

Étendue du vocabulaire et compréhension écrite – le français à l'université en Suède

Mårten Ramnäs (Université de Gothenburg, Suède)

Abstract: This paper investigates the amount of French vocabulary needed to read (with reasonable comprehension) three authentic novels used in a first semester course of French at a Swedish university. The analysis applies the concept of lexical coverage and uses the frequency lists of Lonsdale & Le Bras (2009). The counting unit is the lemma. This study also examines the common French word list used in all Swedish universities offering French degrees. The results show that 7 000 to 9 000 lemmas will provide the 98 % coverage needed for reasonable comprehension. This is at least twice the 4 000 lemmas contained in the university list. The results therefore point to the necessity of including a vocabulary component also in higher-level French courses.

Keywords: vocabulary, coverage, reading comprehension, SLA, French

1. Introduction

Il existe aujourd'hui quantité d'études qui montrent une forte corrélation entre l'étendue du vocabulaire (le nombre de mots qu'on connaît) d'un apprenant de langue seconde et ses compétences linguistiques générales. Cela vaut en particulier pour les compétences passives, mais également pour les compétences productives (Laufer, 1992, 1998; Milton *et al.* 2010; Stæhr, 2008). En d'autres mots, un apprenant qui a un vocabulaire étendu a le plus souvent aussi une bonne maîtrise générale de la langue qu'il apprend. Malgré les nombreuses études qui mettent en évidence ce lien, le vocabulaire peine à faire sa place dans l'enseignement des langues, et ce depuis plus d'un demi-siècle (Bogaards 1994). Or, de plus en plus de voix se font entendre pour souligner sa grande importance. En Suède, l'initiative d'une liste de vocabulaire commune valable pour le premier semestre d'études universitaire de français, introduite il y a une dizaine d'années et adoptée par toutes les universités suédoises est peut-être révélatrice de cet intérêt.

1.1 But

Cet article reprend et développe un sujet déjà abordé dans une publication récente (Ramnäs 2017). En effet, nous nous proposons d'analyser ici le vocabulaire de trois textes littéraires qui sont au programme du premier semestre d'études de français à l'université de Göteborg pour essayer de déterminer l'étendue du vocabulaire nécessaire pour les lire avec fluidité et sans recours au dictionnaire. Nous examinerons aussi différents aspects de la liste de vocabulaire commune des universités suédoises mentionnée ci-dessus en la comparant avec les résultats de notre analyse: les étudiants, sont-ils capables de lire les romans s'ils

connaissent les mots de la liste? Pour conclure, nous discuterons quelques implications pédagogiques de cette étude pour l'apprentissage et l'enseignement du vocabulaire.

1.2 Cadre théorique et terminologique

Avant d'arriver à la présentation des résultats, il est utile de rappeler la signification de quelques termes qui seront employés dans ce qui suit. Comme la notion de **mot** manque de clarté, nous nous restreindrons ici à deux définitions bien distinctes. Ce terme sera tantôt employé dans le sens de **token** (mot-forme ou occurrence), "un ensemble de morphèmes formant une unité sémantique, et, le plus souvent isolé par deux espaces à l'écrit" (Le Querler 1994), tantôt comme synonyme du terme **lemme** pour désigner les mots tels que l'on les trouve dans un dictionnaire, là où ils regroupent les différentes formes flexionnelles d'un même mot (entrée). Pour les verbes par exemple, le lemme est représenté par l'infinitif (sa forme vedette ou canonique). Dans ce dernier cas, par souci de clarté, nous préférons le plus souvent le terme de lemme, mais comme d'autres chercheurs emploient parfois "mot" dans le sens de lemme, nous n'avons pas pu complètement éviter cet emploi. (En français, on trouve aussi les termes unité lexicale, lexie ou item lexical pour désigner la notion de lemme.)

Les chercheurs anglophones vont souvent plus loin en utilisant le terme de **famille de mots** (*word family*) sous lequel on regroupe les mots qui partagent le même radical (la même racine), l'idée étant qu'avec des connaissances morphologiques élémentaires, on peut deviner les significations des autres membres de la famille de mots si on connaît la signification d'un membre. Pour l'anglais, il existe une définition de famille de mots à 6 niveaux proposée par Bauer et Nation (1993). Pour définir les niveaux ces chercheurs se sont servis de critères tels que la fréquence, la productivité, la transparence et la régularité des morphèmes dérivationnels. En étudiant cette définition de près, on s'aperçoit que le premier niveau correspond au terme **type** ("a different form is a different word", 270) et le deuxième niveau à peu près au terme lemme (regroupant les formes flexionnelles). Dans le corpus qui soutient l'étude de Bauer et Nation, le nombre de types (niveau 1) est sensiblement plus élevé que le nombre de lemmes (le niveau 2). En effet, cette différence est de 39 % alors que la différence en nombre entre lemmes (niveau 2) et familles de mots dans sa définition la plus étendue (niveau 6) est nettement moindre (22 %). Quand on compare une étude où l'unité de mesure est le lemme avec une étude où l'unité de mesure est la famille de mots, il faut tenir compte de la différence qui existe entre les deux, mais cette différence est peut-être moins importante que ce que l'on pourrait penser au premier abord. Pour le français, il n'y a, à notre connaissance, pas de définition semblable proposée, ce qui ajoute à la complexité des comparaisons.

Dans l'introduction, nous avons évoqué le lien qui existe entre l'étendue du vocabulaire d'un apprenant et ses compétences linguistiques générales. **L'étendue** peut se définir comme le nombre de mots pour lesquels un locuteur peut donner au moins un sens correct (Schmitt 2014, 942). Or, la connaissance d'un mot comprend une multitude d'autres aspects que l'on regroupe souvent sous le terme de **profondeur**. Nation (2001, 48) propose de diviser ces différents aspects en trois catégories: la forme, le sens et l'utilisation. Les catégories sont ensuite divisées en sous-catégories pour lesquelles il y a toujours une face productive et une autre face réceptive. Schmitt (2014) constate que la profondeur est un concept utile pour expliquer la nécessité d'apprendre tous les aspects d'un mot (prononciation, orthographe, réseaux sémantiques, collocations, etc.), mais que ce concept est très difficile à cerner et à

mesurer vu sa grande complexité. D'autre part, l'étendue et la profondeur sont en relation puisque la connaissance passive d'un mot peut être considérée comme un aspect de la profondeur. En effet, pour les apprenants moins avancés, l'étendue et la profondeur montre une corrélation forte (Schmitt 2014). En dehors de ces deux concepts, Daller *et al.* (2007) font usage d'une troisième dimension qui est celle de la **fluidité** et qu'ils définissent comme la facilité et la rapidité avec lesquelles le locuteur est capable de reconnaître et d'utiliser les mots. Dans cette étude, nous nous intéresserons uniquement à l'étendue du vocabulaire en raison de la relation particulièrement étroite qui existe entre celle-ci et la compréhension écrite.

En anglais, on entend par le terme *coverage* la proportion de mots d'un texte qui sont couverts par un certain vocabulaire. Nous emploierons en français le terme **couverture**. Comme les mots les plus fréquents de la langue reviennent dans tous les textes, on constate qu'un millier de lemmes suffisent souvent pour assurer une couverture atteignant les 80 %. La couverture nécessaire pour lire et comprendre un texte est une question sur laquelle se sont penchés de nombreux chercheurs. Différents chiffres ont été avancés, mais plusieurs études récentes concordent pour dire qu'au moins 98 % de couverture est nécessaire pour arriver à une lecture fluide et plaisante (Nation 2001, 2006; Schmitt *et al.* 2011). Une couverture de 98 % implique que 2 % des mots sont inconnus. Pour lire et comprendre un texte raisonnablement bien on n'a donc pas besoin de comprendre tous les mots qui le constituent. Les exemples ci-dessous offrent une illustration simple du phénomène de couverture. Le passage est tiré du roman *La Petite Bijou* (Modiano 2001) et les exemples adaptés sont tirés de Ramnäs (2017). Même si on arrive à une couverture de 80 % avec mille lemmes¹, le texte reste incompréhensible puisque les mots couverts sont, à l'exception de quelques mots lexicaux très fréquents, pour la plupart des mots grammaticaux (voir le premier exemple ci-dessous). La couverture augmente vite avec un vocabulaire plus étendu comme nous le voyons dans le deuxième et le troisième exemple ci-dessous, mais au-delà des premières tranches de mille mots, chaque nouvelle tranche de mille mots ne fait augmenter la couverture que de très peu. Dans le troisième exemple, 4 mots sur 85 sont inconnus, ce qui correspond à une couverture d'environ 95 %. Pour arriver à 98 %, deux mots au maximum peuvent être inconnus dans cet extrait.

1 000 lemmes (18 mots inconnus)

Une m'était revenue en, l'une des quelques que j'ai gardées de ma mère. Son visage est comme si un l'avait fait de la nuit. J'ai toujours une devant cette Dans mes, chaque fois, c'était une que quelqu'un me — un de police, un de la — pour que je puisse cette personne. Mais je restais Je ne savais rien d'elle.

2 000 lemmes (7 mots inconnus)

Une *photo* m'était revenue en *mémoire*, l'une des quelques *photos* que j'ai gardées de ma mère. Son visage est comme si un l'avait fait de la nuit. J'ai toujours *éprouvé* une devant cette *photo*. Dans mes *rêves*, chaque fois, c'était une *photo* que quelqu'un me tendait — un *commissaire* de police, un *employé* de la — pour que je puisse *identifier* cette personne. Mais je restais Je ne savais rien d'elle.

¹ Les travaux de Lonsdale et Le Bras (2009) servent de référence (voir aussi sous 1.3).

3 000 lemmes (4 mots inconnus)

Une *photo* m'était revenue en *mémoire*, l'une des quelques *photos* que j'ai gardées de ma mère. Son visage est *éclairé* comme si un l'avait fait *surgir* de la nuit. J'ai toujours *éprouvé* une *gêne* devant cette *photo*. Dans mes rêves, chaque fois, c'était une *photo* que quelqu'un me tendait — un *commissaire* de police, un *employé* de la — pour que je puisse *identifier* cette personne. Mais je restais Je ne savais rien d'elle.

L'extrait sans trous

Une *photo* m'était revenue en *mémoire*, l'une des quelques *photos* que j'ai gardées de ma mère. Son visage est *éclairé* comme si un *projecteur* l'avait fait *surgir* de la nuit. J'ai toujours *éprouvé* une *gêne* devant cette *photo*. Dans mes rêves, chaque fois, c'était une *photo anthropométrique* que quelqu'un me tendait — un *commissaire* de police, un *employé* de la *morgue* — pour que je puisse *identifier* cette personne. Mais je restais *muette*. Je ne savais rien d'elle. (Modiano 2001, 9–10)

Comme nous allons essayer de discuter quelques implications didactiques à partir des résultats de notre analyse, il est utile de faire ici quelques remarques préliminaires. Dans les cours de français à l'université en Suède, il existe depuis longtemps des épreuves qui évaluent les connaissances lexicales des étudiants. Pour réussir aux épreuves, les étudiants sont renvoyés aux cours de littérature. Jusqu'à très récemment, cela a été, de loin, la pratique la plus courante et l'idée qui la sous-tend est qu'on apprend des mots implicitement en lisant. C'est donc une méthode implicite ou indirecte pour apprendre le vocabulaire.

En effet, en langue maternelle, la plupart des chercheurs sont d'accord pour dire que l'enrichissement du vocabulaire se fait avant tout par la lecture et ce dès que l'enfant commence à lire (Cunningham 2005, Sternberg 1987). La rencontre fortuite et répétée de mots dans des textes est sans aucun doute une source importante pour le développement du vocabulaire en langue maternelle. Or, appliquée à l'apprentissage et à l'enseignement d'une langue seconde comme l'a proposé par exemple Krashen (1983), cette approche se heurte à plusieurs problèmes. Comme la proportion de mots inconnus que l'on apprend de cette manière est très faible, il faut lire énormément de textes (Cobb & Horst 2004). C'est ce qui se passe généralement pour la langue maternelle (même si les différences individuelles concernant la quantité de lecture peuvent être très importantes), mais pour la langue seconde, c'est très rarement le cas. D'autre part, pour lire tant de textes, il faut que les textes ne soient pas trop difficiles par rapport au niveau de l'apprenant, ce qui complique l'apprentissage quand il s'agit d'une langue seconde (Tréville & Duquette 1996, 69). Aux méthodes implicites ou indirectes, on peut donc opposer les méthodes explicites (ou directes), qui impliquent que l'apprentissage du vocabulaire nécessite « d'élaborer un programme qui vise à développer systématiquement la compétence lexicale » (Tréville & Duquette 1996, 95).

1.3 Matériaux et méthode

Une fois que nous avons déterminé la couverture qui correspond au niveau requis pour arriver à une lecture aisée, on peut se demander combien de lemmes sont nécessaires pour y atteindre. Pour l'anglais, Nation (2006, 71) arrive à la conclusion qu'il faut connaître de 8 000 à 9 000 familles de mots pour lire des romans (98 % de couverture). Pour le français, il existe très peu d'études à ce sujet. C'est pourquoi nous nous sommes proposé d'analyser les romans qui font partie des lectures du premier semestre d'études de français. À l'université de Göteborg, quatre romans sont au programme dont trois obligatoires et un au choix de

l'étudiant. Nous en avons retenu les trois romans étudiés par l'ensemble des étudiants. Ces romans sont *La Petite Bijou* (Modiano 2001), *L'étranger* (Camus 1995 [1942]) et *No et moi* (de Vigan 2007). Le tableau 1 présente quelques données relatives au corpus constitué des trois livres.

	Nombre de mots	Nombre de lemmes
La Petite Bijou	35 388	2 293
L'étranger	34 380	2 643
No et moi	55 977	3 745

Tableau 1 : Le corpus des textes littéraires analysés.

Quel vocabulaire est en effet nécessaire pour lire ces livres choisis pour leur relative simplicité ? Pour répondre à cette question, les livres ont été scannés et soumis à un programme de reconnaissance de caractères (*Nuance OmniPage 18*). Les textes électroniques ainsi obtenus ont ensuite été lemmatisés grâce à *TreeTagger* (Schmid 2016). Ce lemmatiseur produit une liste qui comprend tous les lemmes et leurs fréquences dans le texte analysé. Le taux de précision de *TreeTagger* est élevé, mais le résultat de l'analyse contient pourtant un certain nombre d'erreurs. Pour y remédier, nous avons entrepris un examen manuel de la liste, ce qui nous a amené à reclasser certains mots. Cette liste a ensuite été comparée à la liste de vocabulaire qui est à la base de *A Frequency Dictionary of French* (Lonsdale & Le Bras 2009). Ce dictionnaire rassemble les 5 000 mots (lemmes) les plus fréquents d'un très vaste corpus équilibré constitué de textes de sources différentes, écrites et parlées. Les lemmes moins fréquents (de 6 000 à 25 000), non-inclus dans le dictionnaire imprimé, sont accessibles sur le site *Compleat Lexical Tutor* (Cobb 2017). Dans le tableau 2, nous trouvons des exemples de lemmes tirés du roman *L'étranger* pour différentes tranches de 1 000 lemmes (1 000–15 000) ou par tranches de 5 000 lemmes (16 000–25 000). Le "Off-list" rassemble les lemmes qui ne sont pas recensés par les matériaux qui soutiennent le dictionnaire cité.

Tranches de mille lemmes	Exemples de lemmes
1 000	aller, comme, encore, famille, propre
2 000	arbre, cinéma, dénoncer, lendemain, prêter
3 000	assassiner, convoquer, éternel, gardien, sommeil
4 000	aile, heurter, intensité, piquer, sourd
5 000	barreau, cordon, humble, marée, triompher
6 000	ardent, canne, cracher, paupière, sentence
7 000	ascension, étoffe, faubourg, funèbre, valide
8 000	acharnement, éparpiller, fracas, gravir, tempe

9 000	brûlure, cloison, croûte, empiler, puéril, vaciller
10 000	bagne, bordure, chalutier, furtif, tonner
11 000	bouée, dévaler, fautif, informe, marmonner
12 000	bourdonner, mauresque, niais, pétale, raie
13 000	blafard, péremptoire, réfectoire, rejaillir, taciturne
14 000	devanture, fougueux, imperturbable, tirade, véranda
15 000	charogne, crosse, pourvoi, prétoire, songeur
20 000	brassard, cabanon, doucereux, punaise, taillader
25 000	encrier, insolation, taquet, traversin, treuil
Off-list	asphodèle, connaissance, dodeliner, vrombir

Tableau 2 : Exemples de lemmes par tranche dans *L'étranger*.

1.4 La liste de vocabulaire commune des universités suédoises

La liste commune a été élaborée en deux étapes. La première version a vu le jour en 2008 et la seconde quelques années plus tard. La première version de la liste a été élaborée par Lindqvist *et al.* (2013) à des buts de recherche. Elle comptait environ 2 700 lemmes et incluait les mots apparaissant plus de 10 fois dans un corpus oral (le CorpAix). Cette première version a ensuite été révisée par Per Förmegård de l'université de Stockholm à des buts d'enseignement, qui a fait des ajouts, mais aussi des exclusions. Aujourd'hui la liste compte non loin de 4 000 mots auxquels s'ajoutent aussi un certain nombre d'expressions où sont employés les mots. Les connaissances lexicales par rapport à la liste sont évaluées au moyen du "Riksprov" (une épreuve nationale). Environ 200 étudiants passent cette épreuve chaque semestre dans les 8 universités où sont donnés des cours de français en Suède.

2. Résultats et analyses

Le tableau 3 ci-dessous résume le résultat des analyses. On y voit la couverture exprimée en pourcentage et le nombre de lemmes couverts dans les textes pour une étendue de vocabulaire donnée par tranches de 1 000 lemmes (ou 5 000 lemmes). Par exemple, un vocabulaire consistant des deux mille lemmes les plus fréquents couvre 91,5 % des mots de *L'étranger* et au total 1 162 de ces deux mille lemmes y sont présents. Pour la tranche spécifique (la deuxième tranche de mille lemmes) 451 lemmes sont représentés. Surtout, il ressort du tableau qu'il faut une étendue du vocabulaire de 7 000 lemmes (*La Petite Bijou*) à 9 000 lemmes (*No et moi*) pour lire ces trois romans jugés suffisamment faciles pour le premier niveau de français à l'université en Suède². Comme nous l'avons vu plus haut, Nation (2006, 71) avance le chiffre de 8 000 à 9 000 familles de mots pour lire des romans, mais il convient de rappeler qu'il s'agit là d'une autre unité de mesure (famille de mots) et d'une autre

² Les noms propres sont envisagés comme des mots (tokens) connus, mais ne font pas partie du répertoire des lemmes.

langue (l'anglais). D'autre part, l'étude de Nation s'intéresse aux romans en général alors que la nôtre ne porte que sur trois livres utilisés dans un cours de français. Or, compte tenu de ces différences, les résultats de notre analyse sont tout de même d'un ordre de grandeur cohérent avec ceux de Nation et indiquent qu'il faut un vocabulaire relativement important pour lire avec fluidité des romans authentiques.

Vocabulaire par tranches	L'étranger (Camus)			La Petite Bijou (Modiano)			No et moi (de Vigan)		
	Couverture	Nombre de lemmes dans le texte	Nombre de lemmes cumulés	Couverture	Nombre de lemmes dans le texte	Nombre de lemmes cumulés	Couverture	Nombre de lemmes dans le texte	Nombre de lemmes cumulés
1 000	86,0%	711	711	85,2%	665	665	84,7%	810	810
2 000	91,5%	451	1162	90,9%	401	1066	90,1%	585	1395
3 000	94,3%	331	1493	94,2%	286	1352	93,2%	429	1824
4 000	95,9%	230	1723	95,9%	200	1552	95,1%	335	2159
5 000	96,7%	163	1886	96,9%	136	1688	96,1%	241	2400
6 000	97,4%	131	2017	97,6%	107	1795	97,0%	206	2606
7 000	97,9%	100	2117	98,0%	73	1868	97,4%	149	2755
8 000	98,2%	64	2181	98,3%	57	1925	97,8%	127	2882
9 000	98,5%	66	2247	98,5%	44	1969	98,1%	122	3004
10 000	98,7%	56	2303	98,8%	50	2019	98,4%	89	3093
11 000	98,9%	41	2344	99,0%	46	2065	98,7%	95	3188
12 000	99,0%	41	2385	99,2%	38	2103	98,8%	67	3255
13 000	99,2%	39	2424	99,2%	15	2118	99,0%	62	3317
14 000	99,3%	25	2449	99,4%	27	2145	99,1%	61	3378
15 000	99,4%	28	2477	99,4%	16	2161	99,3%	40	3418
20 000	99,7%	81	2558	99,7%	65	2226	99,6%	157	3575
25 000	99,8%	30	2588	99,8%	28	2254	99,8%	63	3638
Off-list	100,0%	47	2635	100,0%	39	2293	100,0%	107	3745

Tableau 3 : Couverture et nombre de lemmes par tranche dans les trois textes.

Nous pouvons aussi noter que, quel que soit le texte, les mots les plus fréquents sont employés à un haut degré, d'où l'utilité de définir des listes de vocabulaire où un critère

important est la fréquence. Pour les lemmes de la première tranche, plus de deux lemmes sur trois sont utilisés dans chacun des textes et pour la deuxième tranche de mille lemmes environ la moitié en sont utilisés. Par contre, pour la 15^e tranche de mille lemmes seuls quelques dizaines apparaissent dans les textes. Exiger que les étudiants maîtrisent ces mots en premier serait donc une mauvaise idée. Pour les mots peu fréquents, le nombre de lemmes est seulement légèrement supérieur à celui du nombre de tokens, ce qui s'explique par le fait que la plupart des mots peu fréquents n'apparaissent qu'une fois dans les textes. Si on ne comprend pas un mot peu fréquent, la couverture (la compréhension) n'en est donc pas affectée de la même façon que s'il s'agissait d'un mot appartenant aux premières tranches de mille lemmes. Cela explique aussi le fait que le nombre absolu de lemmes inconnus peut paraître important sans trop nuire à la compréhension du texte. Par exemple, avec une couverture de 98 % de *L'étranger*, les 2 % de mots inconnus représentent environ 600 mots (tokens) et 450 lemmes. Il peut malgré tout y avoir des lemmes de basse fréquence (peu fréquents dans la langue en général), mais qui reviennent souvent dans un texte. Ces lemmes, susceptibles d'être importants pour l'histoire, sont faciles à identifier avec des outils tels que ceux que nous avons décrits plus haut. Ainsi, le tableau 4 donne les lemmes moins fréquents que les 8 000 premiers lemmes, mais qui reviennent au moins quatre fois dans *L'étranger*. L'identification de ce type de mots peut avoir un intérêt pédagogique.

Lemme	Tranche de mille mots	Fréquence
pourvoi	15 000	9
aumônier	16 000	7
charogne	15 000	6
cabanon	18 000	5
huissier	10 000	5
Pensionnaire	11 000	5
plaidoirie	9 000	5
tromperie	11 000	5
Bouée	11 000	4
Brun	11 000	4
éventer	21 000	4
prétoire	15 000	4
verrière	15 000	4

Tableau 4 : Lemmes de basse fréquence qui reviennent plus de quatre fois dans *L'étranger*.

Si on compare la liste de vocabulaire commune des universités suédoises avec nos résultats, on constate que les 4 000 lemmes qu'elle contient sont loin de suffire pour garantir une lecture fluide, même des textes sélectionnés pour leur relative simplicité. En effet, il faut

connaître au moins le double de lemmes (7 000 à 9 000). Il convient aussi de noter que les 4 000 lemmes de la liste commune n'offrent pas une couverture aussi importante que les 4 000 premiers lemmes du dictionnaire de Lonsdale & Le Bras, ce qui s'explique par le mode de sélection des mots. Comme nous l'avons vu sous 1.4, la sélection s'est effectuée en deux étapes distinctes. Il y a d'abord eu une première version de la liste contenant les 2 700 lemmes les plus fréquents d'un corpus oral. Le critère de sélection de cette première version était donc la fréquence, mais la liste était basée uniquement sur des matériaux oraux. L'expansion de la liste (la deuxième version), qui a porté le nombre de lemmes de 2 700 à 4 000, est une bonne chose, mais la sélection des nouveaux mots (et l'exclusion de d'autres mots) est le fruit d'une démarche intuitive (et non scientifique). C'est pourquoi on trouve dans la version actuelle de la liste plusieurs centaines de mots ajoutés, qui sont peu fréquents dans la langue en général (si on prend les travaux de Lonsdale & Le Bras comme référence), comme par exemple *bêche*, *courgette*, *jonquille*, *joufflu*, *osier* et *ouate*. Sous 2.0 nous avons vu que la fréquence est un critère important si on veut définir une liste de mots à apprendre pour les étudiants. Cela est dû au fait qu'un apprenant a tendance à apprendre les mots les plus fréquents avant les mots moins fréquents (Cobb & Horst 2004, 17). En effet, les mots fréquents reviennent dans tous les contextes, comme nous l'avons déjà constaté (voir sous 2.0). Cependant, si on fait de la fréquence le seul critère de sélection on risque de passer à côté de mots qui peuvent être importants pour différentes raisons, mais qui ne s'avèrent pas fréquents dans la langue en général. C'est pourquoi on emploie souvent aussi d'autres critères comme par exemple la portée, la répartition ou la disponibilité (Bogaards 1994, 113–129). Ainsi, avec des critères scientifiques, il serait possible d'évaluer la pertinence d'inclure des mots comme ceux cités plus haut.

En Suède, le français est introduit à l'école comme langue optionnelle au moins trois ans après l'anglais qui est obligatoire pour tous les élèves. Sur le plan lexical, l'anglais et le français montrent une forte proximité; la présence de mots d'origine française dans le lexique anglais est frappante. Ce "cognate advantage" dont parlent certains chercheurs (Bardel *et al.* 2012) est également un élément à prendre en compte dans la sélection des mots, mais aussi susceptible d'être mieux mis à profit dans l'enseignement du français en Suède de façon générale. En effet, nous pensons que toutes les considérations exposées ci-dessus devraient être prises en compte dans une éventuelle révision de la liste commune de vocabulaire. Finalement, il nous faut tout de même ajouter que les 4 000 lemmes de la liste procurent une couverture qui monte à presque 95 %, ce qui est une bonne base et fait baisser sensiblement le nombre de mots inconnus lors de la lecture d'un nouveau texte.

2.1 Implications pédagogiques

Le développement du vocabulaire est primordial pour la réussite des études de langue, mais comment offrir aux étudiants les meilleurs moyens de parvenir à l'apprendre? C'est une question vaste et complexe à laquelle nous nous contentons d'apporter quelques bribes de réponse suites aux résultats de notre analyse. Comme nous l'avons fait remarquer dans ce qui précède, une pratique très répandue consiste à intégrer le vocabulaire dans les cours de littérature en pensant que la lecture constitue la meilleure manière de s'appropriier le vocabulaire. Si on compare le vocabulaire des trois textes analysés avec les 4 000 lemmes de la liste, on constate qu'environ un quart ne sont représentés aucune fois dans les textes et beaucoup n'apparaissent qu'une ou deux fois. À cela, il faut ajouter que les textes, déjà au départ, sont trop difficiles pour permettre aux étudiants de ce niveau un apprentissage

implicite. Cela exclut ainsi totalement l'idée que la lecture de ces romans fournit à elle seule la possibilité d'apprendre tous les mots de la liste implicitement. Dans le cadre de cours comme celui de l'université de Göteborg, la lecture doit être complétée par un enseignement explicite (direct) et structuré intégrant des méthodes qui sont reconnues pour leur efficacité (Laufer 2017; Tréville & Duquette 1996).

3. Conclusion

Pour bien maîtriser une langue étrangère le vocabulaire est un facteur clé. Dans cette étude, qui focalise un domaine spécifique de la compétence linguistique, à savoir la compréhension écrite, nous avons montré que le nombre de lemmes nécessaire pour arriver à lire avec fluidité et sans recours au dictionnaire trois textes littéraires qui font partie des lectures obligatoires d'un cours de français langue étrangère à l'université en Suède est de 7 000 – 9 000. C'est au moins le double du nombre de lemmes présents dans la liste de vocabulaire qui accompagne ce même cours. La lecture fluide de la littérature contemporaine est un objectif pertinent pour les étudiants qui poursuivent leurs études après le premier semestre. C'est pourquoi il serait nécessaire d'intégrer l'enseignement du vocabulaire aussi dans les cours aux niveaux supérieurs. La lecture de textes littéraires présente de nombreux avantages pédagogiques dans les cours de langues. En ce qui concerne le vocabulaire, la lecture permet notamment l'exposition et la découverte de mots en contexte. Par contre, la lecture de quelques romans ne suffit pas pour que l'apprentissage implicite de nouveaux mots ait lieu dans une mesure significative, ce qui rappelle la nécessité d'un enseignement explicite du vocabulaire. L'établissement d'une liste de vocabulaire commune est positif et répond à un besoin pédagogique en ce qu'elle définit une base et souligne l'importance d'apprendre des mots. La liste commune favorise aussi l'échange entre les universités et la réflexion sur les pratiques pédagogiques. De manière générale, le vocabulaire apparaît dans ce qui vient d'être exposé comme un domaine où le lien entre recherche et enseignement mériterait d'être développé.

Bibliographie

Corpus

Camus, Albert. [1942] 1995. *L'étranger*. Paris. Gallimard.

Modiano, Patrick. 2001. *La Petite Bijou*. Paris. Gallimard.

Vigan (de), Delphine. 2007. *No et moi*. Paris. J. C. Lattès.

Ouvrages consultés

Bardel, Camilla, Anna Gudmundson & Christina Lindqvist 2012. « Aspects of lexical sophistication in advanced learners' oral production. Vocabulary acquisition and use in L2 French and Italian ». *Studies in Second Language Acquisition* 34, p. 269-290.

Bauer, Laurie & I. S. Paul Nation. 1993. « Word families ». *International Journal of Lexicography* no. 6 (4): 253–279.

- Bogaards, Paul. 1994. *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris : Didier.
- Cobb, Tom & Marliese Horst 2004. « Is there room for an AWL in French? », *Vocabulary in a second language* edited by Paul Bogaards & Batia Laufer, 15–38. Amsterdam: John Benjamins.
- Cobb, Tom. *Compleat Lexical Tutor*. (2017). Accessed June 11, 2017. <http://www.lextutor.ca/>.
- Cunningham, Anne. E. 2005. « Vocabulary growth through independent reading and reading aloud to children. » *Teaching and learning vocabulary: Bringing research to practice*, edited by Elfrieda H. Hiebert & Michael. L, 45–68. Kamil. Mahwah, New Jersey: Erlbaum.
- Krashen, Stephen. 1989. We Acquire Vocabulary and Spelling by reading: Further Evidence for the Input Hypothesis. *The Modern Language Journal*, 73, 440-464.
- Laufer, Batia. 1992. « How much lexis is necessary for reading comprehension? » *Vocabulary and applied linguistics*, edited by Henri Béjoint & Pierre J.L. Arnaud, 126–132. London: Macmillan.
- Laufer, Batia. 1998. « The development of passive and active vocabulary in a second language: Same or different? » *Applied Linguistics*, no. 19, 255–271.
- Laufer, B. 2017. The Three ‘I’s of second language vocabulary learning: Input, instruction, involvement., *Handbook of research in second language teaching and learning, Volume III* edited by Eli Hinkel, 343–354. London : Routledge.
- Lindqvist, Christina, Anna Gudmundson & Camilla Bardel. 2013. « A new approach to the measuring of vocabulary in L2 oral production ». *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis*, edited by Camilla Bardel, Christina Lindqvist & Batia Laufer, Eurosla Monograph Series 2, 109–126. <http://eurosla.org/monographs/EM02/EM02home.php>
- Lonsdale, Deryle. & Yvon Le Bras. 2009. *A Frequency Dictionary of French*. New York: Routledge.
- Meara, Paul. 1980. « Vocabulary acquisition: A neglected area of language learning ». *Language Teaching and Linguistics: Abstracts*, no. 4, 221–246.
- Milton, James. 2006. « Language lite? Learning French Vocabulary in School ». *Journal of French Language Studies*, no.16 (2), 187–205.
- Milton, James. 2009. *Measuring Second Language Vocabulary Acquisition*. Clevedon : Multilingual Matters.
- Milton, James, Jo Wade & Nicola Hopkins. 2010. « Aural word recognition and oral competence in a foreign language ». *Further insights into non-native vocabulary teaching and learning*, edited by Rubén Chacón-Beltrán, Christian Abello-Contesse & María del Mar Torreblanca-López, 83–98. Bristol : Multilingual Matters.
- Nation, I.S. Paul. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I.S. Paul. 2006. « How Large a Vocabulary is Needed For Reading and Listening? ». *Canadian Modern Language Review* no. 1, 59–82.
- Ramnäs, Mårten. 2017. Vokabulären i franskundervisningen på svenska universitet. *Lingua*, no. 3, 26–30.

- Schmid, Helmut. 2016. *TreeTagger*, Downloaded May 25, 2016. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- Schmitt, Norbert. Xiangying Jiang & William Grabe. 2011. « The percentage of words known in a text and reading comprehension ». *The Modern Language Journal*, no. 95, 26-43.
- Stæhr, Lars S. 2008. « Vocabulary size and the skills of listening, reading and writing ». *The Language Learning Journal*, no. 36 (2), 139–152.
- Sternberg, Robert J. 1987. « Most Vocabulary is Learned From Context ». *The Nature of Vocabulary Acquisition*, edited by McKeown, Margaret & Mary E. Curtis., 89–105. Hillsdale, New Jersey: Erlbaum.
- Tréville. M.-C. & Duquette L. 1996. *Enseigner le vocabulaire en classe de langue*. Paris: Hachette.