



GERFLINT

ISSN 1901-3809

ISSN en ligne 2261-2807

La compétence lexicale en français langue étrangère – quel est l'impact des mots cognats ?

Christina Lindqvist

Université de Göteborg, Suède
christina.lindqvist@sprak.gu.se

Mårten Ramnäs

Université de Göteborg, Suède
marten.ramnäs@sprak.gu.se

Résumé

Le but de la présente étude est d'examiner la taille du vocabulaire d'un groupe d'étudiants universitaires suédois, et d'analyser le rôle de mots cognats. Trente étudiants suédois, tous en première année d'études de français à l'université en Suède, ont participé à l'étude. Ils avaient tous appris l'anglais avant d'apprendre le français comme troisième langue. Le *Test de la Taille du Vocabulaire* (Batista, Horst, 2016) a été utilisé afin de mesurer la taille du vocabulaire. Les résultats font ressortir que les étudiants connaissent environ 7 200 mots en moyenne. De plus, les résultats montrent que le taux de réponses correctes est significativement plus élevé pour les mots cognats que pour les mots non cognats, ce qui indique un effet facilitateur pour les mots cognats, surtout pour les mots les moins fréquents.

Mots-clés : taille du vocabulaire, mots cognats, français, suédois, anglais

Lexical competence in French as a foreign language - what is the impact of cognates?

Abstract

This study aims to provide an overview of Swedish university students' receptive vocabulary size in French and to discuss the role of cognates in the estimated vocabulary sizes. Thirty Swedish students, who were all enrolled in their first year of French studies at a Swedish university, participated in the study. They had all learned English before they started to learn French as their third language. The *Test de la Taille du Vocabulaire* (Batista et Horst, 2016), was used in order to measure their vocabulary size. The results show that the students know around 7 200 words on average. Furthermore, the results show that the proportion of correct answers is significantly higher for the cognate words, suggesting a facilitative cognate effect, especially among the low-frequent words.

Keywords: vocabulary size, cognates, French, Swedish, English

Introduction

Cet article traite un aspect crucial des connaissances lexicales de l'apprenant de langue étrangère, à savoir la taille du vocabulaire. Cet aspect est habituellement défini, de manière relativement vague, comme le nombre de mots connus par un individu. Pour ce qui est de la taille du vocabulaire de l'apprenant d'anglais, un grand nombre d'études révèle un lien fort entre cet aspect et d'autres compétences langagières, telles la compréhension écrite et la production orale et écrite (Milton, 2009), voilà pourquoi il nous semble important de mener une étude similaire sur les apprenants de français langue étrangère, d'autant plus qu'il y a peu d'études antérieures à ce sujet, en comparaison avec les études portant sur l'anglais (voir pourtant Batista et Horst, 2016 ; David, 2008 ; Lindqvist, 2018 ; Milton, 2006, 2008). L'étude présente vise à contribuer aux recherches précédentes en se concentrant sur la taille du vocabulaire d'apprenants suédophones de français troisième langue (L3). L'étude a été effectuée dans une université en Suède avec des étudiants en première année d'études de la langue française. Ils ont tous des connaissances en anglais comme deuxième langue (L2), d'où la dénomination de L3 pour le français. D'ailleurs, comme nous le verrons par la suite, cette distinction se révélera primordiale pour l'interprétation des résultats de l'étude, dans le sens où les connaissances de l'anglais L2 jouent un rôle décisif. Le rôle joué par des L2 déjà apprises a rarement été abordé dans ce genre d'études, ce qui fait ressortir le survol récent de Peters (2020), voir pourtant Lindqvist (2020). Nous discuterons notamment le statut des mots cognats, à savoir des mots avec des formes aussi bien que des significations similaires dans deux ou plusieurs langues, par rapport aux estimations de la taille du vocabulaire. L'objectif principal est d'avoir un aperçu de l'état des lieux concernant la taille du vocabulaire réceptif en français d'apprenants universitaires suédophones. De plus, nous cherchons à mieux comprendre l'effet des mots cognats sur les estimations des tailles du vocabulaire.

1. La taille du vocabulaire en français langue étrangère et le rôle de mots cognats

Les recherches précédentes sur la taille du vocabulaire en français langue étrangère sont peu nombreuses, et ont été effectuées dans d'autres pays, ce qui implique d'autres langues maternelles, d'autres méthodes pédagogiques, d'autres systèmes scolaires, etc. De ce fait, il est parfois difficile de les comparer avec la nôtre. David (2008) a étudié le développement de la taille du vocabulaire du français à travers les années scolaires jusqu'à la dernière année d'études à l'université en Angleterre. Les étudiants universitaires de première année connaissent en moyenne 2 524 lemmes selon les estimations de David. Mentionnons aussi que le groupe le plus avancé dans l'étude de David (2008), les étudiants de quatrième année, ont un vocabulaire de 3 359 lemmes en moyenne.

Les études de Milton concernent elles aussi les apprenants anglophones ; les données ayant été collectées au Pays de Galles. Selon Milton (2006), les apprenants ayant étudié le français pendant sept ans, âgés de 18 ans, ont un vocabulaire de 1 930 lemmes en moyenne. Milton (2008) a examiné la taille du vocabulaire d'étudiants universitaires après quatre ans d'études de français à l'université. En fin d'études, ils connaissent 3 326 lemmes en moyenne. Milton (2009) a aussi examiné la taille du vocabulaire par rapport au CECRL (Cadre européen commun de référence pour les langues, Conseil d'Europe, 2001). Nous ne ferons pas de telle analyse dans notre étude, mais nous pouvons constater que selon le programme du cours, les étudiants sont censés atteindre le niveau B1/B2 vers la fin de la première année d'études. Pour le niveau B1, Milton (2009, p. 187) présente les chiffres suivants : 850 lemmes pour les apprenants anglais, 2 422 pour les apprenants grecs et 2124 pour les apprenants espagnols. Pour le niveau B2 : 1 920 pour les apprenants anglais, 2 630 pour les apprenants grecs et 2 450 pour les apprenants espagnols. Notons aussi qu'au niveau le plus élevé du CECRL, C2, les apprenants dans l'étude de Milton (2009) atteignent entre 3 300 et 3 721 lemmes en moyenne.

La seule étude publiée portant sur la taille du vocabulaire d'apprenants suédois de français est celle de Lindqvist (2018)¹. L'étude concerne 156 apprenants à l'école, dès la première année d'études jusqu'à la quatrième année, et montre un développement régulier de la taille du vocabulaire. Vers la fin de la quatrième année d'études à l'école, à l'âge de 15 ans, les apprenants connaissent 1 150 lemmes en moyenne. Il est à noter que cette moyenne dépasse celle des apprenants anglais de niveau B1 dans l'étude de Milton (2009). Nous pouvons ainsi constater qu'il existe une variation importante par rapport aux estimations de la taille du vocabulaire réceptif des apprenants de français dans les études précédentes.

Batista et Horst (2016) examinent la taille du vocabulaire chez des apprenants immigrants au Canada, divisés en quatre groupes sur la base d'un test portant sur la compréhension et la production orale et écrite. Les résultats font ressortir les tailles du vocabulaire suivantes : *beginners*: 2 699 lemmes en moyenne, *low-intermediate* 4 068, *high-intermediate* 5 274, *advanced* 6 871, donc des estimations nettement plus élevées que celles des autres études abordées ci-dessus. Ceci dit, il faut admettre que les participants dans l'étude de Batista et Horst (2016) se trouvent tous en situation d'acquisition d'une langue seconde en milieu naturel, alors que les participants dans les autres études étudient le français dans des pays où la langue n'est normalement pas utilisée dans la société en général.

L'effet des mots cognats a souvent été discuté dans le cadre de la perspective de la fréquence, selon laquelle les mots les plus fréquents de la langue cible seraient acquis avant les mots de basse fréquence. Néanmoins, pour ce qui est de

l'acquisition du français, les recherches précédentes ont montré que les apprenants moins avancés, qui sont normalement censés ne pas connaître une proportion importante de mots de basse fréquence, utilisent tout de même ce genre de mots en production orale, dû au fait, semble-t-il, que certains de ces mots ont des mots cognats dans d'autres langues connues par les apprenants (Bardel et Lindqvist, 2011, Bardel, Gudmundson et Lindqvist, 2012). Ces résultats sont cruciaux pour l'étude présente puisque les tests de la taille du vocabulaire sont habituellement basés sur des listes de fréquence de la langue cible. En effet, c'est aussi le cas pour le test que nous utilisons, le TTV (Batista et Horst, 2016). Pour ce qui est de l'effet des mots cognats sur le vocabulaire réceptif, il est intéressant de noter que Batista et Horst (2016, p. 228) affirment que les apprenants de français ayant une L1 romane surpassent de loin les apprenants avec d'autres langues maternelles. Elles démontrent des différences statistiquement significatives entre le groupe « L1 romane » par rapport aux groupes « L1 asiatique » et « L1 autre ». Par contre, il n'existe pas de différence entre les deux derniers groupes. Les chercheurs suggèrent ainsi que les apprenants ayant une langue romane pour L1 ont un avantage dans ce genre de test, grâce aux similarités entre ces langues au niveau lexical. Milton (2009, p. 36) admet, lui aussi, que les mots cognats peuvent jouer un rôle dans les mesures de connaissances lexicales, en disant que ces mots peuvent être reconnus plus facilement par les apprenants, surtout à l'écrit. Néanmoins, selon lui, le facteur le plus puissant guidant l'acquisition du vocabulaire serait toujours la fréquence. Les apprenants dans l'étude de Cobb (2000) ont eu de meilleurs scores sur les mots cognats français/anglais que sur les autres mots dans le *Vocabulary Levels Test*. Lindqvist (2020), pour sa part, ré-examine les résultats de l'étude de Lindqvist (2018) en focalisant sur les mots cognats. En effet, comme nous l'avons vu plus haut, les apprenants dans cette étude ont des résultats qui surpassent les groupes correspondants dans les études de David (2008) et de Milton (2006). Selon Lindqvist (2018), une des explications pourrait être le fait que les apprenants bénéficient de leur faculté de reconnaître les mots cognats. Cette hypothèse est en effet confirmée dans Lindqvist (2020), les résultats de l'étude indiquant que la majorité des mots connus par les apprenants ont des mots cognats dans l'anglais et le suédois. On pourrait alors se demander dans quelle mesure les mots « connus » représentent au fond le vocabulaire français des apprenants, ceux-ci ayant éventuellement reconnu certains mots de l'anglais ou du suédois et pas du français en premier lieu. Notons, enfin, que très peu d'études sur la taille du vocabulaire s'intéressent au rôle joué par d'autres langues que la L1. Cette étude se propose de remplir cette lacune dans les recherches antérieures.

Il semblerait alors que les apprenants peuvent bénéficier de la cognation entre deux ou plusieurs langues. Il est cependant à noter qu'il existe aussi des cognats (*partially*) *deceptive* (Granger, 1993), ayant des formes similaires et des sens divergents.

2. Questions de recherche

Nous posons les questions de recherche suivantes :

- 1) Quelle est la taille du vocabulaire d'apprenants universitaires suédophones de français ?
- 2) Quel est l'effet de la connaissance de mots cognats sur les estimations de la taille du vocabulaire ?
- 3) Y a-t-il une différence d'effet concernant les mots cognats L1-L2-L3 par rapport aux mots cognats L2-L3 ?

3. Méthodologie

3.1. Instrument de mesure et méthode d'analyse

Nous avons utilisé le TTV (*Test de la Taille du Vocabulaire*), élaboré en plusieurs étapes par Batista et Horst (2016). Ce test, qui comprend 120 mots au total, estime par extrapolation la taille du vocabulaire (de 1 à 10 000 mots) de la personne évaluée. Le test est divisé en quatre parties représentant quatre tranches de mille mots différentes (2K, 3K, 5K et 10K). Il y a donc 30 mots par tranche. Dans le test, il faut relier un mot avec la définition qui correspond. Les définitions ont été rédigées avec des mots de fréquence égale ou plus élevée que les mots testés. Ce principe a aussi guidé le choix des mots distracteurs. Le résultat obtenu sur les mots de la tranche de 3K est censé refléter les connaissances des étudiants sur cette tranche de fréquence précise (les 2 001 à 3 000 mots les plus fréquents), alors que les résultats obtenus sur les tranches de 2K, 5K et 10K sont censés représenter plusieurs tranches, 1-2K, 4-5K et 6-10K respectivement. Par exemple, un étudiant qui obtient un résultat de 15 sur 30 sur la tranche de 10K est donc supposé connaître 50 % des mots dans les tranches 6-10K, c'est-à-dire 2 500 mots. Pour arriver à estimer la taille globale du vocabulaire, les résultats des différentes tranches sont additionnés.

Une partie de notre analyse sera consacrée aux mots cognats dans ce test. Afin d'identifier ces mots, nous avons suivi la définition selon laquelle les mots cognats sont des mots avec des formes et des significations similaires dans deux ou plusieurs langues. Il n'y a donc pas forcément de lien étymologique entre les

mots, même si c'est le plus souvent le cas. Pour savoir s'il y a un effet cognat dans nos données, nous avons divisé les 120 mots du test en trois groupes : ceux ayant un cognat en anglais et en suédois (44 mots), ceux ayant un cognat en anglais (34 mots) et ceux n'ayant pas de cognats, ni en anglais, ni en suédois (39 mots). Il reste 3 mots qui ont un cognat uniquement en suédois, mais nous ne les considérons pas ici. Il est tout à fait intéressant de noter que sur les 120 mots du test environ 40 % ont un mot cognat en suédois et 65 % en anglais. Rappelons aussi que tous les mots du test ont été choisis au hasard, sans prise en compte d'une cognation éventuelle, et que le test est censé refléter le vocabulaire français à différentes tranches de fréquences. Il est à noter que nous n'avons pas analysé le degré de transparence (ou opacité) des paires de mots français-suédois ou français-anglais. Certaines paires de mots manifestent une transparence graphique totale, comme par exemple *brutal*, *division*, *global*, *incident*, mais la plupart du temps l'orthographe diffère. Le plus souvent, cette différence est relativement petite (fr. *actif*, ang. *active*, sué. *aktiv*). En cas de doute sur l'usage d'un mot dans la langue contemporaine, nous avons eu recours à des dictionnaires électroniques, à savoir *Cambridge dictionary*² pour l'anglais et *Svenska Akademiens Ordlista*³ pour le suédois. Si un mot donné n'existe pas dans un de ces dictionnaires, il n'a pas été retenu en tant que mot cognat. Ajoutons aussi que la recherche des mots cognats se fait au niveau du lemme. Les mots cognats doivent donc appartenir à la même catégorie grammaticale.

3.2. Participants

Nous avons fait passer le TTV à 30 étudiants, dont 5 hommes et 25 femmes, inscrits en première année de français à une université en Suède. Ils avaient tous entre 20 et 25 ans. Le test a été administré environ à mi-parcours de leurs études. Avant de commencer les études supérieures de français, ces étudiants avaient fait en moyenne environ 600 heures de français. Le nombre d'heures minimum pour commencer les études supérieures de français est de 420, ces heures s'étalant le plus souvent sur 4 ans au collège et un an au lycée. En dehors du minimum requis, la plupart des étudiants avaient fait de 100 à 300 heures supplémentaires au lycée.

3.3. Collecte des données

Les étudiants ont fait le test dans le cadre d'un cours de grammaire, l'un des chercheurs étant aussi leur enseignant et donc présent dans la salle de classe. Ils ont été rassurés que le test ne faisait pas partie de l'enseignement du cours, mais d'un projet de recherche. Il n'y avait pas de limite de temps, mais la plupart

des étudiants ont terminé le test en 40-45 minutes. Ils ont fait le test sur papier, sans accès à d'autres moyens qu'un crayon. Avant de commencer le test, le chercheur les a informés des considérations éthiques, selon lesquelles ils participaient volontairement et de manière anonyme. Ils avaient le droit de se retirer à n'importe quel moment de la collecte des données, ou même après celle-ci. Une fois le test terminé, les étudiants ont répondu à quelques questions concernant leur background (études préalables de français et d'autres langues, âge, sexe, etc.).

4. Résultats

4.1. La taille du vocabulaire des apprenants

La première question de recherche était : 1) Quelle est la taille du vocabulaire d'apprenants universitaires suédophones de français ? Le Tableau 1 montre la taille du vocabulaire des trente étudiants évalués. Il est important de noter que ces chiffres sont obtenus par extrapolation (voir section 4.1). La moyenne du groupe est de 7 192 mots avec une médiane de 7 250 mots. Il faut aussi noter que seuls deux étudiants ont une taille inférieure à 5 000 mots.

Taille du vocabulaire estimée (max. 10 000)	
Max	9933
Min	4233
Médiane	7250
Moyenne	7192

Tableau 1 : Taille du vocabulaire estimée

Le Tableau 2 donne le résultat global des étudiants par tranches de fréquence. Il en ressort que les étudiants maîtrisent les mots des tranches 2K et 3K à hauteur de presque 90 %, avec très peu de différences entre les deux tranches. Pour la 5K, le score moyen descend à environ 80 %, alors que celui de la 10K est sensiblement plus bas, 60 % environ. La différence relativement faible entre le 2K, 3K et 5K est probablement due au fait que la plupart des étudiants ont un résultat supérieur à 5 000 mots. En effet, un test ANOVA suivi du post-test HSD de Tukey nous montre que les différences entre la tranche de 10K et les autres tranches sont significatives ($p < 0,05$), alors qu'aucune des autres comparaisons n'a montré de différences significatives⁴.

Les résultats de nos étudiants sont comparables à ceux du groupe avancé dans l'étude de Batista et Horst, ces derniers ayant un minimum de 1 320 heures de français (cf. ci-dessus). Rappelons que nos étudiants avaient fait environ 600 heures de français avant de commencer les études de français à l'université. Nous pouvons donc en conclure que nos étudiants ont une taille de vocabulaire relativement grande par rapport au nombre d'heures d'études de français.

4.2. Effet cognat

La deuxième et la troisième question de recherche concernent la cognation :

- 2) Quel est l'effet de la connaissance de mots cognats sur les estimations de la taille du vocabulaire ?
- 3) Y a-t-il une différence d'effet concernant les mots cognats L1-L2-L3 par rapport aux mots cognats L2-L3 ?

Comme nous venons de le faire remarquer plus haut, 120 mots au total sont testés au TTV. Ces mots ont été tirés au hasard des tranches de 2K, 3K, 5K et 10 K dans deux listes de fréquences de référence. Sur ces 120 mots, 44 ont un mot cognat correspondant à la fois en suédois et en anglais, 3 mots ont un cognat en suédois⁵ (mais pas en anglais) et 34 mots ont un cognat en anglais (mais pas en suédois). Dans le test, 39 mots n'ont pas de cognats en anglais ou en suédois. Ce groupement permet de faire deux constats immédiats. Premièrement, il semble y avoir un effet cognat de façon générale puisque les étudiants ont un meilleur score sur les mots ayant un cognat. Deuxièmement, le fait qu'un mot français ait un cognat suédois et anglais semble avoir un effet plus facilitateur que le fait qu'un mot français ait un cognat anglais (voir le total du Tableau 3). Afin de vérifier si ces différences sont statistiquement significatives, nous avons fait un test ANOVA à mesures répétées sur le pourcentage de réponses correctes totales de chaque catégorie. En effet, le test montre une différence générale entre les catégories ; et le test post-hoc de Bonferroni indique qu'il y a des différences statistiquement significatives, $p < 0,05$, entre les trois catégories au total. Une séparation des mots par tranche de fréquence, permet d'apporter une description plus nuancée à l'effet facilitateur général constaté ci-dessus. En effet, cet effet se manifeste surtout au-delà des 5 000 mots, c'est-à-dire dans un intervalle fréquentiel qui est moins bien maîtrisé par les apprenants (voir les tranches 2K-10K du Tableau 3).

Batista et Horst (2016) ont séparé les informateurs par L1 (romane, asiatique et autres), ce qui leur ont permis de constater que le résultat des apprenants ayant une L1 romane était largement supérieure à ceux des deux autres groupes (asiatique et autres), voir le Tableau 4, où nous avons ajouté nos résultats à ceux de Batista et Horst (2016). Du Tableau 4, il ressort que les résultats de nos étudiants sont au niveau de ceux des L1 romane de Batista et Horst (2016). Il faut aussi noter que la distribution des scores est plus homogène dans les groupes suédois et romane par rapport aux groupes asiatique et autres (voir l'écart-type [SD] dans le Tableau 4).

Tranche	Suédois (SD) N=30	Romane (SD) N=27	Asiatique (SD) N=67	Autres (SD) N=81
2K	26,3 (2,8)	25,4 (4,2)	17,9 (7,5)	17,8 (5,7)
3K	26,5 (2,8)	24,5 (4,5)	14,3 (7,1)	14,4 (6,1)
5K	23,9 (4,4)	21,9 (5,0)	11,8 (8,2)	13,0 (6,8)
10K	17,7 (5,5)	14,0 (4,5)	6,4 (5,3)	7,0 (5,1)
Total	96,4 (13,9)	85,8 (16,1)	50,3 (25,7)	52,3 (25,7)

Tableau 4 : Les résultats des apprenants suédois par tranche comparés à ceux de Batista et Horst (2016)

5. Discussion

Les résultats nous ont montré que les scores des étudiants suédois sont élevés par rapport à ceux rapportés par David (2008), Milton (2009) et Batista et Horst (2016). Comme la comparaison inclut des étudiants de systèmes scolaires différents et de L1 différentes, il est cependant difficile de tirer des conclusions définitives. Il faut aussi prendre en compte le fait que les apprenants dans Batista et Horst (2016) sont des immigrants, qui suivent leur formation en milieu naturel, contrairement aux nôtres. Ainsi, la quantité d'*input* de la langue française devrait être considérablement plus importante pour leurs apprenants. En effet, les apprenants suédois ont rarement l'occasion de s'exposer à la langue française dans la société suédoise en général. Cependant, le fait que les scores des étudiants suédois soient proches de ceux obtenus par le groupe de langue maternelle romane dans l'étude de Batista et Horst nous amène à penser que la L1 est un facteur important, ce qui corrobore d'ailleurs les résultats de plusieurs études antérieures sur la taille du vocabulaire (Cobb, 2000 ; Lindqvist, 2020).

En outre, une analyse préalable a montré que, dans le test, 37 % des mots ont un cognat en suédois et 68 % des mots ont un cognat suédois et anglais. Si on ajoute à

cela, que, de façon générale, il a souvent été constaté que les Suédois ont un bon niveau d'anglais, nous pensons que la moyenne élevée de nos participants s'explique en grande partie par l'avantage que donne l'accumulation des connaissances des deux langues anglaise et suédoise dans ce test. Il y a, en effet, une différence significative entre les scores des étudiants obtenus sur les mots qui ont un cognat à la fois en suédois et en anglais par rapport à ceux qui ont un cognat en anglais uniquement, et aussi par rapport aux mots sans cognat. Notre étude souligne donc l'importance de mener de la recherche où les connaissances de langues étrangères supplémentaires sont prises en compte.

L'analyse des différentes tranches par rapport aux mots avec ou sans cognats nous a montré que l'effet cognat s'explique de façon presque exclusive par les mots de la tranche de 10K. Ces résultats permettent alors de mettre en question, au moins partiellement, l'importance de la fréquence lors de l'apprentissage du vocabulaire, dans la mesure où le facteur de cognation semble l'emporter sur le facteur de fréquence. En d'autres termes, les apprenants arrivent à reconnaître les mots moins fréquents grâce à leurs connaissances d'autres langues. Ainsi, nos résultats vont dans la même direction que ceux de Bardel et Lindqvist (2011) et de Bardel, Gudmundson et Lindqvist (2012), qui, nous l'avons vu plus haut, ont montré cette même tendance. Il est aussi à noter que ces dernières études traitent de la production orale et spontanée de la langue, tandis que la présente étude concerne les connaissances réceptives du vocabulaire. Nous pouvons supposer que l'effet des mots cognats soit encore plus prononcé lors d'un test du vocabulaire réceptif dans le sens où l'apprenant a plus de temps pour réfléchir et ainsi pour se servir de ses connaissances préalables à la fois en langue maternelle et en langue(s) seconde(s).

Dans les tranches de 2K, 3K et 5K les étudiants connaissent la plupart des mots, qu'ils aient un cognat anglais et suédois (*ambassadeur, global, incident, nuance, qualification, tragique, tourisme*) ou uniquement anglais (*cage, joie, législation, persuader, transition, véhicule*). Les mots sans cognats sont également relativement bien maîtrisés (*aube, bâtiment, goutte, instituteur, mentir, rêver*). Les différences que l'on constate entre mots avec et sans cognat ne sont pas significatives dans cette zone de fréquence. En effet, si on regroupe tous les mots ayant un cognat anglais et/ou suédois et les compare aux mots sans cognats, on constate que la différence entre ces deux catégories dans les tranches de 2K, 3K et 5K confondues est petite (environ 5 %) alors que dans la tranche de 10K, la différence s'élève à 32 %.

Alors comment expliquer le fait que les participants n'ont pas de meilleurs résultats sur les mots qui ont un cognat dans la zone inférieure à 5K? Tout d'abord, il faut noter que les participants ont un score moyen qui dépassent les 5 000 mots.

Ces mots sont donc relativement bien connus des participants quel que soit leur statut (cognat ou pas). Or, nous pensons aussi qu'un autre facteur à considérer est la nature potentiellement *deceptive* ou *partially deceptive* de certains cognats (Granger, 1993). En effet, il y a notamment quelques mots avec cognats anglais pour lesquels les étudiants ont donné de fausses réponses, comme par exemple *distinguer* (24 mauvaises réponses sur 30). Si on étudie de près les réponses des étudiants, on constate que, pour certains de ces mots, c'est parfois un même distracteur qui explique toutes les mauvaises réponses. On peut donc penser que les étudiants ont cru savoir la bonne réponse, mais que quelque chose les a trompés. Pour *distinguer*, il fallait relier ce mot avec la définition 'remarquer', mais les 24 étudiants qui ont fourni une fausse réponse ont tous fait le lien entre un des mots distracteurs (*mentionner*) et la définition. Les deux mots *mentionner* et *remarquer* ont des mots cognats correspondant en anglais (*mention* et *remark*), mais ces paires de mots anglais et français ont des emplois partiellement divergents, ce qui a pu dérouter les étudiants. En effet, la signification de *He mentioned that* est proche de *Il a fait remarquer que*, mais pas de *Il a remarqué que*. Le fait que la plupart des étudiants donnent une mauvaise réponse ne veut donc pas forcément dire qu'ils ne connaissent pas du tout le sens de ces mots français ; simplement, ils n'en connaissent pas tous les aspects du sens (voir Nation, 2013, p. 49). En effet, Granger (1993, p. 50) affirme que l'effet facilitateur des mots cognats est important au stade initial de l'apprentissage dans le sens où il est susceptible de faciliter la compréhension. Or, pour les apprenants plus avancés, quand il s'agit d'apprendre les contraintes quant à l'emploi et les nuances de sens, elle dit que cet effet aura moins de poids et peut être source d'erreurs.

Sur les 30 mots de la tranche de 10K, il y a un certain nombre de mots qui sont inconnus aux apprenants. On peut donc supposer que, pour un mot totalement inconnu, le mot suédois ou anglais correspondant permet aux participants d'en deviner le sens et de donner la bonne réponse. Cet effet fait monter le taux de bonnes réponses par rapport à celui des mots sans cognats. En revanche, dans un intervalle fréquentiel où les participants ont déjà rencontré la plupart des mots et où ils ont déjà une idée plus ou moins précise du sens de la plupart des mots, l'effet facilitateur lié au mot cognat suédois ou anglais ne sera pas aussi déterminant. Dans la tranche de 10K, les dix mots sur lesquels les participants font le plus de fautes sont tous des mots qui n'ont pas de cognats comme par exemple *fût*, *trapu*, *foisonner*, *scier*, *ahuri*, *crouler* ou âpre alors que les mots qui ont un cognat suédois et/ou anglais s'avèrent plus faciles (*syntaxe*, *persévérance*, *charpentier*, *recrue*).

Pour les mots sans cognats, on constate que le taux de bonnes réponses décroît en fonction de la fréquence des mots (de 94,3 % pour la tranche de 2K à 44,4 % pour

la tranche 10K) attestant une tendance générale valable pour l'apprentissage d'une L2 : on apprend en général les mots fréquents avant les mots moins fréquents. Or, pour les mots qui ont un cognat, le taux de bonnes réponses varie peu en fonction de la fréquence. Le lien qui existe de manière générale entre la fréquence des mots dans la L2/L3 et sa difficulté se trouve donc en partie annulé pour les mots qui ont un cognat.

Conclusion

Cette étude avait pour but d'examiner la taille du vocabulaire d'étudiants universitaires suédophones de français. Les résultats ont montré que les étudiants connaissent en moyenne environ 7 200 mots, un chiffre relativement élevé, vu le nombre d'années et d'heures d'études des apprenants. En outre, il est clair que la proportion de réponses correctes diminue en fonction de la zone de fréquence testée, allant de 88 % pour les zones 2K et 3K à 59 % pour la zone 10K. Par contre, il semblerait que la taille du vocabulaire élevée des étudiants peut en partie s'expliquer par le fait que les étudiants connaissent le suédois et l'anglais, et qu'ils en bénéficient en passant le test. En effet, nous avons pu constater un effet cognat en général, dans le sens où les scores pour les mots cognats sont meilleurs que ceux pour les non cognats. De plus, l'effet est renforcé lorsque les trois langues impliquées partagent les cognats. En même temps, notre étude a montré qu'une analyse des mots cognats par fréquence permet de nuancer l'effet facilitateur. Face à des mots qui leur sont pour beaucoup inconnus ou méconnus, les étudiants sont en quelque sorte débutants ; la cognation leur est très utile en permettant de deviner la bonne réponse. Mais, face à des mots qui sont pour la plupart bien connus, le rôle de la cognation a moins d'importance. Pour ces mots, il faut approfondir les connaissances, ce qui implique notamment d'apprendre tous les aspects du sens des paires de mots cognats de sémantisme partiellement divergent.

Bibliographie

- Bardel, C., Gudmundson, A., Lindqvist, C. 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.
- Bardel, C., Lindqvist, C. 2011. Developing a lexical profiler for spoken French and Italian L2. The role of frequency, cognates and thematic vocabulary. In Roberts, L., Pallotti, G., Bettoni, C. (Ed.) *Eurosla Yearbook* 11, p.75-93. Amsterdam: John Benjamins.
- Batista, R., Horst, M.-L. 2016. « A New Receptive Vocabulary Size Test for French ». *Canadian Modern Language Review* 2(2), p.211-233.
- Cobb, T. 2000. « One size fits all ? Francophone learners and English vocabulary tests ». *The Canadian Modern Language Review* 57(2), p. 295-324.

- Dahlman, G. 2019. *L'étendue du vocabulaire réceptif des apprenants de français*. Mémoire de licence non publié. Université de Göteborg.
- David, A. 2008. « Vocabulary breadth in French L2 learners ». *The Language Learning Journal* 36(2), p.167-180.
- Granger, S. 1993. « Cognates : an aid or a barrier to successful L2 vocabulary development ? ». *ITL Review of Applied Linguistics* 99/100, p. 43-56.
- Ingeland, E. 2018. *La taille du vocabulaire chez des apprenants de français de niveau universitaire*. Mémoire de licence non publié. Université de Stockholm.
- Lindqvist, C. 2018. « Le développement de la taille du vocabulaire en français L2 chez les élèves suédophones ». *Synergies Pays Scandinaves* 11-12, p.151-161. [En ligne]:http://gerflint.fr/Base/Paysscandinaves11_12/lindqvist.pd [consulté le 20 octobre 2020].
- Lindqvist, C. 2020. First and second language cognate effects in third language vocabulary size estimates. In: Booth, P., Clenton, J. (Ed.) *First language influences on multilingual lexicons*. New York: Routledge.
- 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), p. 333-348.
- Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Nation, I. S. P. 2013. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Peters, E. 2020. Factors affecting the learning of single-word items. In Webb, S. (Ed.), *The Routledge Handbook of Vocabulary Studies*, New York: Routledge, p. 125-142.

Notes

1. Deux mémoires non publiés d'étudiants de deux universités suédoises (Ingeland 2018 et Dahlman 2019) ont récemment été écrits à ce sujet. Les résultats d'Ingeland (2018) suggèrent que les étudiants en première année ont un vocabulaire de 6 831 mots en moyenne, tandis que les étudiants en deuxième année connaissent 8 224 mots en moyenne. Dahlman, qui a inclus des étudiants en deuxième et troisième années, a obtenu un résultat similaire : les participants de son étude ayant en moyenne un vocabulaire de 8 360 mots. Il faut cependant savoir que ces études incluent toutes les deux un nombre assez restreint de participants.
2. Cambridge dictionary. <https://dictionary.cambridge.org/>
3. Svenska Akademiens Ordlista. <https://svenska.se/saol/>
4. Nous tenons à remercier Amanda Edmonds, Université Paul-Valéry Montpellier, pour les conseils concernant la statistique.
5. Nous ne considérons pas ici les trois mots qui ont un cognat uniquement en suédois.