

Else Maria Rolsted Lie

La taille du vocabulaire français chez des élèves au niveau lycéen en Norvège

Masteroppgave i lektorutdanning i språkfag

Veileder: Kjersti Faldet Listhaug

November 2022

Else Maria Rolsted Lie

La taille du vocabulaire français chez des élèves au niveau lycéen en Norvège

Masteroppgave i lektorutdanning i språkfag
Veileder: Kjersti Faldet Listhaug
November 2022

Norges teknisk-naturvitenskapelige universitet
Det humanistiske fakultet
Institutt for språk og litteratur



NTNU

Kunnskap for en bedre verden

Remerciements

Tout d'abord, je voudrais remercier ma directrice Kjersti Faldet Listhaug pour ses précieux conseils et son accompagnement inestimable durant mon projet. Vos encouragements étaient très pertinents. Je suis également très reconnaissant pour les conversations plus personnelles qui ont eu lieu pendant le travail, et j'apprécie la cordialité que vous m'avez témoignée.

Puis, j'aimerais faire des remerciements à Maren Kjellmark qui m'a aidée généreusement et m'a donné des bons conseils pour mon travail.

Je veux aussi remercier les personnes et les enseignants qui ont accepté de participer à l'étude, et Stéphanie Robin qui a fait la relecture de ce mémoire.

J'aimerais remercier mon petit ami Tor-Håkon qui m'a également donné des bons encouragements. Merci pour ta patience.

Pour finir, j'adresse mes remerciements à mes parents Johan Inge et Birgitte, et à mon frère Kristian. Vous avez toujours donné des encouragements généreux pendant mes cinq ans d'études. Et bien sûr, un remerciement supplémentaire et bien mérité à ma maman qui a fait des relectures de tous mes travaux écrits pendant les études, ce travail inclus.

Résumé

Dans ce mémoire, nous examinons la taille du vocabulaire en français chez les élèves du secondaire en Norvège. Nous utilisons le test X-lex (Meara et Milton, 2003), un test de oui/non avec 100 mots français et 20 mots inventés. Les mots inventés sont des mots qui ressemblent à des mots français orthographiquement ou phonétiquement, et les 100 vrais mots sont sélectionnés parmi une liste des 5000 mots les plus fréquents dans la langue française. Les 80 participants, âgés 15 à 18 ans, et issus des 4 niveaux de classe du lycée norvégien, ont rempli trois versions du test. Les niveaux varient de débutant (moins de 6 mois de français) à intermédiaire (4,5 ans de français). 30 participants sont omis du résultat. Les scores moyens à chaque niveau ont été calculés, et comparés les uns aux autres. Les scores moyens obtenus sont 745 (apprenants de 1ère année), 853 (apprenants de 2ème année), 773 (apprenants de 3ème année) et 1325 (apprenants de 4ème année). Une analyse ANOVA a également été menée pour examiner les différences statistiques entre les groupes. Il y a une grande variance parmi les scores des participants des trois premiers groupes, et des petites tailles d'échantillons, et conséquemment, il n'y a pas une différence statistique entre ces groupes. Le dernier groupe a obtenu un score significativement plus élevé, et il y avait des différences significatives entre ce groupe et les autres. Nous comparerons nos résultats avec des études antérieures similaires utilisant le même test X-Lex, une étude suédoise (Lindqvist, 2017), deux études britanniques (Milton, 2008; David, 2008) et une étude norvégienne (Kjellmark, 2021). Nos scores sont généralement plus élevés que les scores des autres études aux niveaux débutants, mais nous n'observons pas la même croissance régulière.

Table de matières

Remerciements	I
Résumé	II
Chapitre 1. Introduction	1
1.1 Les objectifs de l'étude	2
Chapitre 2. Définitions	3
2.1 Le terme mot	3
2.1.1 <i>Types and tokens</i>	4
2.1.2 <i>La lemmatisation et les familles de mots</i>	5
2.2 La notion de la connaissance d'un mot	7
2.2.1 <i>Le lexique mental et les propriétés de la connaissance du vocabulaire</i>	8
2.2.2 <i>Ce qui est impliqué dans la connaissance d'un mot</i>	10
2.3 L'acquisition d'une langue seconde et les apprenants d'une langue seconde	11
2.4 Mots cognats	12
Chapitre 3. La taille du vocabulaire réceptif: Pourquoi et comment mesurer?	13
3.1 Pourquoi faut-il mesurer la taille du vocabulaire ? L'importance du lexique dans l'apprentissage d'une langue	13
3.2 Combien de mots faut-il connaître ?	15
3.2.1 <i>La couverture</i>	15
3.2.2 <i>La connaissance du vocabulaire définie par le CECRL</i>	16
3.3 Comment mesurer la taille du vocabulaire	19
3.3.1 <i>La fréquence des mots</i>	19
3.3.2 <i>Les tests de taille de vocabulaire réceptif</i>	21
3.3.2.1 Le Test de la Taille de Vocabulaire	21
3.3.2.2 Le test X-Lex	22
3.3.3 <i>La qualité et validité des testes</i>	23
Chapitre 4. Études précédentes de la taille du vocabulaire aux apprenants de français	25
4.1 Les études utilisant le x-lex	25
4.1.1 Études anglophones	25
4.1.2. Études scandinaves	26
4.2 Les cognats et leur effet sur des tests de vocabulaire	28
Chapitre 5. Méthode	29
5.1 Participants	29
5.2 La procédure	32
5.2.1 Le questionnaire biographique	33

5.2.2 Le test X-Lex	34
5.2.2.1 La réalisation du test (et les modifications)	34
5.2.3. Calcul de scores	35
Chapitre 6. Résultats	37
6.1. La taille de vocabulaire	37
6.2. Le développement du vocabulaire	38
Chapitre 7. Discussion	41
7.1 Comparaison des résultats	41
7.2 Des conséquences pédagogiques	45
7.3 Facteurs du test X-lex qui peuvent affecter nos résultats	46
7.3.1 <i>L'effet des mots cognats</i>	46
7.3.2 <i>Le nombre des participants qui répond "oui" aux mots inventés</i>	48
7.3.3 <i>La différence entre les trois version du test</i>	49
7.4 Les limites de cette étude	50
Chapitre 8. Conclusion	52
Bibliographie	54
Annexes	58
I Relevans for lektorutdanningen	58
II Bakgrunnsinformasjon for forskningsprosjekt om vokabularstørrelse	59
III Informasjonsskriv	62
IV La mis en page d'une des versions du test X-lex à Nettskjema	65

Chapitre 1. Introduction

Souvent, la grammaire est considérée comme l'élément le plus important dans l'apprentissage d'une langue. Certes, la grammaire nous fournit les règles pour pouvoir créer des phrases, mais pour qu'un système grammatical soit utile, il faut des mots pour remplir ses fonctions. Comme dit Wilkins (1972): "[...] without grammar, very little can be conveyed, without vocabulary *nothing* can be conveyed." (p.111). Ces dernières années, de plus en plus d'études ont examiné le rôle du lexique et du vocabulaire, et ont découvert sa fonction fondamentale dans l'apprentissage d'une langue (Gass, 2013; Laufer, 2021; Milton, 2009). La recherche du vocabulaire est très variée, et inclut des études sur la relation lexique-grammaire (Levelt, 1989; Qian et Lin, 2020), l'effet de la taille du vocabulaire sur la compréhension (Wang et Treffers-Daller, 2017), la production (Miralpeix et Muñoz, 2018; Stæhr, 2008), et comment les erreurs de vocabulaire sont perçues comme les plus graves par des apprenants (Gass, 2013) - pour n'en citer que quelques-uns. Plusieurs de ces études ont montré que le vocabulaire et la trajectoire du développement lexical sont des indicateurs forts sur la maîtrise de la langue en général, et soulignent l'importance de cette branche linguistique. De plus, la nature du vocabulaire évoque un grand avantage dans la recherche du thème, puisqu'il peut être quantifié assez précisément - contrairement aux autres domaines linguistiques.

Ces dernières années, plusieurs tests visant à mesurer la taille du vocabulaire ont été développés. La plupart de ces tests sont créés pour mesurer le vocabulaire des apprenants de l'anglais, et comme conséquence, nous avons un bon nombre de données de ce groupe d'apprenants. Cependant, il nous manque des recherches sur d'autres langues cibles. De plus, il faut aussi mener des recherches sur des langues maternelles variées. En ce moment, de nombreuses études sur la taille du vocabulaire de français comme langue cible examinent les apprenants de l'anglais comme langue maternelle (voir par ex. Milton 2008; David 2008). Pourtant, il est nécessaire d'examiner la taille du vocabulaire en français dans les autres langues et pays aussi. En 2017, Lindqvist a mené une étude visant à mesurer la taille du vocabulaire français des apprenants suédois au niveau du collège, et dans l'étude, Lindqvist encourage la

réalisation d'études similaires dans d'autres pays scandinaves. En utilisant lesdits encouragements comme source d'inspiration, Kjellmark (2021) a très récemment examiné la taille du vocabulaire des étudiants de français à l'université norvégienne, et elle est, à notre connaissance, la seule étude de ce type menée en Norvège. Cette thèse s'inspire donc de ces deux études, et vise à mesurer la taille du vocabulaire des apprenants norvégiens au niveau du lycée.

1.1 Les objectifs de l'étude

L'objectif de cette étude est d'estimer la taille du vocabulaire réceptif des apprenants norvégiens en français au niveau du lycée. Pour mesurer cela, nous utilisons le test X-Lex, développé par Meara et Milton (2003). Les études de Kjellmark (2021), Lindqvist (2017), David (2008) et Milton (2009) mentionnées ci-dessus utilisent aussi le même test. Le test se compose de 120 mots, 100 mots français et 20 mots qui ne sont pas des mots qui existent dans la langue française, mais qui ressemblent à des mots français orthographiquement. Le X-Lex est un test de oui/non, et le participant doit cocher "oui" sur des mots qu'il connaît et "non" sur des mots qu'il ne connaît pas. Comme déjà indiqué, nous nous basons principalement sur l'étude de Lindqvist, et cela inclut aussi les objectifs principaux de cette étude. Les questions de recherche de ce travail sont donc les suivantes :

- I. Quelle est la taille du vocabulaire français chez des lycéens norvégiens?
- II. Y a-t-il un développement de la taille du vocabulaire pendant les années d'études ?

Nous estimons que la taille du vocabulaire réceptif de nos participants est similaire aux résultats dans l'étude de Lindqvist (2017), et qu'elle est inférieure aux scores des participants plus âgés dans l'étude de Kjellmark (2021). Nous estimons aussi que la taille augmente de façon constante pendant chaque année, comme dans l'étude de Lindqvist (2017), David (2008) et Milton (2008).

Chapitre 2. Définitions

L'objectif de cette étude est d'estimer la taille du vocabulaire, ou bien, le nombre des mots connus par les apprenants de français L2. Pour faire cela, il faut premièrement regarder au profond ce qui constitue les notions de *mot*, de *vocabulaire*, du concept *connaître un mot*, ainsi que les termes de *l'acquisition -et l'apprenant d'une langue*. Ces termes peuvent sembler très simples, cependant, dans le recherche du vocabulaire, ils peuvent être définis d'une grand nombre de manières. Les définitions utilisées peuvent fortement affecter nos résultats, puisqu'ils affectent comment compter des unités ; "For example, both Nation (2001) and Schmitt (2000) report estimates of the number of word families in Webster's Third New International Dictionary. One reports 54,000 and the other 114,000 respectively. It is important to be clear at all times what the unit is." (Milton, 2009, p.13). Il est alors très important d'explorer chaque notion et de présenter les définitions que nous choisissons d'utiliser.

2.1 Le terme *mot*

Le Petit Robert donne la définition suivante d'un mot: 1. Élément du langage: Chacun des sons ou groupe de sons correspondant à un sens, entre lesquels se distribue le langage. Les mots écrits sont séparés par des blancs." (*Mot*, 2021). Cependant, si nous explorons cette définition, nous rencontrons plusieurs problèmes. Par exemple, regardez les mots composés de plusieurs unités, comme *salle de bain*. Chaque unité peut se tenir indépendamment, et a son propre sens. De plus, à l'écrit, chaque unité est aussi séparée par des blancs. Donc, est-ce qu'on traite *salle de bain* comme trois mots, ou est-ce que cette combinaison d'unités doit être considérée comme un seul mot étant donné qu'ensemble, ils créent quelque chose de distinct avec son propre sens? Alors, cette définition ne suffit pas, et il faut explorer les autres manières de regarder et compter les mots. Dans la recherche et la littérature de vocabulaire, le mot "mot" est accompagné de plusieurs définitions différentes, chacune adaptée à un objectif spécifique. Dans les deux parties suivantes, nous allons donc examiner les quatre manières les plus courantes de compter les mots, discuter ce qui implique chaque définition, et ensuite, justifier la définition utilisée dans cette étude. Nous commençons avec les notions de types et tokens.

2.1.1 Types and tokens

Pour compter les mots, une des manières la plus simple est l'utilisation des *types* et *tokens* (Nation, 2013; Milton, 2009). Le terme *token*¹ réfère à tous les mots tels qu'ils apparaissent dans un texte, mais la notion de *type* indique le nombre de mots uniques. Regardez la phrase suivante:

La fille mange la pomme rouge.

Dans cette phrase, nous avons six tokens (le mot "la" apparaît deux fois), et notre phrase contient cinq types (fille, mange, pomme, rouge, la). Il est alors évident que les types sont beaucoup plus intéressants en tant que mesure de la connaissance du vocabulaire des apprenants, étant donné que nous voulons savoir combien de mots différents ils ont à leur disposition, plutôt que le nombre total de mots qu'ils peuvent produire (en répétant). Même si la convention de tokens n'est pas très utile pour ce travail, ni pour la notion de la taille de vocabulaire en général, elle peut être appliquée aux cas de nombre des mots d'un corpus, et aussi pour mesurer la richesse lexicale - en mesurant si le lecteur varie ou répète des mots. Il faut quand même noter que compter les mots comme types n'est pas trop utile pour mesurer la taille du vocabulaire, comme le nombre des types ne reflète pas nécessairement le nombre des mots uniques connus par l'apprenant. Prenez la phrase suivante par exemple:

La fille mange la pomme rouge donnée par les filles qui font pousser les pommes rouges.

Comptant les mots comme types, la phrase contient 14 types. Les seuls mots qui se répètent sont les déterminants "la" et "les". Les mots *fille*, *pomme* et *rouge* apparaissent une fois en singulier et une fois en pluriel, mais ils sont quand même comptés comme des types uniques. Pourtant, si nous comptons le nombre des mots connus par une personne, en incluant toutes les formes grammaticales - telle que le pluriel - le résultat sera gonflé, et ne reflète pas nécessairement leur propre connaissance des mots. Nous abordons la relation entre la notion de mot et les formes des mots dans ce qui suit.

¹ Aussi nommé "running words" (Milton, 2009).

2.1.2 La lemmatisation et les familles de mots

Pour éviter d'inclure toutes les formes grammaticales en comptant les mots, il y a des conventions qui définissent un mot de manières différentes des méthodes mentionnées ci-dessus. Les conventions les plus courantes pour notre travail sont soit un regroupement par *lemmes* -la *lemmatisation* - soit un regroupement par *familles de mots*. Il faut cependant noter que les deux notions sont définies de manières très variées parmi les linguistes, donc dans ce qui suit, nous examinons ces deux termes et délimitons leurs définitions telles qu'elles sont utilisées dans ce travail. Comme cette étude utilise le x-lex créé par Mara et Milton (2003), nous nous basons sur ses définitions.

Tout d'abord, il existe deux règles générales qui sont utilisées dans les deux méthodes. Premièrement, les mots composés - comme le mot *salle de bain* comme mentionné dans l'introduction sont comptés comme un seul mot. Deuxièmement, des mots avec plusieurs définitions comme le mot *voler* (se mouvoir dans l'air, ou prendre quelque chose de quelqu'un qui n'a pas donné son accord), sont comptés comme deux mots séparés.

La convention de la lemmatisation (mesurer les mots en lemmes) est souvent appliquée aux tests qui mesurent la taille du vocabulaire. Un lemme se compose d'un "mot tête"², aussi nommé un mot principal, et peut inclure les formes morphologiques grammaticales/flexionnelles les plus fréquentes. Selon Riegel et al. (2018):

On distingue, d'une part, la morphologie flexionnelle ou grammaticale qui traite des variations de la forme des mots selon les catégories du nombre, du genre, de la personne, etc. ; d'autre part, la morphologie lexicale qui décrit les mécanismes, notamment de dérivation et de composition, qui président à la formation des mots. (p. 887)

Ainsi, un verbe comme *travailler* forme avec plusieurs formes flexionnelles un seul lemme; *travaillais*, *travaillent*, *travaille*, etc. Cependant, le substantif *travail* n'apparaît pas dans ce lemme, comme il fait partie d'une classe de mot différente. Conséquemment, le lemme de *travail* inclut aussi d'autres formes flexionnelles, mais celles du substantif. La lemmatisation est très utile pour la recherche sur la taille du vocabulaire parmi les apprenants débutants puisqu'on

² *Headword* en anglais, (Milton, 2009, p.9)

introduit normalement les inflexions les plus fréquentes tôt dans l'apprentissage (Milton, 2009), et le lemme est donc considéré l'unité la plus fiable pour compter les mots (Vemeeer, 2004). De plus, Kremmel (2016) propose que “[...]because the members of a lemma differ only in grammatical form rather than lexicosemantic properties (at least in some cases), knowledge of the lemma representative on the test would most likely imply knowledge of the other lemma members.” (p. 978). Même si l'apprenant ne maîtrise pas productivement tout le paradigme flexionnel, il peut se servir de sa connaissance des formes fréquentes pour reconnaître un lemme ou accéder à son vocabulaire réceptif (expliqué dans la partie 2.2.2). L'exemple du verbe *travailler* au-dessus fait partie des verbes du premier groupe et est un bon exemple d'une morphologie grammaticale très régulière, et qui conséquemment est souvent inclus dans l'enseignement des débutants.

Une autre convention utilisée pour mesurer des mots applique l'idée de diviser les mots en familles de mots (*word family* en anglais). Cette convention utilise un plus grand éventail des inflexions, mais aussi des dérivations des mots. Les familles de mots incluent plus des classes des mots qu'un lemme, et conséquemment aussi plus des inflexions différentes. Le tableau 1 montre les différences principales entre le lemme et la famille des mots, avec des exemples.

Tableau 1

Les exemples des mots communs et comment ils sont définis comme lemme ou famille des mots³

<i>Mot de base</i>	<i>Formes qui peuvent être inclus dans un lemme</i>	<i>Formes qui peuvent être inclus dans une mot de famille</i>
(un) vol	(des) vols	voler, survoler
gouverner	gouverne, gouvernait, gouvernant, a gouverné	gouvernement, gouverneur, ingouvernable
grand	grande, grandes, plus grand	grandir, grandesse

Puisqu'une famille de mots inclut plus de mots sous la même famille, et un lemme inclut seulement des mots d'une seule classe de mots et les formes morphologiques flexionnelles les

³ Tableau 1 est basé du tableau de Milton, 2009, p.11

plus fréquentes, compter les mots comme des lemmes au lieu de familles de mots donne un plus grand nombre de mots. Une division en familles de mots est particulièrement utile pour estimer la taille du vocabulaire chez des locuteurs natifs. On peut s'attendre à ce qu'ils connaissent presque toutes les manières de dériver et conjuguer les mots, contrairement aux apprenants d'une nouvelle langue. Cependant, il faut noter que c'est difficile de comparer les résultats des tests qui n'utilisent pas la même manière de compter les mots. Il existe une règle empirique qui est de multiplier le score de la taille de vocabulaire dans des mots de familles par 1.6 pour rendre le score équivalent en lemmes, mais il faut utiliser cette règle avec prudence (Milton, 2009). Parce que nous allons examiner la taille du vocabulaire des apprenants d'une nouvelle langue, nous comptons les mots comme des lemmes et pas comme des mots de famille. Nous utilisons la définition suivante d'un lemme ; une lemme comprend un mot de tête et des formes morphologiques grammaticales les plus fréquentes dans une seule classe de mot (Milton, 2009). Désormais, nous continuons à utiliser le terme *mot*, avec la définition des lemmes dans la suite (sauf au cas où une terminologie différente est requise et soulevée).

2.2 La notion de la connaissance d'un mot

Ayant discuté de ce en quoi consiste un mot, et aussi les implications des définitions du terme, nous continuons avec la notion de la *connaissance* d'un mot. Connaître un mot, qu'est-ce que cela implique? Cette question est très complexe, et les études qui examinent la connaissance des mots impliquent plusieurs domaines scientifiques, tels que la neurolinguistique et la psycholinguistique. En plus, nous allons traiter la connaissance de mots dans une deuxième langue, et la question devient encore plus complexe. Est-il nécessaire d'utiliser le mot dans une phrase pour le connaître, ou suffit-il de reconnaître le mot, ou comprendre le mot seulement quand il est utilisé dans un contexte ? Faut-il savoir toutes les définitions différentes du mot, ou suffit-il de connaître une seule définition ? Dans cette partie, nous présentons quelques manières courantes de connaître un mot et discutons de ses avantages et de ses faiblesses.

2.2.1 *Le lexique mental et les propriétés de la connaissance du vocabulaire*

Une notion centrale pour la compétence lexicale et celle de *lexique mental*. Dans la plupart des vues théoriques concernant la taille du vocabulaire, le lexique est un élément clé pour comprendre comment fonctionne notre langue (Juffs, 2009). Cette théorie propose que notre lexique se compose des archives de connaissances sur les mots de notre langue. Ces référentiels sont structurés d'une manière qui nous permet de récupérer les mots rapidement, et qui organise toute l'information sur des entrées lexicales pour communiquer d'une manière efficace (Levelt, 1989). On n'a pas encore tout découvert sur cette organisation, mais il y a quand même quelques théories derrière le processus. Dans ce travail, nous n'allons pas trop dans les détails, parce que la théorie est très complexe, en particulier concernant le bilinguisme. Voir par exemple le travail de Listhaug (2015) pour des explications plus extensives.

Pour connaître un mot, le mot doit être inséré dans le lexique mental. Du point de vue de la production de langue, chaque entrée du lexique a plusieurs caractéristiques, par exemple la signification du mot, des propriétés syntaxiques, des spécifications de formes (des segments phonologiques et syllabiques) et des spécifications morphologiques (Levelt, 1989). Cependant, un mot ne doit pas être accompagné par toutes ces caractéristiques pour être connu. D'un niveau très simple, un mot est connu lorsque il est inséré dans le lexique, et se développe quand le mot est reconnu et utilisé. Autrement dit, il y a différents niveaux de la connaissance d'un mot, et dans la prochaine partie nous examinons deux théories visant à expliquer les différents types de connaissance d'un mot.

Tout d'abord, Anderson et Freebody (1981) introduisent la théorie qui fait une division entre *l'étendue* du vocabulaire et *la profondeur* du vocabulaire.⁴ L'étendue fait référence au nombre de mots qu'un apprenant connaît, tandis que la profondeur fait référence à ce que l'apprenant connaît de ces mots. Si nous lions ces termes au lexique mental, l'étendue du mot est quand le mot est inséré dans le lexique, alors que la profondeur d'un mot est quand on connaît plus des caractéristiques du mot, et les relations lexicales aux autres mots. La division nous permet d'établir une distinction entre des apprenants qui ont peut-être appris beaucoup de mots sans savoir comment les utiliser proprement, et les apprenants qui ont également appris

⁴ *Breadth* et *depth* en anglais (Anderson et Freebody, 1981).

comment les mots s'associent aux autres mots, ou bien les nuances de sens qu'ils portent. Bien que ce soit une commodité pour caractériser les différentes qualités des apprenants, les termes ne définissent pas à quel point un mot fait partie d'un vocabulaire étendu ou profond. Par exemple, l'étendue du vocabulaire pourrait d'un côté impliquer la reconnaissance passive des mots tout à fait séparée du sens, donc la capacité de reconnaître un mot sans pouvoir donner la traduction. De l'autre côté, ce terme pourrait aussi impliquer que l'apprenant peut en fait définir le mot. En plus, le terme de la profondeur de vocabulaire est encore plus difficile à cerner, car il peut d'une part impliquer la connaissance des mots associés et des différentes significations d'un mot, ou d'autre part il peut faire référence à la connaissance des caractéristiques morphologiques ou orthographiques d'un mot (Milton, 2009).

Ensuite, une des conventions la plus courantes et simples est de faire la distinction entre *la connaissance réceptive* et *la connaissance productive*⁵ du vocabulaire. La connaissance réceptive concerne les mots qui sont reconnus lorsqu'ils sont entendus ou lus, tandis que la connaissance productive concerne les mots qui peuvent être rappelés à l'esprit et utilisés dans la production orale ou écrite. En général, on estime que la connaissance réceptive est plus grande que la connaissance productive (Milton, 2009 ; Gass, 2013). Nous revenons à ces notions plus loin, en définissant chacune plus en détail.

Les idées courantes telles que la profondeur et l'étendue du vocabulaire ne suffisent pas pour précisément définir la complexité de la connaissance du vocabulaire. Les distinctions de la connaissance réceptive et passive fonctionnent, mais sont aussi trop simples pour donner une image complète de la complexité de la connaissance du vocabulaire. Teichroew (1982) l'argument que "Lexical knowledge cannot be captured by means of a simple dichotomy [...] vocabulary knowledge can best be represented as a continuum, with the initial stage being recognition, and the final being production (Gass, 2013, p. 197)". D'un côté, cette citation convient qu'il existe au moins deux processus différents derrière la connaissance d'un mot, mais elle souligne également que le processus est probablement beaucoup plus compliqué que cela et qu'il y a d'autres divisions qui doivent être faites. La convention élaborée par Nation (2001), présentée ci-dessous, essaie de donner un sommaire plus complet et systématique des différents types de connaissance de vocabulaire.

⁵ Aussi nommé *la connaissance passive et active* (voir Milton 2009 par ex.)

2.2.2. Ce qui est impliqué dans la connaissance d'un mot

Dans les dernières années, les distinctions et définitions créées par Nation (2001) sur la connaissance du vocabulaire sont très courantes, et sont souvent citées dans des études concernant le vocabulaire. Tout d'abord, Nation (2001) suggère que la connaissance d'un mot se compose de trois éléments: *la forme, le sens et l'utilisation*. Ces catégories sont ensuite divisées encore en neuf types de connaissances. Le tableau 2 montre ces neuf types de connaissance, avec des questions qui expliquent chaque échelle de connaissance. Chacune des subdivisions est soit d'un type *réceptif* (indiquée par un R) ou *productif* (indiquée par un P) et retient la convention qu'il existe une distinction mesurable entre ces deux types de connaissances. Daller et al. (2007) suggèrent également une troisième catégorie: *la fluidité*. Cette catégorie est définie comme la capacité de reconnaître et utiliser du vocabulaire avec facilité et rapidité. L'idée est qu'il y a des apprenants qui peuvent accéder à leurs langues facilement et couramment, tandis que d'autres apprenants ont des problèmes pour accéder à leur vocabulaire d'une même manière.

Tableau 2

Tableau traduit de ce qui est impliqué dans la connaissance d'un mot

<i>Ce qui est impliqué dans la connaissance d'un mot</i>				
R = connaissance réceptive ; P = connaissance productive				
Forme	Orale	R	Quels sont les sons du mot ?	
		P	Comment le mot est-il prononcé ?	
	Écrite	R	À quoi le mot ressemble-t-il ?	
		P	Comment le mot est-il écrit et orthographié ?	
		R	Quelles sont les parties reconnaissables dans ce mot ?	
Parties des mots	P	Quelles sont les parties de mots nécessaires pour exprimer le sens ?		
	R	Quel sens la forme de ce mot signale-t-elle ?		
Sens	Forme et sens	P	Quelle forme de mot peut être utilisée pour exprimer ce sens ?	
		R	Qu'est-il inclus dans ce concept ?	
	Concept et référents	P	À quels items ce concept peut-il se référer ?	
		R	À quels autres mots cela nous fait-il penser ?	
		Associations	P	Quels autres mots pourrions-nous utiliser à la place de celui-ci ?
			R	Dans quelles structures le mot apparaît-il ?
Utilisation	Fonctions grammaticales	P	Dans quelles structures devons-nous utiliser ce mot ?	
		R	Quels mots ou types de mots apparaissent avec celui-ci ?	
	Collocations	P	Quels mots ou types de mots devons-nous utiliser avec celui-ci ?	
		R	Où, quand et à quelle fréquence pourrions-nous nous attendre à rencontrer ce mot ?	
	Contraintes d'utilisation	P	Où, quand et à quelle fréquence pouvons-nous utiliser ce mot ?	

(Tableau traduit⁶ tiré de Vinet, 2011)

⁶ Version en anglais rappelée par Nation (2013, p.27)

Évidemment, quand la notion de la connaissance du vocabulaire est si complexe, il s'avère difficile de trouver une méthode qui puisse mesurer un seul élément de la connaissance du vocabulaire, et en même temps, un seul test ne pourrait jamais mesurer tous les éléments de la connaissance de vocabulaire. Parce que les connaissances linguistiques doivent être testées indirectement, des éléments variés de la connaissance du vocabulaire sont souvent évalués en même temps, et d'autres types de connaissances et de compétences peuvent être utilisés par l'apprenant, qui peuvent affecter les résultats des tests et le vocabulaire produit. Néanmoins, dans cette étude, nous allons mesurer la taille du vocabulaire du type *réceptif*, en utilisant la distinction faite par Nation (2001) au-dessus. Nous revenons aux méthodes pour spécifiquement mesurer la taille du vocabulaire réceptif dans le chapitre qui suit, et nous allons aussi examiner comment mesurer ce type de vocabulaire peut donner des informations sur des autres connaissances linguistiques dans l'apprentissage d'une nouvelle langue.

2.3 L'acquisition d'une langue seconde et les apprenants d'une langue seconde

Dans cette étude, nous utilisons le terme *l'acquisition d'une langue seconde*, pour parler du *processus* d'apprendre une langue. Ce terme n'est pas utilisé d'une seule manière, et nous définissons donc notre usage du terme comme utilisé par Ortega (2009, p.1-2) :

Second language acquisition (SLA, for short) is the scholarly field of inquiry that investigates the human capacity to learn languages other than the first, during late childhood, adolescence or adulthood, and once the first language or languages have been acquired.

Désormais, nous utilisons le terme des *apprenants d'une langue seconde* pour parler de (I) les participants de cette étude, et (II) les apprenants d'une langue seconde en général. Quelques clarifications sont nécessaires pour ce terme cependant. Généralement, dans la discipline de SLA, des termes tels que langue maternelle ou première langue (L1, en abrégé) sont utilisés pour

la langue (ou les langues au pluriel) qu'un enfant apprend pendant les années critiques du développement de la langue, tandis qu'une deuxième langue (L2, en abrégé) est utilisée pour toutes les langues apprises après la/les L1. Comme cette étude examine la taille du vocabulaire parmi les apprenants avec norvégien comme leur seule L1, les données des participants avec plusieurs L1 sont omises de l'étude. Nous parlons plus de ce processus dans la partie de la méthode. Parfois, nous utilisons aussi le terme L3 quand c'est nécessaire de diviser entre l'anglais comme L2, et le français comme la langue apprise après la deuxième langue - donc le L3.

2.4 Mots cognats

Dans l'acquisition d'une nouvelle langue, la recherche sur l'influence interlinguistique⁷ est devenue plus importante. Cette domaine linguistique examine l'effet des langues connues dans l'acquisition d'une nouvelle langue et vice versa. Une idée centrale est que des similarités entre les langues connues et la langue cible peut aider l'apprenant dans la compréhension générale, mais aussi dans l'apprentissage des nouveaux mots (Bardel et Falk, 2010; Lindquist, 2020). Dans notre travail, nous allons donc utiliser le terme *cognats*. Cette notion est définie comme la suivante:

[...] orthographically and phonologically identical or similar words that overlap semantically—*cognates* as it were—provide potential affordances that facilitate the acquisition of novel target items, consequently reducing the burden of learnability, enabling recognition and retrieval, and aiding both processing and production in the target language. (Szabo, 2020, p.163. Mot met en cursive par nous).

D'un sens lexical, on suppose donc que les personnes qui connaissent plusieurs langues ont un avantage potentiel en apprenant une nouvelle langue. Cependant, cet avantage dépend des similarités typologiques entre les langues connues et la langue cible (Otwindowska, 2016). Nous revenons sur la notion des cognats et leur effet sur des tests de la taille de vocabulaire dans la partie 4.2.

⁷ *Cross-linguistic influence* en anglais (Szabo, 2020)

Chapitre 3. La taille du vocabulaire réceptif: Pourquoi et comment mesurer?

En recherche, le domaine lexical a généralement suscité moins d'intérêt que les autres domaines de la linguistique. Cependant, dans les dernières années, le nombre d'études sur le thème est croissant. Contrairement aux croyances antérieures, la recherche récente suggère que le lexique peut être un des éléments les plus importants dans l'apprentissage d'une nouvelle langue (Gass, 2013, p. 194.). De plus, un grand nombre d'études ont montré qu'il y a un lien important entre la taille du vocabulaire et la maîtrise de la langue. Dans ce chapitre, nous allons donc premièrement examiner comment la taille de vocabulaire réceptif reflète d'autres compétences linguistiques, et pourquoi la recherche sur le lexique est nécessaire dans le domaine linguistique. Deuxièmement, nous allons discuter combien de mots il faut connaître pour bien maîtriser une nouvelle langue, et combien de mots nous pouvons supposer que les apprenants cibles de cette étude connaissent. Finalement, nous allons présenter quelques méthodes pour mesurer la taille du vocabulaire.

3.1 Pourquoi faut-il mesurer la taille du vocabulaire ? L'importance du lexique dans l'apprentissage d'une langue

Les estimations de l'étendue du vocabulaire ont été démontrées comme des bons indicateurs sur des compétences linguistiques en général (Batista et Horst, 2016 ; Laufer et Nation 1999), et plusieurs études qui sont consacrées à examiner le rôle du vocabulaire ont montré que la compétence lexicale affecte fortement les compétences spécifiques suivantes : (I) la syntaxe et la grammaire, et (II) les quatre compétences du langage d'écriture, de lecture, d'écoute et de parler.

(I) Premièrement, prenez le lien très important entre le vocabulaire et la grammaire et syntaxe. Sur un niveau très simple, c'est évident que ce n'est pas possible de créer des phrases sans mots. Cependant, la relation lexico-grammaticale semble être plus compliquée que prévu. Levelt (1989, p.181) suggère que le lexique est en fait la force primaire de la production des phrases. Il

dit que le processus de la formulation⁸ des phrases et leur syntaxe sont définis par le lexique. Son hypothèse se base sur l'idée que le message verbal déclenche le lexique, qui - avec ses propriétés morphologique, phonologique et syntaxique - déclenche à son tour la grammaire sous-tendue. Le lexique fonctionne donc comme le médiateur entre la conceptualisation et le code grammatical et phonologique. Cette idée est nommée *l'hypothèse lexique*, et elle propose:

The lexical hypothesis entails, in particular, that nothing in the speaker's message will *by itself* trigger a particular syntactic form, such as a passive or a dative construction. There must always be mediating lexical items, triggered by the message, which by their grammatical properties and their order of activation cause the Grammatical Encoder to generate a particular syntactic structure. (p.181)

Il faut noter qu'avec cette hypothèse, Levelt (1989) fait référence à la L1, mais nous pouvons néanmoins constater que - si le lexique est tellement important dans la langue native - c'est un facteur très important aussi dans des L2 (Gass, 2013). De plus, le vocabulaire n'est pas que important que dans la production syntaxique, mais aussi dans la compréhension syntaxique : une étude menée par Altman (1990) examine les facteurs qui affectent la compréhension des phrases, et montre que la compréhension de l'information lexicale est cruciale pour déterminer les relations syntaxiques d'une phrase, et pour bien comprendre des paroles. Enfin, lorsque les deux branches linguistiques de syntaxe et vocabulaire sont comparées, en particulier dans le processus d'acquisition de la L2, la compétence lexicale est souvent considérée comme plus importante que la connaissance syntaxique pour parvenir à la communication (Qian et Lin, 2020).

(II) Deuxièmement, la taille du vocabulaire semble correspondre avec les quatre compétences d'écriture, de lecture, d'écoute et de parler. Par exemple, une étude de Stæhr (2008) examine la relation entre la taille du vocabulaire et les compétences d'écoute, d'écriture et de lecture parmi 88 élèves danois de l'enseignement secondaire, avec anglais comme L2. Il a trouvé que la taille du vocabulaire réceptif des élèves s'est avérée fortement associée avec leurs compétences de lecture et d'écriture, et modérément associée à leur capacité d'écoute. En même

⁸ *Formulation process* en anglais

temps, il souligne que la taille du vocabulaire pourrait encore expliquer une partie substantielle de la variance des scores d'écoute. Une étude très similaire examine cette relation chez 42 apprenants avec une compétence de langue plus élevée que les participants du travail de Stæhr (2008), et les résultats révèlent que la taille du vocabulaire est étroitement liée à l'écriture et est modérément corrélée avec la lecture, la parole et l'écoute (Miralpeix et Muñoz, 2018). De plus, l'étude de Wang et Treffers-Daller (2017) trouve un lien entre la compétence d'écoute et le vocabulaire. Menée auprès de 151 étudiants (non anglophones) d'une université chinoise, cette étude examine quelle proportion de la variance de la compréhension orale s'explique par trois facteurs : la maîtrise générale de la langue, la taille du vocabulaire et la conscience métacognitive. Les résultats montrent que la taille du vocabulaire était le facteur le plus important, suivi de la maîtrise générale de la langue et finalement la conscience métacognitive.

3.2 Combien de mots faut-il connaître ?

Jusqu'ici dans ce chapitre, nous avons vu pourquoi le lexique est très important dans l'apprentissage d'une langue, et que le vocabulaire affecte plusieurs aspects de la maîtrise de la langue. Lorsque l'on aborde ce sujet, il est également nécessaire de discuter de la taille de ce vocabulaire nécessaire pour réussir à communiquer et à comprendre une langue. Afin de pouvoir définir combien de mots il faut connaître à différents niveaux de langue, nous examinerons tout d'abord la notion de couverture (la pourcentage des mots compris dans un texte), puis nous présenterons les niveaux de langue tels que définis par le Cadre Européen Commun de Référence Pour les Langues (CECRL), et enfin nous lierons la couverture aux niveaux de la CECRL.

3.2.1 La couverture

La couverture est définie par le pourcentage des mots connus dans un texte. Un grand nombre des études qui examinent la couverture⁹ et la taille du vocabulaire nécessaire étudient des apprenants d'un niveau relativement haut (à l'université, souvent utilisant des textes

⁹ En anglais, le mot *coverage* est utilisé, mais comme il n'y a pas un terme équivalent en français, nous utilisons le mot couverture comme Kjellamark (2021).

académiques et authentiques), et les participants ont souvent l'anglais pour L1 ou L2. Les résultats de plusieurs de ces études montrent que c'est nécessaire avec une couverture de 95% (Laufer, 1989), et 98% (Hu et Nation, 2000; Schmitt, Jiang et Grabe, 2011). Cela correspond aux découvertes de Laufer et Ravenhorst-Kalovski (2010) qui suggèrent deux niveaux de couverture: un niveau optimal de 98% et un niveau minimum de 95%. Ces chiffres correspondent à une connaissance de 8000 et 4000-5000 mots de famille, respectivement. Une autre étude de Laufer (1992) montre qu'un niveau de 3000 prédit un score de couverture de 56%, un niveau de 4000 est égal à une couverture de 63 % et un niveau de 5000 correspond à une couverture 70%. Encore une fois, ces chiffres peuvent être trop larges pour des apprenants de notre étude. Cependant, il faut être conscient que ces chiffres représentent la compréhension des textes au niveau universitaire, et de façon réaliste nous ne pouvons pas nous attendre à une telle taille parmi les apprenants débutants au lycée. C'est néanmoins pertinent d'examiner des tailles de vocabulaire différentes et la couverture auxquelles elles correspondent, puisque ça nous permet de donner une idée de la taille optimale pour des apprenants. Prenez par exemple l'étude de Stæhr (2008) présentée au-dessus, dans laquelle il a été constaté que la majorité des participants ne connaissaient pas les 2000 mots les plus fréquents en anglais. Les participants qui les connaissaient, cependant, obtenaient également de bons résultats aux tests d'écoute, de lecture et d'écriture, et il suggère donc qu'un niveau de vocabulaire de 2000 est un objectif d'apprentissage crucial pour des apprenants d'une langue seconde (Stæhr, 2008). Cette idée est aussi supportée par Milton (2009), disant que les 2000 mots les plus fréquents sont le plus pertinents pour l'apprenant, et lui permet de communiquer d'une manière "suffisante".

3.2.2 La connaissance du vocabulaire définie par le CECRL

Le Cadre européen commun de référence pour les langues (CECRL) est un document écrit par le Conseil de l'Europe qui divise la maîtrise d'une langue en une échelle de 6 niveaux différents (A1-C2), en définissant la connaissance exigée pour maîtriser ce niveau. Nous n'étudions pas trop dans le détail ce document, mais nous allons brièvement examiner leurs notions sur l'étendue du vocabulaire. Le tableau 3 montre les descriptions et notions-clés du vocabulaire nécessaire à chaque niveau, selon le CECRL.

Tableau 3*CECRL: L'étendue de vocabulaire*

C2	Possède une bonne maîtrise d'un vaste répertoire lexical des expressions idiomatiques et des termes familiers ; est conscient des niveaux de connotation sémantique.
C1	Possède une bonne maîtrise d'un vaste répertoire lexical lui permettant de surmonter facilement les lacunes par des périphrases sans recherche évidente d'expressions et de stratégies d'évitement. Peut choisir entre plusieurs possibilités lexicales dans pratiquement toutes les situations en utilisant des synonymes même pour des mots non familiers. Peut comprendre et utiliser de façon appropriée la gamme de vocabulaire technique et d'expressions idiomatiques propres à son domaine de spécialité.
B2	Peut comprendre et utiliser les termes techniques généraux de son domaine, quand il(elle) en discute avec d'autres spécialistes. Possède une bonne gamme de vocabulaire pour les sujets relatifs à son domaine et les sujets plus généraux. Peut varier sa formulation pour éviter des répétitions fréquentes, mais des lacunes lexicales peuvent encore provoquer des hésitations et ø'usage de périphrases. Peut produire assez systématiquement de nombreux mots ou signes adéquats dans la plupart des contextes. Peut comprendre et utiliser une grande partie du vocabulaire spécialisé de son domaine mais a des difficultés avec la terminologie d'une spécialité différente de la sienne.
B1	Possède un bon vocabulaire lié à des sujets familiers et à des situations de la vie quotidienne. Possède un vocabulaire suffisant pour s'exprimer à l'aide de périphrases sur la plupart des sujets relatifs à sa vie quotidienne tels que la famille, les loisirs et les centres d'intérêt, le travail, les voyages et l'actualité.
A2	Possède un vocabulaire suffisant pour mener des transactions quotidiennes courantes dans des situations et sur des sujets familiers. Possède un vocabulaire suffisant pour satisfaire les besoins communicatifs élémentaires. Possède un vocabulaire suffisant pour satisfaire les besoins primordiaux.
A1	Possède un répertoire élémentaire de mots/signes isolés et d'expressions relatifs à des situations concrètes précises.
Pré-A1	Pas de descripteur disponible.

Note: Tableau tiré de CECRL, 2020.

Le CECRL est pertinent pour notre étude, puisque le plan d'éducation des langues étrangères en Norvège a "un lien clair avec le CECRL, mais les caractéristiques doivent être adaptées à la langue étrangère dans laquelle la formation est dispensée." (Utdanningsdirektoratet, 2020b. Traduction par nous). Le niveau I et le niveau II de français au lycée norvégien (voir la partie 5.1 pour plus de détails) correspond à un niveau A1 et A2 au CECRL, respectivement (Utdanningsdirektoratet, 2020b, 2020c). Pour lier la taille du vocabulaire exigé à chaque échelle de CECRL, nous allons voir le travail de Milton et Alexiou

(2009 dans Milton, 2010). Il a collecté des données auprès de plus de 500 apprenants d'anglais, de français et de grec comme L2, en vue de comparer les niveaux de connaissance du vocabulaire à chaque niveau du CECRL. Les niveaux du CECRL de chaque apprenant ont été déterminés par des enseignants des participants. Les scores moyens et les écarts-types collectés auprès des apprenants à chaque niveau sont résumés dans le tableau 4. L'écart type est inclus pour indiquer le degré de chevauchement entre les groupes à chaque niveau. Les chiffres dans le tableau indiquent que l'échelle de A1 correspond à une taille moyenne de 1400 mots, et en raison des scores bas des apprenants anglophones, le niveau A2 a aussi une moyenne de 1400 mots. Il faut noter que la taille de l'échantillon de ces études est petite et il faut faire des conclusions avec prudence (Milton et Alexiou, 2020).

Tableau 4

Sommaire des scores moyens (et des écart-types en parenthèse) pour chaque niveau CECRL (CEFR en anglais) dans 3 langues différentes

CEFR Level	EFL in Hungary	EFL in Greece	French in Spain	French in Greece	Greek in Greece
A1		1477 (580)	894 (604)	1125 (620)	1492 (705)
A2		2156 (664)	1700 (841)	1756 (398)	2237 (538)
B1	3135 (434)	3263 (434)	2194 (717)	2422 (517)	3338 (701)
B2	3668 (666)	3304 (666)	2450*	2630 (251)	4012 (415)
C1	4340 (471)	3690 (471)	2675 (643)	3212 (473)	
C2		4068 (261)	3721 (416)	3525 (883)	

* in this field there is only one student score and no SD can be calculated

Note: Tableau réimprimé de Milton 2010, The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe, *Eurosla Monographs Series, 1*, p. 225

Puisque différentes langues sont comparées dans le tableau 4, il faut noter qu'on ne peut pas simplement comparer la taille de vocabulaire des langues différents, comme la connaissance de la même taille des mots ne signifie pas nécessairement la même compétence (Milton, 2009). "The figures for French as a foreign language are consistent with each other and suggest that at each CEFR level the learners have smaller vocabulary sizes than would be the case for English" (Milton, 2010, p.226). Quelques langues peuvent fournir une couverture plus élevée que d'autres langues, même avec la même taille de vocabulaire (Cobb et Horst, 2004). Néanmoins, à notre connaissance, il n'existe pas d'autres études qui examinent la relation entre les échelles de CECRL et la taille du vocabulaire, avec les participants de français comme L2. Le moyen de 1400 mots nous donne une *indication* de la taille du vocabulaire qui peut être attendu à ces échelles, mais avec prudence.

3.3 Comment mesurer la taille du vocabulaire

Ayant discuté pourquoi examiner la taille du vocabulaire, et comment cette branche linguistique peut nous fournir des informations importantes sur la maîtrise de la langue en

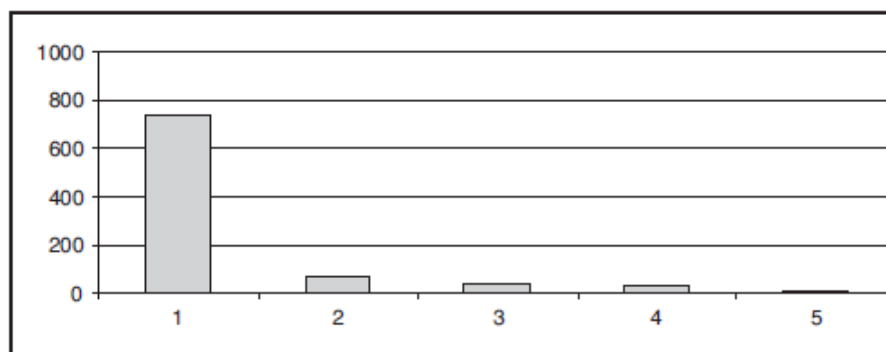
général, nous allons voir comment mesurer cette unité. Pour commencer, nous examinons la théorie de la fréquence des mots, et expliquons comment cette théorie peut être appliquée dans la recherche lexicale.

3.3.1 *La fréquence des mots*

Dans la partie 2.2.1 nous avons expliqué la théorie du lexique mental, et nous avons vu qu'il y a plusieurs niveaux de connaissance de mots. Dans la recherche de la taille du vocabulaire, un facteur central qui va affecter cette connaissance est *la fréquence* du mot. En termes simples, on suppose que quand un mot est utilisé et reconnu fréquemment, c'est plus probable que le mot est inséré dans le lexique mental (Milton, 2009). Pour déterminer la probabilité de rencontrer un mot - c'est-à-dire la fréquence d'un mot - les grandes sélections de corpus d'une langue sont analysées, et les mots sont divisés aux différents niveaux de fréquence, déterminés par combien de fois ils apparaissent dans le corpus. Pour identifier et organiser la fréquence des mots, il est courant de diviser les mots en groupes de fréquence, souvent avec 1000 mots à chaque niveau, et de nommer ces bandes de fréquence 1K, 2K, 3K etc. Dans la première bande, c'est-à-dire la bande désormais nommée 1K, on trouve les 1000 mots les plus fréquents. Autrement dit, on trouve le mot le plus fréquent jusqu'au 999ème mot le plus fréquent. Dans la bande 2K, on trouve donc les 1000 mots les plus fréquents suivants, et ainsi de suite. La figure 1 présente une vue d'ensemble des recherches à la fois en anglais et français, où les apprenants de L2 ont obtenu le score le plus élevé dans la section du test de lexique le plus fréquent, avec un schéma de scores décroissants sur les sections du lexique moins fréquent (Milton, 2009). Généralement, des mots comme des chiffres, des noms propres, et des fautes de frappe dans des corpus ne sont pas inclus dans les listes.

Figure 1

Fréquence moyenne des mots aux 5 premières bandes de fréquence du corpus de Baudot (1992)



Note: Tableau réimprimé de Baudot, 1992, dans Milton, J., 2009, *Measuring second language vocabulary acquisition*, p. 25.

L'hypothèse du lien entre la fréquence et l'acquisition du vocabulaire n'est pas nouvelle (voir Palmer 1917 par exemple), mais elle était démontrée jusqu'à très récemment (voir Paribakht et Wesche, 1996; Horst et Collins, 2006). Un des avantages de cette idée est qu'elle peut être transformée en un modèle qui peut ensuite être testé empiriquement. Deux mises en garde s'imposent ici. L'une est qu'il s'agit d'une tendance et non d'une règle absolument rigide. Premièrement, les manuels d'apprentissage sont souvent organisés en des thèmes et des blocs composés de mots de plusieurs fréquences différentes. Conséquemment, les apprenants d'une langue seconde apprennent souvent des mots qui sont en réalité peu fréquents, comme les mots désignant des vêtements par exemple (Milton, 2009). Deuxièmement, la fréquence d'un mot est un indicateur plus correct sur l'apprentissage d'un mot que les cognates ou la longueur du mot (Milton, 2009). Il est, cependant, aussi clair que dans le cas de certains mots et certains apprenants, l'apprentissage peut ne pas toujours suivre un ordre de fréquence stricte. Par exemple, Lindqvist (2012) montre que les apprenants de L2 français (avec suédois comme L1) peuvent facilement reconnaître certains mots français peu fréquents (*vocation* par ex.) en raison de leurs ressemblances avec des mots suédois (nous y revenons dans la partie 4.2). L'autre mise en garde est que les informations de fréquence ne fournissent pas d'informations sur la difficulté. Ce que l'information sur la fréquence nous dit, est la probabilité qu'un apprenant puisse rencontrer un mot, et le faire répéter assez souvent, pour qu'il soit appris.

3.3.2 Les tests de taille de vocabulaire réceptif

Nous avons maintenant vu pourquoi c'est important de mesurer la taille du vocabulaire dans la recherche linguistique ainsi que la méthode. Il y a donc eu plusieurs tests développés dans le but de mesurer la taille du vocabulaire d'une manière relativement facile. Puisque l'anglais est une des langues les plus utilisées dans le monde, il existe plusieurs tests visant à mesurer la taille de vocabulaire réceptif des apprenants d'anglais comme L2, comme par exemple Le test *Eurocentres Vocabulary Size Test* (Meara et Jones, 1990), le *Vocabulary Levels Test* (Nation, 1983), et le test *X-lex* (Milton, 2008). Des tests pour les apprenants des autres langues n'est pas si nombreux ; en français, il existe principalement deux tests : *Le Test de la Taille du Vocabulaire* (TTV), et la version française du test *X-Lex*. Dans les prochaines parties, nous allons expliquer ces deux tests, leurs avantages et faiblesses et pourquoi nous utilisons X-lex dans cette étude.

3.3.2.1 Le Test de la Taille de Vocabulaire

Batista et Horst (2016) ont développé *Le Test de la Taille du Vocabulaire* (TTV), un test de vocabulaire français qui se base sur un système de grappes. La figure 2 est un exemple tiré du test. Comme le VLT, le TTV utilise des bandes de fréquence de 2K, 3K, 5K, et 10K. La liste de fréquence de Lonsdale et Le Bras (2009) était utilisée pour les mots aux niveaux 2K, 3K et 5K. Cette liste se base sur un corpus de 23 millions de mots, et contient les 5000 lemmes les plus fréquents. Le corpus se compose de textes contemporains écrits, mais aussi de 50% de textes oraux du français international. Les mots dans la bande de fréquence 10K étaient tirés du travail de Baudot (1992). Dans la langue française, il n'existe encore pas une liste des mots français académiques, et ce niveau n'est donc pas inclus dans le TTV.

Batista et Horst (2016) incluent un principe important dans le TTV, notamment que les définitions ne se composent que des mots aux niveaux plus fréquents que les mots qui sont testés. C'est-à-dire que les mots de la bande de fréquence 2K ont des définitions qui appartiennent à la bande de fréquence 1K. Les mots des bandes de fréquence plus hautes (3K, 5K et 10K) sont définis par des mots à la fois de la bande 1K et 2K. Conséquemment, on peut - en théorie - s'assurer que les participants comprennent les définitions. Il faut cependant noter

que, même si un participant comprend un mot moins fréquent, cela ne garantit pas qu'il comprend un mot plus fréquent.

Figure 2

Exemple d'une question du test TTV

Box 4. A noun cluster from 5K frequency section of the TTV	
1. brouillard	
2. coïncidence	
3. farce	_____ une histoire qui fait rire
4. instituteur	_____ ce qui empêche de voir loin
5. pneu	_____ un professionnel de l'éducation
6. soumission	

Note: Figure réimprimé de Batista, R. et Horst, M., 2016, A New Receptive Vocabulary Size Test for French, *Canadian modern language review*, 72, s.p.

3.3.2.2 Le test X-Lex

Le test *X-Lex* a été développé par Meara et Milton (2003). Originellement, le test *X-Lex* était créé pour l'anglais, mais Milton (2009, p. 257-259) a aussi développé trois versions en français. Le *X-Lex* est créé pour faire une estimation sur la connaissance des 5000 mots les plus fréquents dans la langue cible. Chaque version du test se compose de 120 mots de 5 bandes de fréquence différente, et aussi des mots faux. Chaque catégorie - ou bande - consiste en 20 mots chacune, avec des mots échantillonnés des bandes de fréquence 1K, 2K, 3K, 4K et 5K. Finalement, il y a aussi 20 mots inventés, qui ressemblent à des vrais mots. Les mots ressemblent à de vrais mots français, soit par la dérivation des vrais mots (par exemple *arguable** de *argument*), soit en utilisant les mots anglais et les rendant plus français (par exemple *disabilité**, anglais: *disability*), soit en créant des mots composés des syllabes et des affixes qui existent déjà dans la langue (par exemple *précont**)

Les mots de la version française étaient tirés de la liste de Baudot (1992). Le corpus utilisé pour créer des listes de fréquence se compose de 803 échantillons de textes, chaque extrait ayant entre 1000 et 1500 mots. La plupart des textes étaient écrits dans les années 1960-1967 et se composaient de livres et manuels (25%), de journaux (24%), de bulletins et de rapports (7%), de revues et magazines (42%) et de brochures (2%). Quant aux pays d'origine, 62% viennent de la France, 37% viennent du Canada, et 1% viennent d'autres pays francophones.

3.3.3 La qualité et validité des tests

Quand nous avons expliqué le concept du vocabulaire, et du vocabulaire réceptif dans la section 2.2.2, il était clair que ce concept est ambigu pour plusieurs raisons. Conséquemment, il est difficile de concevoir les tests de vocabulaire réceptifs sans faiblesse. Le format le plus utilisé est les tests à choix multiple réceptif et les tests de oui/non - comme les tests mentionnés ci-dessus¹⁰. En effet, ce type de test de vocabulaire n'est naturellement pas sans défauts. Dans cette partie, nous allons mettre en évidence deux principales faiblesses bien connues avec le X-Lex et le TTV: (I) le taux d'échantillonnage et (II) le problème potentiel de surestimation des résultats (Gyllstad, Vilkaitė et Schmitt, 2015). Il faut noter que notre but n'est pas d'examiner la fiabilité ou la validité des tests dans cette étude, et qu'il y a plusieurs études qui sont consacrées à analyser les tests en profondeur (voir par ex. Beeckmans, Eyckmans, Janssens, Dufranne et Velde, 2001; et Batista et Horst, 2016). Nous sommes donc consciente des problèmes et critiques qui sont discutés dans cette partie. Cependant, nous justifions aussi l'utilisation de ce type de test dans ce qui suit.

(I) Un premier défaut des tests est l'échantillonnage des mots de la langue inclus dans le test, et les listes de fréquence. L'étude de Gyllstad, Vilkaitė et Schmitt, 2015 examine le test VLT (*Vocabulary Levels Test*, un test à choix multiple réceptif, très proche au TTV), et souligne l'importance qu'un taux d'échantillon d'une certaine taille est nécessaire pour mieux représenter la population sous-jacente de mots dans les bandes de fréquences correspondantes. En même temps, le test ne peut pas être trop long, afin d'éviter les risques de fatigue de test¹¹ (Nation, 1993), et l'échantillonnage des tests de vocabulaire est absolument nécessaire, puisque ce n'est pas possible de tester tous les mots dans un vocabulaire. Il faut donc trouver l'équilibre entre une taille qui représente les mots d'un vocabulaire, et un échantillon qui n'est pas trop grand. Ensuite, les listes de fréquences utilisées aux tests de vocabulaire peuvent être un défaut potentiel, de plusieurs manières. Rappelons que le TTV se base sur les listes de Lonsdale et Le Bras (2009), quand le X-Lex se base sur l'ouvrage de Baudot. Comme expliqué dans la partie

¹⁰ Le test X-lex est un test auto-rapport à choix binaire, mais l'étude de Gyllstad, Vilkaitė et Schmitt, (2015) inclus ce type de test dans leur liste de test à choix multiple réceptif. Beeckman et al. (2001) utilise le nom "Yes/No" test.

¹¹ Le terme utilisé par Nation (1993) est *test fatigue*

3.3.2.1, la liste de Lonsdale et Le Bras utilise un corpus plus moderne et extensif que Baudot. Prenez par exemple des mots modernes comme *portable* et *internet*, ils sont beaucoup utilisés dans notre vie quotidienne, mais ne sont pas inclus dans le corpus de Baudot, de la même façon qu'il inclut des mots qui sont maintenant dépassés. Néanmoins, leur fréquence ne va probablement pas considérablement changer, puisque les mots les plus fréquents sont souvent des mots de fonction ou de structure (Milton, 2009). De plus, comme mentionné précédemment, le corpus de Baudot est seulement basé sur des sources écrites, contrairement au corpus de Lonsdale et Le Bras qui a 50% de sources orales. Cette caractéristique est importante, comme la langue et le vocabulaire ont une valeur importante à l'écrit et à l'oral, et l'usage des mots peut différer. Comme mentionné par Milton (2009), les informations de fréquence des corpus basés sur des sources écrites peuvent montrer tout à fait autre chose que celles basées sur les sources orales, une des raisons étant que la langue orale est moins formelle, et qu'on trouve certains mots qui sont plus fréquents dans la langue orale et quotidienne que dans la langue écrite.

(II) Le deuxième défaut principal est la surestimation des réponses et des résultats. La même étude de Gyllstad, Vilkaitė et Schmitt (2015) mentionnée au-dessus, souligne que la réponse correcte sur les tests comme TTV et X-lex peut parfois être atteinte par la connaissance du mot cible, mais aussi par l'élimination (au cas de choix multiples) ou par la supposition. En comparant le TTV et le test X-lex, le premier a l'avantage d'avoir moins de risque qu'une réponse correcte soit due à une supposition. Même avec l'inclusion des mots faux, on ne peut pas être sûr qu'ils cochent seulement des mots qu'ils savent. Cependant, pour être capable de compléter le TTV, il est nécessaire avec des bonnes compétences en français ; il ne suffit pas de comprendre les mots cibles, il faut aussi comprendre les mots utilisés dans les définitions pour être capable de donner des réponses correctes. Ce test ne convient pas aux participants très jeunes, ou aux débutants, quand le test X-lex est généralement considéré comme bon pour ces groupes (Meara 1992; Milton et Alexiou, 2020). Il faut aussi noter que le TTV ainsi que le X-lex ne conviennent peut-être pas aux apprenants avec des difficultés d'apprentissage comme la dyslexie, étant des tests à l'écrit où l'orthographe est très importante.

Malgré les avantages d'un corpus plus grand et moderne, et des réponses qui sont plus faciles à valider, nous avons choisi d'utiliser le test X-lex pour cette étude. Nous justifions ce choix pour deux raisons importantes : (I) pour avoir la possibilité de comparer les résultats de

plusieurs d'études antérieures, et (II) parce que le TTV est trop difficile pour les apprenants jeunes et/ou débutants comme dans cette étude. Malheureusement, à notre connaissance, il n'existe pas d'autres alternatives pour tester la taille du vocabulaire réceptif en français.

Chapitre 4. Études précédentes de la taille du vocabulaire aux apprenants de français

Nous passons maintenant à quelques études précédentes qui ont pour but d'estimer la taille du vocabulaire réceptif parmi des apprenants de français L2, utilisant le test X-lex. Nous allons lier les découvertes à nos questions de recherche, en regardant les tailles de vocabulaire découvertes et la manière dont elles augmentent, pour montrer quels résultats peuvent être attendus et pour comparer les résultats dans la partie de discussion. Enfin, nous présentons aussi une étude qui montre comment les mots cognats peuvent affecter les résultats des tests de vocabulaire comme le test X-Lex.

4.1 Les études utilisant le x-lex

4.1.1 *Études anglophones*

En Grande-Bretagne, Milton (2008) réalise un grand test de la taille de vocabulaire réceptif, en utilisant le test X-Lex. Il y avait 449 participants de collèges et lycées britanniques (âgés de 12 à 18 ans). Au lycée, il testait tous les niveaux différents sur une seule école. Le test était fait à la fin de l'année scolaire. En moyenne, les participants de la première année d'apprentissage (year 8, âgé de 12 ans) connaissent 311 mots. Après quatre ans d'apprentissage, le year 10 (14-15 ans) connaît 592 mots. Après un total de 7 ans d'apprentissage (year 13, participants âgés de 17 à 18 ans), Milton (2008) a estimé une taille d'environ 1930 mots. La même méthodologie était utilisée dans une classe de 29 participants (en première année d'études de français) de l'université. Après un an à l'université, leur taille moyenne du vocabulaire a augmenté d'environ 500 mots. Donc, les étudiants d'université apprennent des mots au même rythme que les élèves des deux dernières années de lycée.

Aussi au Royaume-Uni, David (2008) utilise le test X-Lex pour estimer la taille du vocabulaire parmi 483 participants au lycée et à l'université britanniques (âgés de 12 à 23 ans). 66 participants ont été exclus de l'étude, pour avoir coché plus de cinq mots inventés, et les analyses se basent donc sur les données de 417 participants. David (2008) suggère que plus de cinq mots faux peuvent indiquer que le participant devine et leurs résultats ne montrent pas leur connaissance réelle. Le test a été accompli au milieu de l'année scolaire. Nous voyons que les résultats de David (2008) correspondent avec ceux de Milton (2008): les élèves en year 8 (12 ans) connaissent 439 mots en moyenne, quand les élèves en dernière année de lycée (18 ans) ont obtenu un résultat moyen de 2108 mots. Les étudiants en première année à l'université connaissent en moyenne 2524 mots. Le nombre des mots augmente d'une manière régulière au fur et à mesure de leurs études françaises.

4.1.2. Études scandinaves

Dans ce travail, nous avons défini une langue L2 comme toutes les langues apprises après la/les L1. Dans la plupart des cas de nos participants, le français est leur troisième langue apprises, et même si nous utilisons le terme L2 pour la simplicité, nous pouvons donc dire que le français est leur L3. Cependant, la plupart des études, comme celles de Milton (2008) et David (2008), ont des participants qui ne connaissent qu'une seule langue avant d'avoir appris le français, et le français peut donc être défini comme L2, mais pas comme L3. Il y a peu de recherches sur la taille du vocabulaire sur des participants qui ont le français comme troisième langue, comme cette étude. Cependant, l'étude de Lindqvist (2017) examine la taille du vocabulaire parmi les élèves suédois, le français étant leur L3.

Utilisant le test X-lex, Lindqvist (2017) examine la taille du vocabulaire chez 156 apprenants suédophones. Les participants ont 12-15 ans, étant des élèves de la 6ème à la 9ème classe. Pour être capable de comparer des résultats, le modèle proposé par David (2008) était suivi, et Lindqvist (2017) ne prend pas en compte les tests des participants qui ont coché plus de cinq mots inventés. Conséquemment, 152 sont inclus dans l'analyse après la mise à l'écart de quatre tests. La moyenne dans toutes les classes était comparée, et les résultats montrent des similarités aux études ci-dessus, c'est-à-dire que la taille augmente de manière consistante chaque

année. Cependant, parmi les participants les plus âgés dans l'étude de Lindqvist (2017) (15 ans, 9ème classe), ayant quatre ans de français à l'école, certains avaient une taille de vocabulaire moyenne de 1150 mots. Nous pouvons comparer ces résultats à ceux de David (2008) et Milton (2008), où les élèves de quatre et cinq ans d'études avaient une moyenne de 592 et 564 mots, respectivement, et nous voyons une différence considérable. De plus, la taille du vocabulaire des élèves suédoises augmente à un rythme plus rapide.

L'étude de Kjellmark (2021) examine la taille du vocabulaire des étudiantes en première année d'études de français en Norvège. L'étude suit le modèle de Lindqvist (2017) et David (2008), et 16 participants restent après l'élimination de quelques participants. Le but de l'étude était entre autres d'examiner leur taille de vocabulaire et de discuter si la taille suffit pour évaluer le niveau français scolaire des universitaires norvégiens. Les participants ont appris le français pendant 0 à 5 ans¹². Le score moyen des mots connus était de 2451 mots. Les nombres de mots connus par les participants correspondent avec les résultats des étudiants qui participaient aux études de David (2008) et Milton (2008), qui avaient une moyenne de 2524 mots et 2555 mots, respectivement.

Les études de Milton (2008), David (2008) et Lindqvist (2017) ont beaucoup de variations au sein des niveaux testés. Nous voyons que leurs découvertes correspondent, et que la taille du vocabulaire augmente au fur et à mesure. Nous revenons aux scores de chaque groupe testé dans un tableau complet dans la partie 7.1, où nous comparons les tests plus en détail. L'augmentation de la taille est plus rapide chez les participants suédois. Lindqvist (2017) remarque que :

Ces études concernent [David, Milton], comme la nôtre, des apprenants à l'école dans un pays où le français n'est normalement pas utilisé, sauf à l'école. Il est probablement juste de dire que ni les élèves anglophones, ni les élèves suédophones, n'entendent souvent le français hors l'école. La situation d'apprentissage est donc similaire pour les

¹² 10 étudiants ont eu le français pour 5 ans, 4 étudiants pour 2-3 ans, et 2 étudiants pour 0 ans (3 moins plus précisément)

deux groupes, c'est-à-dire qu'il s'agit dans les deux cas d'apprentissage d'une langue étrangère dans un contexte formel. (p. 153)

Il est donc difficile de dire pourquoi les élèves suédois obtiennent un tel score, mais une raison peut être l'effet des mots cognats, discuté dans la prochaine partie.

Dans ce travail, nous avons peu de raisons de croire que nos résultats sont trop différents des études précédentes. Un facteur qui peut affecter notre score différemment est l'âge des participants. L'âge des débutants des autres études est d'environ 12 ans, comparé à 16 ans dans notre étude. Il est difficile de dire si ce facteur va donner des scores plus hauts ou plus bas, puisque plusieurs facteurs affectent l'apprentissage de langues à des âges différents. Concernant la croissance de la taille du vocabulaire, nous pouvons argumenter que nos participants sont plus similaires aux participants de Lindqvist (2017) et qu'il est possible que la croissance rapide montrée dans son étude soit aussi présente dans la nôtre.

4.2 Les cognats et leur effet sur des tests de vocabulaire

Dans le domaine de l'acquisition d'une troisième langue et de l'influence interlinguistique, il a souvent été montré que la L2, plutôt que la L1, a tendance à influencer la L3 (Bardel et Falk, 2010). Cependant, même s'il y a un grand nombre d'études qui examinent la relation entre deux langues, elles concernent souvent la L1 et la L2. L'anglais étant presque toujours la langue cible, nous trouvons peu de recherches sur l'inclusion d'une L3, ou de français comme langue cible. Nous trouvons, néanmoins, le travail récent de Lindqvist (2020) sur l'influence de deux langues cibles - le suédois L1 et l'anglais L2—sur une troisième langue, L3 français. L'étude s'appuie sur les précédentes recherches de Lindqvist (2017), Milton (2008) et David (2008) - présentes ci-dessus - comme les résultats de Lindqvist (2017) montraient une taille relativement plus grande que les recherches avec des participants anglais. Une différence très importante des trois études est que le français est la L3 pour les apprenants suédois, alors que pour les apprenants britanniques du français, cette langue est la L2. Dans son étude, Lindqvist (2020) montrait que la L1 et la L2 semblaient influencer le vocabulaire connu dans la L3. Simplement dit, des participants peuvent indiquer qu'ils connaissent un mot en français, plutôt grâce aux compétences en anglais ou en suédois, que grâce à leurs connaissances en

français. Conséquemment, les résultats des participants de l'étude de Lindqvist en 2017 peuvent être une surestimation par rapport à leurs compétences linguistiques en français en général. En même temps, les mots cognats en L2 anglais pour les suédois sont aussi des mots cognats pour des anglais, mais dans leur L1, donc on peut se demander si on trouve le même avantage pour ce dernier groupe. Pour éviter l'affect des cognats sur des tests de vocabulaire, Szabo (2010) suggère que les mots choisis pour les tests devraient inclure des mots cognats de manière proportionnelle dans toutes les langues, ce qui signifie que les tests devraient être adaptés spécifiquement pour des langues cibles et les langues maternelles testées.

Chapitre 5. Méthode

5.1 Participants

Les participants de notre étude sont des élèves de 1ère, 2ème et 3ème année de lycée norvégien, et font partie de cinq classes individuelles, dans deux écoles différentes. Les élèves sont âgés de 15 à 18 ans, et sont répartis sur quatre niveaux de français différents. Au total, 80 élèves ont participé à l'enquête. Pour l'analyse des résultats, nous avons exclu tout participant 1. ayant une autre langue maternelle que le norvégien, et 2. dyslexique (nous avons également exclu les participants en raison d'un trop grand nombre de réponses incorrectes dans le test, cela est expliqué plus en détail dans la section sur la procédure). Après ces exclusions, il nous reste 50 participants.

En Norvège, l'apprentissage d'une troisième langue commence soit au collège (à l'âge de 12-13 ans), soit au lycée (à l'âge de 15-16 ans), et c'est l'élève qui décide quand il veut commencer l'apprentissage. Le nombre d'années d'apprentissage et le nom du niveau dépendent de ce facteur, voir le tableau 5 pour des détails sur chaque niveau. Il est difficile de dire exactement combien d'heures de français chaque élève a eu au moment de l'étude, mais le plan éducatif national stipule que l'élève doit avoir respectivement 222 heures et 225 heures de français à la fin du collège et du lycée (Utdanningsdirektoratet, 2020a).

Tableau 5

Aperçu des niveaux et années de l'apprentissage des élèves de langues étrangères en Norvège

	Année à l'école	Année de début ¹³	Niveau ¹⁴	Ans d'étude à la fin
--	-----------------	------------------------------	----------------------	----------------------

¹³ Au collège les élèves peuvent choisir soit une langue étrangère (comme le français, par exemple), soit choisir un niveau plus avancé d'anglais ou de norvégien. Quand les élèves choisissent la première option, ils ont trois ans de français au collège (nommé français niveau 1). En commençant au lycée (âgés 15-16 ans), ils peuvent continuer avec le français pour deux (nommé français niveau 2) ou trois ans (nommé français niveau 3). Les élèves qui ont eu l'anglais ou norvégien au collège doivent commencer avec une langue étrangère au lycée (aussi nommé le niveau 1). Les élèves de niveau 2 de français lycéen peuvent choisir d'avoir le français leur dernière année au lycée, mais c'est obligatoire pour les élèves de niveau 1 de français lycéen d'avoir la langue leur dernière année au lycée (nommé le niveau 1+2)

¹⁴ Les niveaux comme ils sont nommés dans le système norvégien.

				d'année
Collège	8ème	8ème	1	1
	9ème			2
	10ème			3
Lycée	11ème	8ème	2	4
	12ème *	8ème		5*
	13ème *	8ème	3	6*
	11ème	11ème	1	1
	12ème	11ème		2
	13ème	11ème	1+2	3

* ne sont pas inclus dans notre étude

Dans cette étude, la durée d'apprentissage du français est la plus importante, et conséquemment nous avons désormais choisi de nommer les quatre groupes de participants selon le nombre d'années d'études en français. Voir le tableau 6 ci-dessous pour voir à quel niveau cela correspond au lycée norvégien et à combien d'années correspond chaque niveau. Dans cette étude, il y a 10 élèves dans leur première année d'apprentissage (année 1), 15 élèves dans leur deuxième année (année 2), 17 élèves dans leur 3ème année (année 3) et 8 élèves dans leur 4ème année.

Tableau 6

Aperçu des participants de l'étude et leur niveau

Nom du niveau dans cette étude	Nombre des participants	Nombre d'années d'études (vers la fin d'année)
Première année d'apprentissage I	10	1
Deuxième année d'apprentissage II	15	2
Troisième année d'apprentissage III	17	3

En Norvège, les élèves commencent obligatoirement l'apprentissage de l'anglais dans leur première année d'école (âgés de 6 ans), et l'enseignement de la langue dure normalement jusqu'à la fin du lycée (12 ans d'apprentissage, en total). De plus, les Norvégiens sont régulièrement exposés à l'anglais, grâce à la prévalence de la langue dans les réseaux sociaux, les films et les médias. Par conséquent, leur niveau d'anglais est généralement élevé. L'auto-évaluation du questionnaire biographique de notre étude confirme cette idée, 75 participants sur 80 ayant indiqué une maîtrise correspondant aux deux niveaux les plus soutenus (voir l'annexe II pour plus de détails). Comme dans l'étude de Lindquist (2017), l'anglais est donc considéré comme la L2 des participants.

5.2 La procédure

Nous avons contacté des professeurs des classes de deux manières différentes ; soit en contact direct, soit en contact avec l'administration du lycée. Avant de participer, les enseignants étaient informés des objectifs de l'étude, de la durée et du type de test, et les instructions pour effectuer le test ont été données lors de leur participation. Les cinq classes ont fait le test la même semaine à la fin de novembre, au début de leurs cours de français. Cela nous permet de comparer les résultats entre les groupes. Au moment de l'enquête, les élèves de la première année avaient suivi des cours de français depuis environ 4 mois.

Les enseignants ont reçu des informations sur l'anonymat et le traitement des informations des étudiants, sur le fait que l'étude était volontaire et sur la manière dont les étudiants pouvaient se retirer de l'étude, ainsi que des informations sur la manière dont ils devraient pratiquement passer le test et comment aider les étudiants s'ils avaient des questions sur le déroulement du test.

Comme Kjellmark (2021), nous avons effectué l'étude en utilisant Nettskjema¹⁵. L'enquête se compose de trois parties ; I. un formulaire d'informations sur l'étude et un formulaire de consentement (voir annexe III), II. le test de vocabulaire (voir annexe IV), et III. un questionnaire biographique (voir annexe II). Après avoir signé le formulaire de consentement, les élèves sont envoyés vers le test de vocabulaire. Nous avons trois versions différentes du test, et le site décide de façon aléatoire quelle version du test est attribuée vers l'élève. À la fin de l'enquête, les élèves remplissent un questionnaire biographique concernant les informations supplémentaires des participants, décrit davantage dans ce qui suit.

5.2.1 Le questionnaire biographique

Après le test, les participants ont répondu à un questionnaire biographique. Nous l'avons inclus afin d'acquérir les informations de base nécessaires sur chaque participant. Le questionnaire biographique est basé sur le questionnaire de Kjellmark (2021). Quelques questions et formulations étaient gardées presque comme références grâce à la pertinence et aux bonnes formulations, mais nous avons adapté les questions pour le groupe cible dans cette étude. Le questionnaire se compose de questions concernant le genre, l'âge et le niveau d'apprentissage des participants, une auto-évaluation sur leurs compétences en français, anglais (et d'autres langues), et des questions concernant combien d'heures par semaine ils sont exposés au français et à l'anglais (les cours au lycée non inclus). L'auto-évaluation se compose de cinq niveaux : basique, intermédiaire, indépendant, compétent et courant¹⁶ (voir X), avec des courtes descriptions de ce qui indique chaque niveau. Deux questions au questionnaire en particulier sont importantes pour décider quels participants seront inclus ou exclus des résultats ; (I) la langue maternelle et (II) la dyslexie.

(I) Cette étude examine le taille du vocabulaire des élèves avec norvégien comme leur seule langue maternelle. Dans son étude, Lindquist (2017) inclut des élèves ayant deux langues maternelles. Cependant, comme des études ont montré que les langues déjà acquises peuvent affecter l'apprentissage du vocabulaire (Batista & Horst, 2016, pp.18-19), nous avons décidé

¹⁵ <https://www.uio.no/tjenester/it/adm-app/nettskjema/>

¹⁶ *grunnleggende, middels, selvstendig, kompetent, flytende* en norvégien

d'appliquer ce critère strict. Il faut quand même noter que, parmi les participants, il semble y avoir une confusion générale quant à ce que signifie la langue maternelle. Plusieurs participants ont choisi l'anglais comme langue maternelle (ainsi que le norvégien), mais ont indiqué plus tard qu'ils ne le parlaient pas avec leur famille ou leurs amis proches ou qu'ils ne le parlaient qu'avec leurs connaissances. Nous avons donc examiné chaque réponse en détail pour évaluer s'ils ont vraiment deux langues maternelles. Quand ils ont aussi indiqué qu'ils parlent cette langue avec leur famille (proche ou élargie), et ont un niveau plus haut qu' "indépendant" dans leur auto-évaluation, ils ont été exclus de l'étude. (II) Nous avons aussi exclu les participants qui ont indiqué être atteints de la dyslexie. Puisque le test X-Lex est un test qui est basé sur une compétence et compréhension écrite, le diagnostic pouvait affecter leur résultats. De plus, les faux mots sont basés sur une structure syllabique plausible, ainsi qu'une vraie morphologie dérivationnelle, soulignant l'importance de l'orthographe . Nous avons exclu 7 participants basés sur le critère des autres langues maternelles et 6 basés sur le critère de la dyslexie. 3 des participants qui avaient plusieurs L1 ainsi que tous 6 participants atteints de dyslexie ont aussi coché trop de mots inventés; ces participants sont alors exclus selon deux de nos critères.

5.2.2 Le test X-Lex

Le test X-lex était terminé au cours de français et en ligne, utilisant le site Nettskjema. Nous adoptons le format de Lindquist (2017) où le participant doit cocher soit "oui, je peux reconnaître, comprendre ou utiliser ce mot en français" soit "non, je ne reconnais pas, ne comprends pas ou ne peux pas utiliser ce mot en français" pour chaque mot (voir l'annexe X). Ces réponses - ainsi que les instructions - sont formulées en norvégien pour aider les participants à comprendre le test correctement. Les participants avaient 45 minutes de disponibles pour finir le test, mais en moyenne ils ont terminé en 7 minutes et 50 secondes. Il n'était pas possible de finir le test sans répondre à chaque question. Enfin, nous avons utilisé trois versions du test X-Lex, ce qui nous permet de tester plus de mots, et d'autres montraient qu'il n'y a pas de différence statistiquement significative entre les trois versions (David, 2008). La version du test (1,2 ou 3) était randomisée par le site utilisé. Nous avons utilisé les trois versions de Milton (2009, p.257-259) comme elles étaient adaptées par Kjellmark (2021) , mais avec des modifications et corrections discutées dans ce qui suit.

5.2.2.1 La réalisation du test (et les modifications)

Si la source d'origine du test est de Milton (2009), la réalisation du test a besoin de quelques modifications, comme l'a noté Kjellmark (2021). Les modifications sont décrites en détail par Kjellmark, et nous ne faisons donc qu'un résumé des points les plus importants ici:

1. L'ordre des mots dans chaque version du test était modifié et randomisé. Premièrement, dans la présentation de Milton, les mots sont regroupés selon leur bande de fréquence. De même, tous les mots inventés sont répertoriés dans une liste à part. Il était donc nécessaire de randomiser l'ordre des mots. Nous reprenons l'ordre exact de Kjellmark (2021), qui a utilisé un site internet¹⁷ pour créer un ordre quasi randomisé. Deuxièmement, les tests contiennent quelques mots que nous considérons comme plus connus par les élèves, et ces mots étaient donc placés délibérément au début et vers la fin du test.
2. A la suite de Kjellmark (2021), nous avons exclu quelques mots répertoriés comme des non-mots dans la version de Milton mais qui sont maintenant inclus dans des dictionnaires. Il s'agit plus spécifiquement des mots **pédiment* et **satisfactoire* de la version 2 et 3 du test, et **spirité* dans la version 3 (le mot existe, mais sans l'accent aigu). Ils sont remplacés par des faux mots dans les deux autres versions du test ; **manchir*, **diroir* et **jerette*, respectivement
3. Il y a plusieurs mots dans les trois versions du X-lex de Milton avec des fautes de frappe. Nous les avons corrigées dans notre version. Voici une liste des mots des fautes de frappe et leur versions corrigées en parenthèses: **tourneé* (tournée) **inéressé* (intéressé) **rgant* (gant) **concurrén* (concurrent) **hautemen* (hautement).

5.2.3. Calcul de scores

Le score de chaque participant est calculé individuellement selon le schéma suivant:

1. Une réponse correcte pour un mot vrai donne 50 points (+50 points)
2. Une réponse incorrecte pour un mot inventé (cocher la réponse "oui") donne une déduction de 250 points, (-250 points)

¹⁷ ListRandomizer4

3. Une réponse incorrecte pour un mot vrai (cocher la réponse “non”) ne donne ni de points ni de déductions (0 points).

Donc, le score maximal du test est 5000 points (50 points par mot x 100 mots vrais). Le score maximal correspond avec le fait que le test est créé en choisissant parmi les 5000 mots les plus fréquents dans la langue française. Parmi les participants inclus dans l'étude selon les critères Selon les critères de Lindquist 2017 et David (2008), nous avons exclu les participants qui avaient coché plus de 5 mots inventés. Un nombre de réponses plus haut peut indiquer que le participant devine. A la base de ce critère, nous avons exclu (20 participants au total, 8 participants ayant complété la version 1 du test, 7 de la version 2, et 12 de la version 3). Parmi les participants inclus dans l'étude selon ce critère, il y en avait 3 avec des scores négatifs (-50, -50 et -600, respectivement). A la suite de Lindquist (2017), nous avons corrigé ces nombres à 0, étant donné qu'il est impossible de connaître un nombre négatif de mots.

Chapitre 6. Résultats

6.1. La taille de vocabulaire

Pour répondre à notre première question de recherche, “*Quelle est la taille du vocabulaire français chez des lycéens norvégiens ?*”, nous calculons la taille du vocabulaire selon la procédure décrite dans 5.2.3. De 80 participants, nous excluons 30, et 50 participants sont inclus dans notre étude. 14 participants ont complété la version 1 du test, 23 participants de la version 2 et 13 de la version 3. Les résultats varient de 0¹⁸ points/mot à 2000 points/mot, et le total moyen est de 880 mots. Le tableau 7 montre le score moyen, minimum et maximum dans les années respectives, ainsi que les écarts types.

Tableau 7

Score moyen, minimum, maximum

Année d'études	Score moyen	Min. score	Max score	Écart type
1 (n=10)	745	0	2000	761
2 (n=15)	853	0	1550	441
3 (n=17)	773	50	1850	530
4 (n=8)	1325	550	1750	374

Nous voyons que les participants des trois premières années montrent des scores moyens de vocabulaire qui sont assez similaires, avec des scores moyens de 745, 853 et 773, respectivement. Le score moyen des participants de la 4ème année est plus élevé, avec 1325 mots. Le score minimum de 550 mots dans ce groupe est aussi significativement plus haut que celui des autres groupes. Dans les trois premiers groupes, il y a une grande variation entre la taille minimum et maximum. Cela est refléchi dans l'écart type, qui est relativement haut en 1ère année, et le plus bas en 4ème année.

¹⁸ -600 corrigés à 0, comme expliqué dans la partie 5.2.3

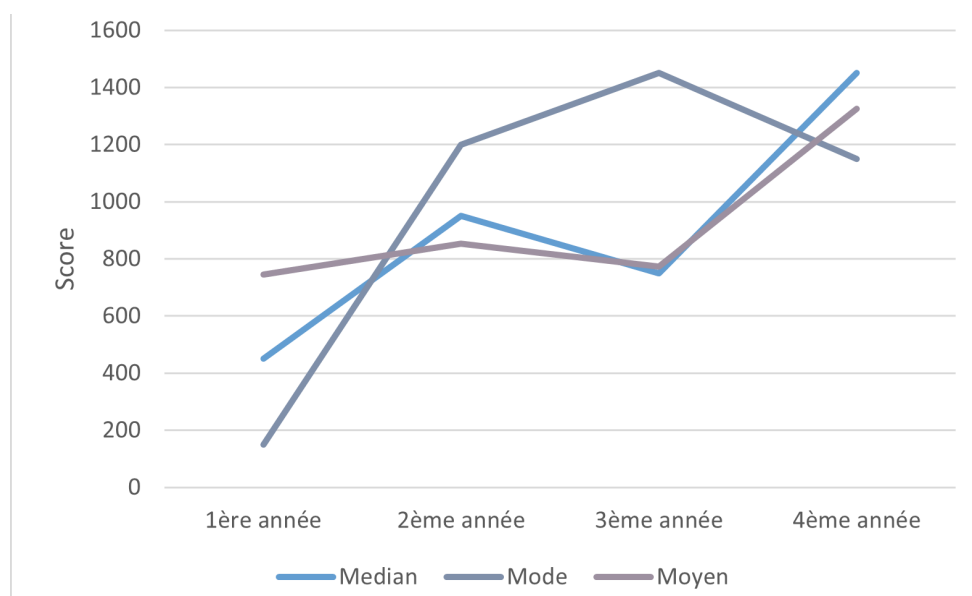
6.2. Le développement du vocabulaire

Pour répondre à notre deuxième question de recherche, “*Y a-t-il un développement de la taille du vocabulaire à travers les années d'étude ?*” une analyse ANOVA à facteur contrôlé était effectuée pour examiner si les différences entre les groupes étaient significatives. Nous effectuons l'analyse avec une mise en garde en raison de notre petit échantillon. L'analyse montrait une différence entre groupes qui n'est pas significative en utilisant 0.05 comme critère $p < 0,1 : F(3,46) = 2.31, p = .09$. En regardant la figure 3, nous pouvons voir qu'il y a un développement d'une année à l'autre, alors que la croissance n'est pas régulière. La 3ème année a un score inférieur au score de la 2ème année.

Les scores des participants de chaque groupe sont très variés. L'écart type de la 1ère année est 761, et le moyen de 745, la médiane de 450, et le mode de 150, varient beaucoup dans ce groupe, comme montré dans la figure 3. En fait, parmi les participants de la 1ère année on trouve le meilleur (2000) et le troisième (1750) scores de l'échantillon, mais aussi des scores de 0 ou 150 mots.

Figure 3

Score de médian, mode et de moyen



Le moyen est relativement similaire entre le 1ère, 2ème et 3ème année d'apprentissage, mais le moyen de la 4ème année est considérablement plus haut. Il est donc intéressant de voir si ce

groupe obtient un score différent du groupe de l'année précédente (la 3ème année) et du groupe avec le score moyen le plus élevé (la 2ème année), et un t-test unilatéral était donc effectué pour montrer s'il y a des différences significatives entre les groupes. Les différences sont significatives:

T-test 4ème et 3ème année $t(23) = -2.63, p = .007$.

T-test 4ème et 2ème année $t(21) = -2.69, p = .006$.

Nous ne pouvons donc pas conclure que la taille du vocabulaire augmente d'une façon constante à travers les années. Nous pouvons cependant constater que la taille du vocabulaire des participants de la 4ème année est en fait significativement plus haute que la 2ème et la 3ème année. Il faut noter que la taille des groupes de participants varie, et que le groupe des apprenants de 4ème année n'est pas nombreux.

Nous avons aussi calculé le pourcentage de connaissance à chaque niveau et pour chaque bande de fréquence. La figure 4 montre les pourcentages des mots connus à chaque niveau, et la figure 5 montre le résultat total de tous les tests. Les nombres suggèrent que les apprenants connaissent le plus de mots dans la bande 1K, avec une diminution progressive des mots connus vers la bande 5K, et cela correspond à la suggestion de Milton (2009), discutée dans la partie 3.3.1. Nous observons aussi dans la figure 4 que le connaissance de chaque bande de fréquence est relativement similaire dans les années 1-3. La différence est la plus élevée dans le bande 1K et 2K, mais s'égalise aux bandes 3K-5K. Dans les bandes 3K et 5K, les apprenants de 1ère année ont le score le plus élevé des trois premiers groupes. Ce groupe a généralement une distribution plate dans des bandes de fréquence, probablement en raison de la grande variation de score des apprenants.

Figure 4

Connaissance des bandes de fréquence de chaque année d'apprentissage

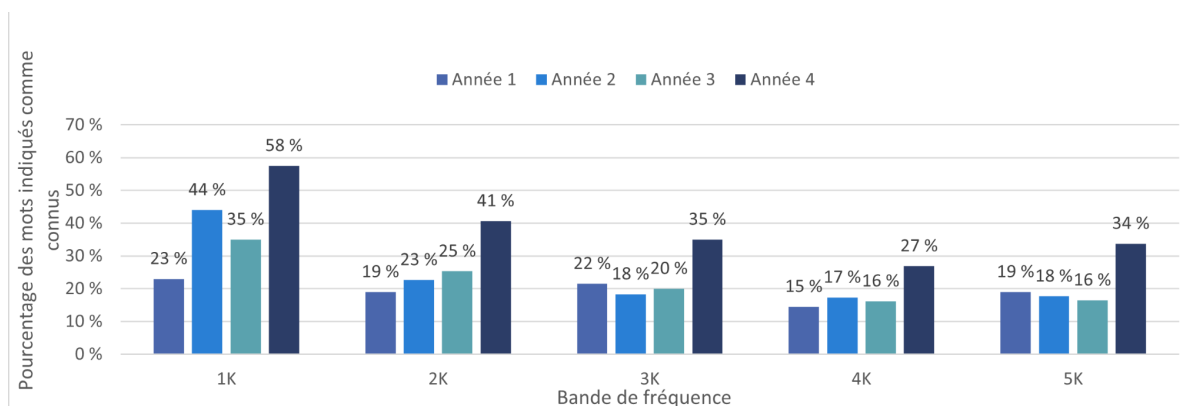
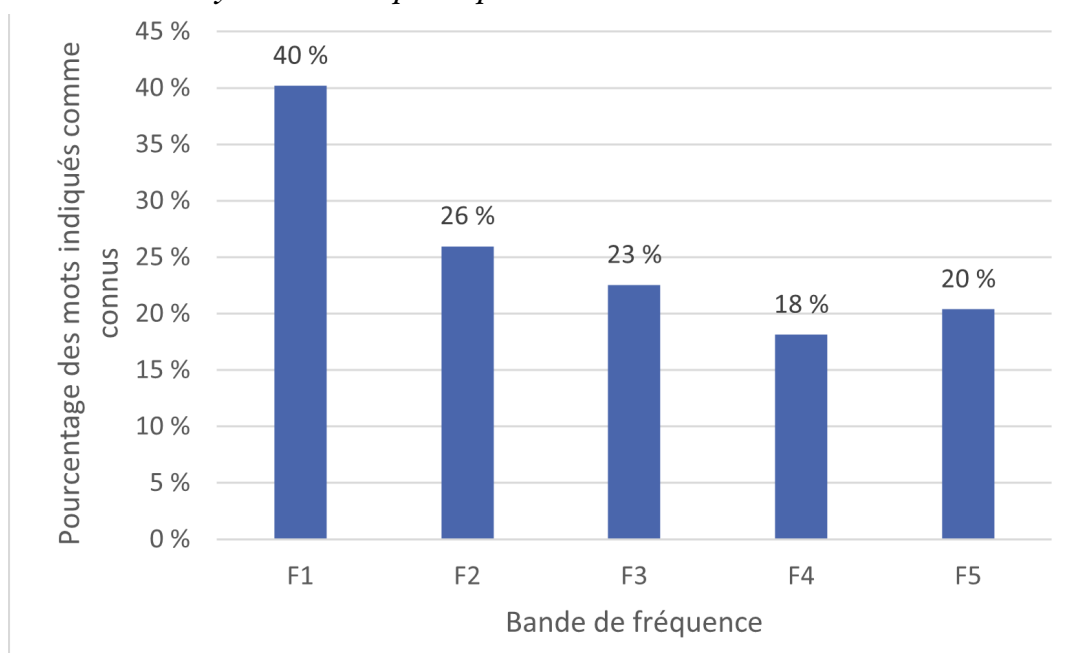


Figure 5

Connaissance moyen de tous les participants



Chapitre 7. Discussion

Dans cette section, nous allons explorer les découvertes et résultats présentés dans la partie précédente. Tout d'abord, nous comparons nos résultats avec des études de Milton (2008), David (2008), Lindqvist (2017) et Kjellmark (2021), en discutant nos questions de recherche. Ensuite, nous discutons des conséquences pédagogiques de nos résultats. Enfin, nous examinons des facteurs du test qui peuvent affecter nos résultats, et finalement, nous traitons des limites de cette étude.

7.1 Comparaison des résultats

Nous revenons à notre première question de recherche: *quelle est la taille du vocabulaire français chez des lycéens norvégiens?* Dans le chapitre 4, nous avons présenté des études antérieures qui examinent la même problématique en utilisant la même méthode et des participants similaires. Nous avons estimé que nos résultats sont comparables aux résultats de David (2008), Milton (2008) et Lindqvist (2017). L'étude de Kjellmark (2021) est la seule étude similaire menée en Norvège, mais puisque ses participants sont plus âgés et d'un niveau universitaire, notre hypothèse était que les participants de notre étude n'obtiendront pas des scores comparables à ceux des participants de Kjellmark (2021). Nous allons donc maintenant principalement comparer nos résultats à ceux des 3 autres études mais nous discutons des découvertes de Kjellmark (2021) dans la section 7.2. Il convient de mentionner que nous comparons les scores avec prudence, car notre nombre de participants est considérablement inférieur à celui des autres études.

Le tableau 8 est un aperçu de scores moyens, maximum, minimum et l'écart type de toutes les études mentionnées précédemment. Nous avons indiqué le nombre d'années d'apprentissage de français à chaque groupe, mais il faut noter que l'âge des participants varie de 12 à 18 ans, et que les participants de notre étude avaient entre 16 et 18 ans. En comparant les scores des niveaux plus bas - c'est-à-dire des scores des élèves d'un an d'apprentissage - les scores

moyens de Milton (2008) et Lindqvist (2017) sont 311 et 232, respectivement¹⁹. Le score moyen des apprenants de 1ère année de notre étude est 745. Cependant, les participants de ces groupes de Milton (2008) et Lindqvist (2017) ont 12 ans, et les nôtres ont 16 ans. Des études récentes ont montré que les apprenants plus âgés acquièrent plus rapidement plusieurs aspects linguistiques, et ils sont aussi plus efficaces dans la réalisation des tâches scolaires (Catalan et al. 2014; Cenoz 2002; Muñoz 2008), et cela peut expliquer la différence de scores des apprenants qui sont au même niveau mais qui ont des âges différents. Nous pouvons aussi comparer nos résultats des groupes d'âge plus similaire, et plus d'années d'études. Les participants de Milton (2008) et David (2008) qui ont 16 ans ont eu 5 ans d'apprentissage, et les participants de Lindqvist qui ont 15-16 ans ont eu 4 ans d'apprentissage, et leurs scores moyens sont 825, 1577 et 1150, respectivement. Le groupe des apprenants de 4ème année de notre étude a un score de 1325. Nous voyons donc que ce score est plus proche des scores des apprenants au même niveau dans les autres études. Cependant, il est aussi probable que la différence du score des élèves de 1ère année d'études n'est due qu'à l'âge des participants, mais que cette différence est due de la grande variation de scores de nos participants de 1ère année. L'écart type de cette groupe est 761, quand Milton (2008) et Lindqvist (2017) ont des écarts types de 193 et 171, respectivement. Cette variance se reflète dans la différence entre les scores les plus bas et les plus hauts. Toutes ces études ont un score minimum de 0, mais le score maximum de notre groupe est 2000, comparé à un score maximum de 900 (Milton, 2008) et 850 (Lindqvist, 2017). Une combinaison des petits groupes et des individuels avec des scores très hauts peut créer des scores moyens gonflés. Les scores de notre étude sont généralement plus élevés que les autres études, mais plus similaires aux résultats des apprenants suédois de Lindqvist (2017) qu'à ceux des apprenants anglais de Milton (2008) et David (2008). Regardez par exemple les scores de 4ème année: 1325 (notre étude) et 1150 (Lindqvist, 2017) contre 529 (Milton, 2008) et 564 (David, 2008). Il faut encore une fois souligner que la différence d'âge et la petite taille de notre échantillon peuvent affecter cet écart.

¹⁹ L'étude de David (2008) n'a pas d'apprenants de 1ère et 3ème année et nous ne pouvons donc pas comparer ces niveaux.

Tableau 8*Aperçu des scores des études antérieures et cette étude*

	années d'apprentiss age*	n	min.	max.	moyen	écart-type
Milton	1 an	80	0	900	311	193
Milton	2 ans	78	0	1150	411	228
Milton	3 ans	106	0	1150	460	261
Milton	4 ans	57	50	1350	592	263
Milton	5 ans	49	0	1800	852	440
Milton	6 ans	10	1100	2000	1555	349
Milton	7 ans	69	650	3100	1930	475
David	2 ans	28	-200	1800	439	532
David	4 ans	26	-200	1650	564	352
David	5 ans	40	350	2500	1577	466
David	6 ans	33	1200	3300	2108	506
Lindqvist	1 an	33	0	850	232	171
Lindqvist	2 ans	52	0	1450	547	293
Lindqvist	3 ans	32	0	1900	678	386
Lindqvist	4 ans	34	400	2500	1150	460
Kjellmark	0-5 ans**	16	1323	3600	2451	659.8
Lie	1 an	10	0	2000	745	761
Lie	2 ans	15	0	1550	853	411
Lie	3 ans	17	50	1850	773	530
Lie	4 ans	8	550	1750	1325	374

*Nous indiquons le nombre d'années d'apprentissage à la fin d'année scolaire pour tout le tableau, dans le but de simplifier la comparaison entre groupes et la discussion qui suit, comme

expliqué dans le chapitre 5. Les élèves du niveau plus bas dans l'étude de Lindqvist (2017), par exemple, ont en réalité seulement eu le français pendant 1 semestre, et les participants du niveau le plus bas de notre étude ont étudié la langue pendant environ 4 mois.

** 10 étudiants ont étudié le français pendant 5 ans, 4 étudiants pendant 2-3 ans, et 2 étudiants pendant 0 an.

Nous revenons aussi à notre deuxième question de recherche: *y a-t-il un développement de la taille du vocabulaire pendant les années d'étude?* Nous n'avions aucune raison de croire que la taille du vocabulaire n'augmentera pas d'année en année, comme l'ont montré des études antérieures. Cependant, nous ne pouvons pas observer la même croissance dans notre étude comme l'étude de Milton (2008) et David (2008), non plus la croissance plus rapide et croissante chez Lindqvist (2017) (nos scores des trois premières années sont 745, 853 et 773). C'est probable qu'une des raisons principales pour cela est la grande variation interne des groupes, comme déjà discuté au-dessus. Dans la section 6.1.2, nous avons vu qu'il n'y a pas une différence significative entre tous les groupes, mais nous pouvons quand même voir qu'il y a un développement, et il y a une différence significative entre la 3ème année et la 4ème année, avec des scores de 773 et 1325, respectivement. Le croissance de 552 mots après un an d'étude est similaire à la croissance rapide de Lindqvist de 678 à 1150 mots (un croissance de 472 mots). De plus, le groupe des apprenants de 4ème année se distingue des autres groupes. En examinant les participants qui sont exclus ayant coché trop de mots inventés, 70% sont des élèves qui ont appris la langue pendant 4 ans²⁰. Conséquemment, si plus de ces élèves avaient été inclus dans l'étude, leur score ne serait peut-être pas aussi élevé, et la croissance n'aurait pas été si importante. C'est difficile d'expliquer pourquoi autant de participants exclus font partie de ce groupe, mais il a récemment été démontré que les apprenants plus âgés ont une plus grande tendance à utiliser et à reconnaître les mots cognats que les apprenants plus jeunes (Catalán et al. 2014). Plusieurs des mots inventés sont créés avec l'intention d'être similaires aux vrais mots - c'est-à-dire des cognats, plus discutés dans la partie 7.3.1. De plus, Beeckman et al. (2001) ont montré que des élèves d'un niveau plus haut ont la capacité de rejeter moins de mots inventés que les débutants, et nous discutons de ces deux facteurs dans la partie 7.3.

²⁰ 18 participants sur 26 élèves de 4ème année ont été exclus

7.2 Des conséquences pédagogiques

Dans la partie 3.2.2, nous avons discuté du nombre de mots qui est atteint aux différents niveaux du CECRL. Nous avons alors conclu qu'un vocabulaire d'environ 1 400 mots est attendu, tant au niveau A1 qu'A2. Pour les apprenants de la 1ère à la 3ème année, ce but n'est pas atteint, mais le score des apprenants de 4ème année est proche. Il faut quand même se demander si une telle taille est réaliste pour les niveaux débutants, pour 3 raisons. Premièrement, les scores des études antérieures ne montrent pas non plus une telle taille. Deuxièmement, contrairement à l'anglais, être exposé au français en dehors de la salle de classe est rare en Norvège. Les réponses au questionnaire biographique ont confirmé cela, et presque tous les participants répondent qu'ils sont exposés au français soit jamais, soit 0 à 1 heure par semaine (contrairement à l'anglais, où la plupart ont répondu 6 heures ou plus). Troisièmement, la taille de 1400 est la moyenne de plusieurs scores, de différentes langues maternelles et il faut comparer avec prudence. Il faut aussi noter que la capacité de communication et la taille du vocabulaire sont deux qualités qui sont liées, mais qu'il y a aussi des qualités qui peuvent être séparées les unes des autres. Comme dit Milton (2010):

It will be recalled that vocabulary size and communicative performance are separable qualities and one interpretation of this variation is that students with the same or similar vocabulary sizes may make different use of this knowledge to communicate more or less successfully (p.225)

C'est-à-dire qu'il y a des apprenants qui peuvent utiliser leur vocabulaire d'une manière plus efficace que d'autres apprenants avec un vocabulaire plus grand. Néanmoins, vu à la lumière des études présentées dans la section 3.1, nous pensons qu'il est important et juste que l'accent soit de plus en plus mis sur le vocabulaire et l'enseignement du vocabulaire explicite. La plupart des étudiants ne poursuivent pas leurs études en français, mais certains le font, et c'est important qu'ils soient suffisamment préparés pour le niveau universitaire. Le score moyen des étudiants au niveau universitaire de Kjellmark (2021) est 2451. Regardant le score de 1325 des participants de 4ème année, il y a un grand écart entre le niveau le plus haut de lycée, et le niveau universitaire. Néanmoins, ces élèves ont toujours un an de français à l'école, et il est possible que

leur niveau s'élève suffisamment pendant cette période. Cependant, les élèves de 3ème année, terminent en effet leur dernière année de lycée, n'ont obtenu qu'un score moyen de 773²¹. Nous pouvons donc nous demander si leur taille de vocabulaire acquis au lycée est suffisante pour être capables d'étudier la langue à l'université sans grandes difficultés.

7.3 Facteurs du test X-lex qui peuvent affecter nos résultats

Dans la section 3.2.3 nous avons traité les avantages et faiblesses générales de test X-Lex (Meara et Milton, 2003). Dans cette partie, nous examinons 3 défis spécifiques liés à l'exécution de notre étude, et qui aura possiblement affecté nos résultats. Premièrement, nous discutons de l'effet des mots cognats, deuxièmement nous discutons des raisons possibles pour le grand nombre des mots inventés qui étaient cochés, et finalement, nous regardons brièvement les différences entre les trois versions du test.

7.3.1 *L'effet des mots cognats*

Dans la section 4.2, nous avons traité des mots cognats, et comment ils peuvent affecter le nombre des mots qui sont connus par un apprenant. Le participant peut indiquer qu'il connaît un mot dans la langue cible, grâce aux compétences dans la L1 (ou bien le L2). Dans l'étude de Kjellmark (2021), elle a signalé une tendance très intéressante concernant les mots inventés : les mots inventés qui ressemblent à des mots anglais - c'est-à-dire des "cognats" - sont les faux mots les plus cochés. Nous observons aussi la même tendance dans nos résultats. Étant donné que ces mots n'existent pas en réalité, ils ne sont pas des vrais cognats, mais on peut observer qu'ils ont un peu le même effet. Nous n'entrons pas trop dans les détails, mais le tableau 9 résume les 3 mots inventés les plus cochés dans chaque version du test. Tous les mots ont une racine similaire à l'anglais (qui est certainement connue des participants), combinée à une morphologie française. De plus, certains ont aussi une tige en français, mais elle est combinée avec la mauvaise morphologie.

²¹ Voir la partie 5.1

Tableau 9

Les mots inventés les plus cochés et leur mot équivalent anglais en parenthèses

Version 1 (n=14)		Version 2 (n=23)		Version 3 (n=13)	
entrance (entrance)	10	ministeur (minister)	9	talenté (talented)	5
expecter (expect)	8	litéracie (literacy)	4	rescuer (rescue)	5
disabilité (disability)	9	fronter (front)	3	ultimation (ultimation)	2

De plus, dans les figures 4 et 5, nous observons une petite augmentation entre la bande 4K et 5K. cette augmentation peut être à cause d'un bon nombre des mots cognats dans la bande 5K. Les mots les plus cochés de cette bande ont tous des équivalents anglais. Le tableau 10 montre les quatre mots les plus cochés dans la bande 5K de chaque version du test, et leur cognat anglais en parenthèses.

Tableau 10

Les quatre mots dans la bande 5K les plus cochés et leur mot équivalent anglais en parenthèses

Version 1 (n=14)		Version 2 (n=23)		Version 3 (n=13)	
élaboration (elaboration)	6	divorce (divorce)	15	financer (finance)	6
innocent (innocent)	8	célébrité (celebrity)	13	habitation (habitation)	8
animation (animation)	11	différemment (differently)	7	réflexe (reflex)	6
radio (radio)	10	succession (succession)	12	attachement (attachement)	9

Comme déjà expliqué, les mots cognats peuvent affecter les résultats des tests de la taille du vocabulaire, et donner des surestimations sur les compétences des participants dans la langue cible. Dans nos résultats, nous avons observé quelques tendances atypiques qui peuvent s'expliquer à cause de ce facteur. Premièrement, la distribution des mots connus selon les bandes de fréquence étaient traités dans la partie 5.2. Généralement, les groupes ont montré des profils

lexicaux typiques ; le pourcentage des mots connus de chaque bande de fréquence diminue vers les bandes des mots moins fréquents, et cela correspond à l'hypothèse bien établie mentionnée dans la partie 3.3.1. Cependant, il y a une tendance dans 3 groupes sur 4 à ce que le pourcentage des mots connus augmente en peu entre les bandes de 4K et 5K. Cette augmentation peut être soit à cause d'un grand nombre de mots cognats anglais et norvégien, soit à cause de mots qui ne sont généralement pas courants, mais souvent enseignés dans les cours de langues étrangères (*pardon* et *habillé*, par exemple). De plus, dans la partie 5.2 nous avons aussi expliqué que ce n'était pas seulement les vrais mots cognats qui sont souvent choisis, mais aussi les mots inventés qui ressemblent aux mots anglais. Nous pouvons nous demander si un tel nombre de mots inventés auront été choisis sans leur statut de cognats. D'un côté, il est possible qu'il y ait des surestimations des scores à cause des cognats, mais de l'autre côté, il est possible que des participants soient évités de l'étude - pas parce qu'ils ont deviné - mais à cause du statut de cognat des mots. Ces deux tendances correspondent à la théorie de Lindqvist (2020), discutée dans la partie 4.2. Les cognats peuvent affecter (I) les nouveaux apprenants ainsi que (II) les apprenants plus âgés. Les facteurs d'âge et de niveau ne sont pas toujours linéaires, mais dans notre étude, les apprenants plus jeunes sont des apprenants de niveau plus bas, et vice versa. Concernant les nouveaux apprenants (I), Meara (1996) dit : "The test does not perform well with low-level learners, who respond unpredictably to the pseudowords. Certain learners obtain very low scores as a result of their overwillingness to claim knowledge of the pseudowords" (cité dans Beeckman et al. 2001, p.240). Concernant les apprenants plus âgés (II), l'étude de Catalán et al. (2014) déjà mentionné indique que les apprenants plus âgés ont aussi le tendance d'être affectés par les mots cognats. Pour éviter des surestimations des scores, Szabo (2008) suggère que les tests de vocabulaire doivent inclure des mots cognats de manière proportionnelle dans toutes les langues (aussi dans les mots inventés). Conséquemment, les tests devraient être adaptés spécifiquement pour des langues cibles et des langues maternelles testées.

7.3.2 Le nombre des participants qui répond "oui" aux mots inventés

Ayant discuté de l'effet des mots cognats sur le nombre des mots indiqués comme connus, nous allons aussi discuter comment les mots inventés peuvent aussi avoir le même effet comme des cognats. Dans notre étude, 26 participants sont exclus pour avoir coché plus de 5

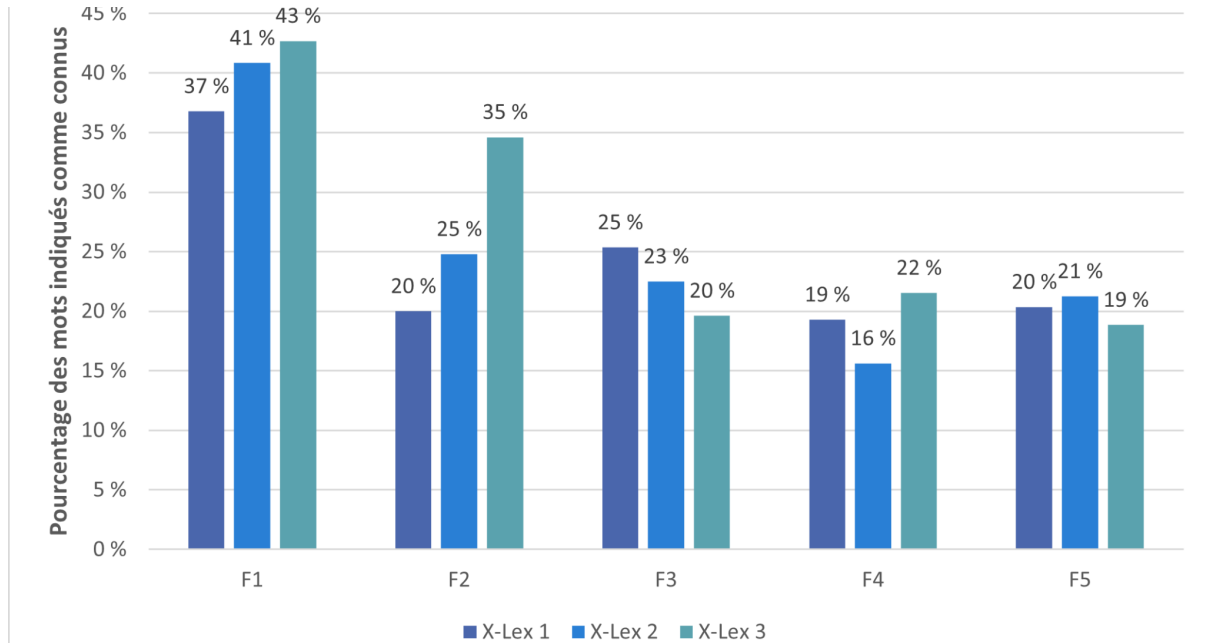
mots inventés. Quelques participants ont coché un peu plus de 5 mots, mais un grand nombre ont indiqué qu'ils connaissent 10 mots inventés, ou plus. Dans l'étude de Beeckman et al. (2001) mentionné dans la partie 3.3.3. sur la qualité et la validité du test, ils ont montré 3 tendances très intéressantes concernant les mots inventés. (I) Un grand nombre de sujets ont affiché un taux élevé de fausses réponses; (II) ces taux élevés de fausses réponses n'étaient pas limités aux élèves faibles ; et (III) il semble y avoir une relation inverse inquiétante entre la capacité à identifier les mots et la capacité à rejeter les mots inventés, alors que la logique du format s'attend à un résultat opposé. Comme mentionné dans la partie précédente, 70% des participants exclus de l'étude étaient des apprenants de 4ème année. Nous nous attendions à ce que les élèves de niveau plus haut aient une meilleure capacité à rejeter les mots inventés, mais dans notre étude, cela n'est pas le cas, et l'effet montré par Beeckman et al. (2001) semble donc en accord avec nos découvertes.

7. 3. 3 La différence entre les trois version du test

La figure 6 montre les résultats en pourcentages des trois versions du test. Rappelons que les différentes versions du test ont été choisies au hasard par le site. Nous observons ici qu'il y a une différence relativement grande entre les trois tests. Cette différence peut s'expliquer par les facteurs suivants: (I) plus des personnes ont été renvoyées vers la version 2 du test (32 participants contre 25 et 23), et cette version du test avait le plus petit nombre de réponses qui ont été retirés, résultant d'un nombre de participants plus haut en global, (II) il n'y a qu'un participant du niveau le plus bas (apprenants de 1ère année) dans la 3ème version du test, (III) la plupart des apprenants de 1ère année ont rempli la 1ère version du test. Nous observons aussi qu'il y a une différence de 5% dans le version 1 du test, entre les bandes de fréquence 2 et 3. Cette différence peut être due aux mêmes raisons discutées au-dessus, notamment en raison d'un bon nombre de cognats norvégiens et anglais dans la 3ème bande (*grupper, silhouette, muscle, formuler, classer et progresser*, par ex.) La tendance est quand même la même, et les participants connaissent plus de mots les plus fréquents.

Figure 6

Pourcentage des résultats de chaque bande de fréquence à trois versions du test



7.4 Les limites de cette étude

Dans la partie 3.3.3 nous examinons des limites et défis du test X-lex, et dans la partie 7.3 au-dessus nous discutons des facteurs spécifiques liés à notre étude qui peuvent avoir influencé notre résultat. Dans cette section, nous discutons des trois limites principales qui peuvent affecter notre étude.

Premièrement, une mise en garde qu'il faut mentionner est les estimations. On ne peut jamais être sûr que les mesures de la taille du vocabulaire sont exactes, et il faut conséquemment comparer les résultats des études antérieures avec prudence. Néanmoins, nous avons essayé d'utiliser des méthodes les plus similaires possibles. De plus, l'âge des participants de notre étude et les études antérieures peuvent affecter les résultats dans des manières déjà mentionnées, et les comparaisons deviennent encore plus complexes. Deuxièmement, même si la réponse de chaque

mot du test est soit “oui”, soit “non”, l’interprétation de la réponse peut varier. Comme dit Kjellmark (2021):

Si un participant est en doute, il peut cocher oui mais un autre peut opter pour non même s’il a une idée de la signification du mot. Plusieurs études ont démontré que les apprenants dans différents pays affichent des tendances différentes quant à l’évaluation de leurs propres connaissances. Il est difficile de dire si ces nuances pourraient ou non avoir un impact sur les résultats du test, mais pour éliminer cette possibilité, il serait peut-être avantageux que les études futures s’en tiennent à un seul format aussi proche que possible afin de faciliter la comparaison de trouvaillles. (p.50-51)

Enfin, un des plus grands défis de notre étude est le nombre de participants. Idéalement, il aurait été préférable avec un échantillon plus large que 50, pour mieux représenter les variations des élèves lycéens en Norvège. Nous espérons que plus d’études similaires seront effectuées en Norvège et dans les autres pays scandinaves à l’avenir.

Chapitre 8. Conclusion

Le présent travail a examiné la taille de vocabulaire en français des élèves norvégiens au secondaire, et a montré qu'en moyenne, les apprenants de 1ère année connaissent 745 mots, les apprenants de 2ème année 853 mots, les apprenants de 3ème année 773 mots et les apprenants de 4ème année 1325 mots. Nous avons comparé nos scores avec d'autres études similaires menées en Angleterre (Milton, 2008 ; Davis, 2008) et en Suède (Lindqvist, 2017). Nous avons comparé les scores moyens pour la même année d'apprentissage dans toutes les études, et nos participants avaient généralement un score plus élevé que dans les autres études. Cependant, nos participants avaient quelques années de plus, nous avons donc également comparé les scores d'apprenants du même âge, mais dans des années d'apprentissage différentes. Dans cette comparaison, les scores des autres études étaient plus similaires au score des apprenants de 4ème année, dans notre étude. Nous avons également comparé la croissance de mots, et nous n'avons pas observé la même croissance dans notre étude comme l'étude de Milton (2008) et David (2008), non plus la croissance plus rapide et croissante chez Lindqvist (2017) Il est probable qu'une des raisons principales à cela est la grande variation interne dans chaque groupe, et de petit échantillon.

D'après une étude de Milton et Alexiou (2009), les élèves d'un niveau CECRL A1-A2 devraient idéalement être capables de connaître 1400 mots. Comme le montre notre étude, nos apprenants de 4ème année ont presque atteint cette taille de vocabulaire, mais nous posons toujours la question de savoir si cette taille de vocabulaire est réaliste avec le peu de temps consacré à l'apprentissage des langues étrangères dans le système scolaire norvégien.

Des tests de vocabulaire ne sont pas sans faiblesses, et comme plusieurs études antérieures, nous constatons aussi qu'il faut les améliorer. Par exemple, nous avons observé que 70% des participants qui étaient omis de l'étude sont des participants de 4ème année, et nous posons la question qu'une des raisons pour cela est le grand nombre des mots inventés qui fonctionnent de la même manière que des mots cognats. Nous suggérons aussi de mettre à jour les listes de fréquences sur lesquels les tests de vocabulaire se basent, et de faire attention à la proportion des mots cognats ; ce nombre doit premièrement être similaire dans toutes les

versions du test utilisées dans une seule et même étude, mais doit aussi être similaire aux tests avec des langues cibles différentes - pour faire de comparaisons plus facilement. Notre étude est de trop petite taille pour pouvoir tirer des conclusions définitives sur le vocabulaire des élèves du secondaire supérieur en Norvège, néanmoins, c'est la première étude à enquêter sur ce sujet en Norvège, et nous espérons qu'elle peut servir de base à de futures études qui approfondiront le sujet important qui est le vocabulaire.

Bibliographie

- Anderson, R.C. and Freebody, P. (1981) Vocabulary knowledge. En J.T. Guthrie(dir.) *Comprehension and Teaching: Research Reviews*(pp. 77-117). International Reading Association
- Alderson, & Alderson, P. J. C. (2006). *Diagnosing Foreign Language Proficiency: The Interface Between Learning and Assessment*. Bloomsbury Publishing Plc.
- Altman, G. T. M., (1990) Cognitive Models of Speech Processing: An Introduction. En Altman, G. T. M., *Cognitive Models of Speech Processing Psycholinguistic and Computational perspectives*. (pp.2-23) Cambridge: MIT Press
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). ASPECTS OF LEXICAL SOPHISTICATION IN ADVANCED LEARNERS' ORAL PRODUCTION. *Studies in second language acquisition*, 34(2), 269-290. <https://doi.org/10.1017/S0272263112000058>
- Baudot, J. (1992). *Fréquences d'utilisation des mots en français écrit contemporain*. Presses de l'Université de Montreal.
- Batista, R. et Horst, M. (2016). A New Receptive Vocabulary Size Test for French. *Canadian modern language review*, 72(2), 211-233. doi: 10.3138/cmlr.2820
- Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M. et Velde, H. V. D. (2001). Examining the Yes/No vocabulary test: Some methodological issues in theory and practice. *Language testing*, 18(3), 235-274. doi: 10.1177/026553220101800301
- Bogaards, P., & Laufer-Dvorkin, B. (2004). *Vocabulary in a second language : selection, acquisition, and testing*. John Benjamins.
- Cenoz. (2007). C. Muñoz (ed.): AGE AND THE RATE OF FOREIGN LANGUAGE LEARNING. *Applied Linguistics*, 28(4), 627–630. Oxford University Press. <https://doi.org/10.1093/applin/amm038>
- Cobb, T. et Horst, M. (2004). Is there room for an academic word list in French? Dans P. Bogaards & B. Laufer-Dvorkin (dir.), *Vocabulary in a second language : selection, acquisition, and testing* (vol.10). John Benjamins.
- Conseil de l'Europe. (2020). Cadre européen commun de référence pour les langues - Apprendre, Enseigner, Évaluer (CECRL).
- Daller, H., Milton, J. et Treffers-Daller, J. (2007). Editor's introduction: conventions, terminology and an overview of the book. Dans H. Daller, J. Milton & J. Treffers-Daller (dir.), *Modelling and assessing vocabulary knowledge* (pp. 1-32). Cambridge University Press
- David, A. (2008). Vocabulary breadth in French L2 learners. *Language learning journal*, 36(2), 167-180. doi: 10.1080/09571730802389991
- Dictionnaires Le Robert (2021) Mot. Dans J. Rey-Debove et A. Rey (dir.) *Le Nouveau Petit Robert de la langue française*. Paris.
- Falk, Y., & Bardel, C. (2010). The study of the role of the background languages in third language acquisition. *The state of the art. IRAL - International Review of Applied Linguistics in Language Teaching*, 48(2-3), 185-219. <https://doi.org/10.1515/iral.2010.009>

- Gass, S. M., Behney, J., & Plonsky, L. (2013). *Second language acquisition : an introductory course* (4e éd.). Routledge
- Gyllstad, H., Vilkkaitė, L., & Schmitt, N. (2015). Assessing vocabulary size through multiple-choice formats: Issues with guessing and sampling rates. *ITL. Instituut voor Toegepaste linguïstiek*, 166(2), 278-306. <https://doi.org/10.1075/itl.166.2.04gyl>
- Horst, N., et Collins, L. (2006) From Faible to Strong: How Does Their Vocabulary Grow? *The Canadian Modern Language Review*, 63 (1), 83-106. DOI: 10.3138/cmlr.63.1.83
- Hu, Marcella & Nation, Paul. (2000). Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language*, 13 (1) 403-430.
- Juffs, A. (2009). Second language acquisition of the lexicon. Dans W.C. Ritchie, et T.K. Bathia, (dir.), *The New Handbook of Second language acquisition*. Emerald Group Publishing Limited.
- Kremmel, B. (2016). Word families and frequency bands in vocabulary tests: Challenging conventions. *TESOL Quarterly*, 50(4), 976–987.
- Laufer, B. (1989). What percentage of text lexis is essential for comprehension? Dans C. Lauren & M. Nordman (dir.), *Special language: From humans thinking to thinking machines* (p. 316–323). Multilingual Matters
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? Dans H. Bejoint & P. Arnaud (dir.), *Vocabulary and applied linguistics* (pp.126–132). Macmillan.
- Laufer, & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, 16(1), 33–51. <https://doi.org/10.1177/026553229901600103>
- Laufer, B., & Goldstein, Z. (2004). Testing Vocabulary Knowledge: Size, Strength, and Computer Adaptiveness. *Language learning*, 54(3), 399-436. <https://doi.org/10.1111/j.0023-8333.2004.00260.x>
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a foreign language*, 22(1), 15.
- Laufer, B. (2021). Lexical Thresholds and Alleged Threats to Validity: A Storm in a Teacup? *Reading in a Foreign Language*, 33(2), 238.
- Levelt, W. J. M. (1989). *Speaking : from intention to articulation*. MIT Press.
- Lindqvist, C. (2020). First and Second Language Cognate Effects in Third Language Vocabulary Size Estimates. Dans P. B. Booth et J. Clenton (dir.) *First Language Influences on Multilingual Lexicons*. (1e éd., p. 149-161). Routledge.
- Lonsdale, D. et Le Bras, Y. (2009). *A frequency dictionary of French: Core vocabulary for learners*. Routledge
- Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, 20, 51–81
- Nation, I.S.P. (1983). Testing and teaching vocabulary. *Guidelines*, 5(1), 12–25.
- Nation, I.S.P. (1993). Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures. *Language Testing*, 10(1), 27-40
- Meara, P. M et Jones, G. (1990) *The Eurocentres Vocabulary Size Tests: 10KA*. Eurocentres.
- Meara, P. (1992) *EFL Vocabulary Tests*. University College Swansea: Centre for Applied Language Studies
- Meara, P. M. et Milton, J. (2003). *X-lex: the Swansea levels test*. Express Publishing

- 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(3), 333-348. doi: 10.1017/S0959269508003487
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. *Eurosla Monographs Series*, 1, 211-231.
- Milton, J. et Alexiou, T. (2020). Vocabulary Size Assessment: Assessing the Vocabulary Needs of Learners in Relation to Their CEFR Goals. Dans M. Dodigovic et M. P. Agustín-Llach (dir.), *Vocabulary in Curriculum Planning: Needs, Strategies and Tools* (p. 9-27). Palgrave Macmillan.
- Miralpeix, I. et Muñoz, C. (2018) Receptive vocabulary size and its relationship to EFL language skills. *IRAL*, 56(1), 1–241. <http://doi.org/10.1515/iral-2017-0016>
- Muñoz, C. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 29(4): 578–596.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2e éd.). Cambridge University Press.
- Ortega, L. (2013). *Understanding second language acquisition*. Routledge.
- Otwinowska. (2018). Cognate Vocabulary in Second Language Acquisition. In *Cognate Vocabulary in Language Acquisition and Use* (pp. 83–93). Multilingual Matters. <https://doi.org/10.21832/9781783094394-009>
- Palmer, H.E. (1917) *The Scientific Study and Teaching of Languages*. Harrap
- Paribakht, T.M. and Wesche, M. (1993) The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, 11, 9-29.
- Pignot-Shahov, V. (2018) *L2 French lexical development of undergraduate students in a UK university [Thèse de doctorat. University of Southampton.]* EPrints. <http://eprints.soton.ac.uk/id/eprint/432082>
- Qian, D. D. et Lin, L. H. F. (2020) The Relationship Between Vocabulary Knowledge and Language Proficiency. Dans Webb, S. (dir.) *The Routledge Handbook of Vocabulary Studies*. Routledge.
- Riegel, M., Pellat, J.-C., & Rioul, R. (2018). *Grammaire méthodique du français* (7e édition.e éd.). Presses Universitaires de France.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal (Boulder, Colo.)*, 95(1), 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language learning journal*, 36(2), 139-152. <https://doi.org/10.1080/0957173080238997>
- Szabo, C. Z. (2020). The Reassessment of Vocabulary Tests Based on Cognate Distribution. Dans *First Language Influences on Multilingual Lexicons* (1e éd., p. 162-190). Routledge.
- Teichroew, F. J. (1982). Receptive versus productive vocabulary: A survey. *Interlanguage Studies Bulletin*, 6, 5–3
- Utdanningsdirektoratet (2020a) *Læreplan i fremmedspråk* (FSP01-02). <https://www.udir.no/lk20/fsp01-02>

- Utdanningsdirektoratet (2020b) Kjennetegn på måloppnåelse – fremmedspråk nivå I (FSP01-02).
<https://www.udir.no/laring-og-trivsel/lareplanverket/kjennetegn/kjennetegn-pa-maloppnaelse-fremmedsprak-niva-i/>
- Utdanningsdirektoratet (2020c) Kjennetegn på måloppnåelse – fremmedspråk nivå II (FSP01-02).
<https://www.udir.no/laring-og-trivsel/lareplanverket/kjennetegn/kjennetegn-pa-maloppnaelse-fremmedsprak-niva-ii>
- Vermeer, A. (2004) The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. Dans P. Bogaards et B. Laufer (dir.) *Vocabulary in a Second Language* (p. 173-189). John Benjamins.
- Vinet, A. (2011). *L'enseignement du vocabulaire en cours de français*. [Thèse de Master, L'Université de Stockholm].
- Wang, Y. et Treffers-Daller, J. (2017) Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, 65, 139-150.
<http://dx.doi.org/10.1016/j.system.2016.12.013>
- Wilkins, D. A. (1972) *Linguistics in Language Teaching*. Edwar Arnold.

Annexes

I Relevans for lektorutdanningen

I dette arbeidet har jeg undersøkt hvor stort vokabularet til norske videregående elever med fransk som fremmedspråk er. Innenfor språkforskning har det de siste årene har det vært et økende fokus på vokabular, og flere studier har vist hvordan vokabularforståelse og vokabularstørrelse er viktige faktorer som har et symbiotisk forhold til flere av språkferdighetene. Vokabular kan ikke lengre sees på som en isolert kunnskap, eller en kunnskap som skal komme i andre rekke til fordel for eksempelvis grammatikk. Som nevnt i introduksjonen til oppgaven, så kan lite formidles uten grammatikk, men *ingenting* kan formidles uten vokabular.

Kunnskapen jeg har opparbeidet meg om denne viktige sammenhengende mellom vokabular og generell språkferdighet er noe jeg kommer til å ta med videre i arbeidet som lærer. Både i den form av at vokabularundervisning må ta en enda større del av undervisningen, men også hvordan ulike former for vokabularkunnskap kan undervises og -som konsekvens - testes på ulike måter. Min nåværende kunnskap om de ulike måtene å kunne et ord på - gjenkjenning, definering og til slutt bruk - har gitt meg en større og mer variert måte å tenke om vokabularkunnskap på, noe som også åpner opp for ulike måter å undervise på. Videre kan jeg også nevne at jeg gjennom arbeidet med kategorisering og bearbeiding av større mengder data har lært hvordan man kan på enklere måter kartlegge elevenes vokabularkunnskaper, for å videre kunne bruke disse dataene til å teste og forbedre egen undervisningspraksis.

II Bakgrunnsinformasjon for forskningsprosjekt om vokabularstørrelse

For å hjelpe oss i å samle så nøyaktige resultat som mulig til dette prosjektet ber vi deg svare på noen nødvendige spørsmål om deg og din språkbakgrunn.

Hvilket kjønn er du?

Hvor gammel er du?

- 15
- 16
- 17
- 18
- 19 eller eldre

Hvilket trinn går du på?

- VG1
- VG2
- VG3

Hvilket nivå fransk har du?

- Fransk nivå I
- Fransk nivå II
- Fransk nivå I+II

Da skoleåret begynte høsten 2021, hvor lenge hadde du hatt fransk på skolen?

- Aldri
- 1 år
- 2 år
- 3 år
- 4 år eller mer

Hvilket trinn begynte du å lære fransk på?

- 8.klasse
- 9.klasse
- 10.klasse
- VG1

Hva er morsmålet ditt? Lærte du to eller flere språk fra fødselen av, oppgir du begge/alle språkene

Har du familie/nære venner som du snakker et annet språk enn norsk med?

- Ja
- Nei

Du svarte ja på at du snakker et annet språk enn norsk med familie/nære venner. Vennligst oppgi hvem du snakker andre språk enn norsk med.

Du må velge minst ett svaralternativ.

Nær familie

Storfamilie

Nære venner

Perifere venner / bekjente

Du svarte at du snakker et annet språk enn norsk med familie/nære venner.

Vennligst oppgi hvilke(t) språk dere bruker

Hvilke språk kan du og hvordan vil du vurdere ferdighetene dine i disse språkene? Sett kryss i tabellen. Hvis du kan flere språk enn engelsk og fransk, fyll inn disse selv i neste spørsmål.

	Grunnleggende: Jeg kan kjenne igjen og bruke enkle ord og uttrykk	Middels: Jeg kan bruke vanlige ord og enkle setninger, og kan kommunisere på grunnleggende nivå.	<u>Selvstendig</u> : Jeg kan forstå hovedinnholdet i standard tale og snakke/skrive sammenhengende om kjente tema.	Kompetent: Jeg kan snakke ganske flytende og forstå sammenhengende tale om vanskelige, kjente tema. Jeg kan skrive detaljert om mange tema.	Flytende: Jeg kan uttrykke meg flytende, spontant og presist. Jeg forstår uten problemer alt jeg leser og hører, og kan bruke språket effektivt i sosiale og faglige sammenhenger
Fransk					
Engelsk					

Hvis du kan andre språk enn fransk og engelsk, vennligst fyll inn i boksen under hvilke(t) og hvordan du selv vil vurdere ferdighetene dine i disse språkene.

Grunnleggende:

Jeg kan kjenne igjen og bruke enkle ord og uttrykk

Middels:

Jeg kan bruke vanlige ord og enkle setninger, og kan kommunisere på grunnleggende nivå.

Selvstendig:

Jeg kan forstå hovedinnholdet i standard tale og snakke/skrive sammenhengende om kjente tema.

Kompetent:

Jeg kan snakke ganske flytende og forstå sammenhengende tale om vanskelige, kjente tema. Jeg kan skrive detaljert om mange tema.

Flytende:

Jeg kan uttrykke meg flytende, spontant og presist. Jeg forstår uten problemer alt jeg leser og hører, og kan bruke språket effektivt i sosiale og faglige sammenhenger.

Svareksempel: samisk- kompetent

Hvor mange timer i uka utenom skole tror du at du blir eksponert for fransk?

Dette kan for eksempel være gjennom serier, youtube, filmer, bøker, podcast, gaming, musikk osv.

- Aldri
- 0-1 time
- 2-3 timer
- 4-5 timer
- Mer enn 6 timer

Hvor mange timer i uka utenom skole tror du at du blir eksponert for engelsk?

Dette kan for eksempel være gjennom serier, youtube, filmer, bøker, podcast, gaming osv.

- Aldri
- 0-1 time
- 2-3 timer
- 4-5 timer
- Mer enn 6 timer

Har du dysleksi?

Vi spør kun om dette fordi det kan tenkes å påvirke resultatet av vokabulartesten.

- Ja
- Nei

Tusen takk for din deltakelse!

Husk å trykke på "Send" for å fullføre og sende inn skjemaet

Med vennlig hilsen

Kjersti Faldet Listhaug

(Veileder)

Else Maria Rolsted Lie

(Masterstudent)

III Informasjonsskriv

Vil du delta i forskningsprosjektet Norske VGS elevers vokabularstørrelse i fransk?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke vokabularstørrelsen hos elever i videregående med fransk som tredjespråk.

Formål

Dette er en masteroppgave der målet er å undersøke vokabularstørrelsen (hvor mange ord man kan på et språk) i fransk hos elever i den norske videregående skolen. Et godt ordforråd er nødvendig for å kunne uttrykke seg på et fremmedspråk, og det er derfor nyttig å se nærmere på hvor stor ordforrådene til norske elever faktisk er. Resultatene fra testene vil sammenligne elever med fransk nivå 1 og 2, samt elever på ulike årstrinn. Resultatene kan også sammenlikning av data fra andre studier, både tidligere studier og studier i fremtiden.

Hvorfor får du spørsmål om å delta?

Du har blitt spurt om å delta fordi du er videregåendelev med fransk som fremmedspråk.

Hva innebærer det for deg å delta?

Hvis du velger å delta, innebærer det at du først gjennomfører en vokabulartest. Du får se lister med franske ord, og blir bedt om å krysse av for om du kjenner igjen eller kan bruke disse ordene på fransk. Blant disse er det også blandet inn «falske» ord som er konstruert for å ligne på franske ord, men som ikke finnes på ekte. Disse falske ordene fungerer altså som en form for kontroll. Etterpå får du spørsmål om deg og din språkbakgrunn.

Det tar ca. 15-20 minutter å svare på undersøkelsen.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Du kan når som helst trekke samtykket - rett og slett ved å la være å fullføre undersøkelsen. Alle dine personopplysninger vil da bli slettet. Kun fullstendig anonymiserte data vil bli beholdt. Det vil ikke være mulig å trekke ditt samtykke etter at skjemaet er sendt inn, da det ikke er mulig å spore svarene tilbake til individuelle deltakere når dette er gjort.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

NTNU har ansvar for prosjektet. Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om her. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Datamaterialet vil bli lagret på NTNUs servere. Det er kun prosjektansvarlig (veileder) og en masterstudent som vil ha tilgang til opplysningene som samles inn. Du vil ikke kunne gjenkjennes i publikasjon av masteroppgaven.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Opplysningene er anonyme og oppbevares anonymt ved prosjektslutt.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke. På oppdrag fra NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med prosjektansvarlig Kjersti Faldet Listhaug (kjersti.listhaug@ntnu.no) eller personvernombud ved NTNU, Thomas Helgesen (thomas.helgesen@ntnu.no). Du kan også ta kontakt med NSD – Norsk senter for forskningsdata AS, på epost (personverntjenester@nsd.no) eller telefon: 55 58 21 17.

Studien gjennomføres anonymt. Dersom du skulle ønske å trekke deg underveis i undersøkelsen, kan du gå ut av skjemaet uten å fullføre det. Hvis du trekker deg underveis, vil dataene dine vil ikke bli registrert. Når du fullfører undersøkelsen, vil dataene dine anonymiseres og du vil ikke kunne identifiseres.

Klikk på neste for å begynne på undersøkelsen.

Samtykke-erklæring

Samtykke til:

- Jeg har mottatt og forstått informasjon om prosjektet *Norske VGS elevers vokabularstørrelse i fransk*, og har fått anledning til å stille spørsmål, og
- Jeg har mottatt og forstått informasjon om prosjektet *Norske VGS elevers vokabularstørrelse i fransk*, og har fått anledning til å stille spørsmål.

gis når du trykker "send" på slutten av spørreskjemaet.

IV La mis en page d'une des versions du test X-lex à Nettskjema

Vokabulartest

Under finnes en liste med ord, noen er ekte, franske ord mens andre er oppfunnet og kun ment for å se ut som franske ord.

Vi ber som at du krysser av "ja" hvis du kjenner igjen, forstår eller kan bruke ordene på fransk. Dersom du verken kan kjenne igjen, forstå eller bruke ordet på fransk, krysser du av for "nei".

Vennligst kryss av for alle ordene, selv om du ikke er sikker

Ja, jeg kan kjenne igjen, forstå eller bruke dette ordet på fransk	Nei, jeg kan ikke kjenne igjen, forstå eller bruke dette ordet på fransk
--	--

existence *

genre *

survivre *

