le niveau de la mer s'élèvera partout d'une centaine de mètres.

En effet, les côtes de la Terre entière seront inondées. Certaines îles et les villes situées en bordure de mer seront englouties sous les eaux.

« Gentside découvertes »

Le magazine au cœur de notre planète (2015)

Je vérifie ma compréhension du texte

1. Qu'est-ce que l'Antarctique et le Groenland ?
2. Que représente leur surface par rapport aux terres de notre planète ?
3. Quelle est la cause de la fonte des glaciers ?
4. Quelle en sera la répercussion sur le niveau de la mer ?
5. Que risquent les côtes de la Terre entière ?

Je retiens

Le réchauffement climatique est la 1^{ère} cause de la montée des eaux. Cela représente un sérieux danger pour une partie importante de la population mondiale.



Le sais-tu ?

50% de la population mondiale vit dans les zones côtières.



Texte support de la 1ère séquence (Projet III)

PROJET

3

Séquence 1

Je comprends mon texte

Pourquoi devrions-nous utiliser des énergies renouvelables ?

texte

Pour ne plus dépendre des carburants fossiles tels que le pétrole, le gaz et le charbon qui, en plus d'être très polluants, vont s'épuiser rapidement, il faut :

- investir dans l'énergie solaire, technique qui consiste à utiliser l'énergie du rayonnement solaire ;
- installer des fermes éoliennes qu'on appelle aussi parc éolien ou centrale éolienne. Cela consiste à regrouper sur un site un grand nombre d'éoliennes. Le site doit impérativement se situer dans un lieu où les vents soufflent fortement et régulièrement ;
- développer la biomasse en utilisant les matières animales ou végétales brûlées pour produire de la chaleur.

Pour satisfaire la demande croissante en énergie de la population mondiale, le développement des énergies renouvelables devient incontournable.

Rappelons que ces énergies sont inépuisables et moins dangereuses pour l'homme et l'environnement.

*D'après « Les progrès de la science »,
Éditions LLC (2007)*

Je vérifie ma compréhension du texte

1. Quels sont les carburants fossiles cités dans le texte ?
2. Quelles sont les énergies dont parle l'auteur dans le texte ?
3. D'après l'auteur, que doit-on faire pour ne plus dépendre des carburants fossiles ?
4. Pourquoi le développement des énergies renouvelables est-il devenu incontournable ?

Je retiens

Les énergies renouvelables sont la solution du futur pour deux raisons :

- elles sont inépuisables ;
- elles sont moins polluantes que les énergies fossiles.



Le sais-tu ?

Un texte **prescriptif** sert à prescrire, conseiller, donner des indications ou des ordres.
Pour cela, on utilise souvent une suite d'actions à réaliser dans un certain ordre.

Texte support de la 2^{ème} séquence (Projet III)

PROJET

3

Séquence

Je comprends mon texte

Les bons réflexes pour devenir un éco-citoyen

texte

Pour réaliser des économies et éviter le gaspillage, voici une liste de réflexes à avoir dans la vie de tous les jours :

- couper les veilles de la télévision, de l'ordinateur et de tous les appareils électriques après utilisation.
- débrancher le chargeur de batterie une fois le téléphone portable rechargé.
- éteindre les lumières d'une chambre, d'une classe, dans les couloirs et les toilettes quand il n'y a personne.
- ne pas laisser couler l'eau inutilement, pendant le lavage de la vaisselle, le rasage ou encore le brossage des dents.
- prendre une douche rapide plutôt qu'un bain.
- boire de l'eau du robinet et non pas de l'eau minérale.
- éviter d'acheter son goûter, le préparer à la maison.
- pousser les proches et amis à agir pour un meilleur environnement.

*D'après « Des gestes pour la nature »
bienchezsoi.net*

Je vérifie ma compréhension du texte

1. A qui s'adresse l'auteur de cette liste de consignes ?
2. A quoi sert-elle ?
3. Que faut-il faire pour devenir un éco-citoyen ?

Je retiens

Pour éviter le gaspillage et réaliser des économies, il est facile de s'habituer à avoir des gestes simples. Ces gestes deviendront à la longue des réflexes qui feront de toi un éco-citoyen.



Le sais-tu ?

La quantité d'eau que tu gaspilles en prenant un bain est l'équivalent de 4 à 5 douches.

