

APPRENDRE DES MOTS EN FLE AVEC CHATGPT

Questo studio si interessa all'uso di ChatGPT per l'apprendimento del vocabolario in francese come lingua straniera. Sapendo che imparare il vocabolario utilizzando schede (flashcard) costituisce uno dei metodi più efficaci, lo confrontiamo con un approccio di apprendimento basato sull'impiego di ChatGPT. I partecipanti sono 51 giovani adulti in formazione (livello terziario, universitario o pedagogico HEP), per lo più germanofoni.

In un primo momento, i partecipanti dovevano imparare 20 parole con due supporti diversi (Quizlet e ChatGPT) e in quattro liste distinte, al fine di variare le parole studiate con entrambi i supporti. Una metà dei partecipanti ha lavorato prima con Quizlet e poi con ChatGPT, mentre l'altra metà ha seguito la procedura in ordine inverso. Ai partecipanti è stato poi chiesto di utilizzare le parole apprese per scrivere una lettera di reclamo riguardante la Mensa della loro alta scuola. La ritenzione delle parole è stata infine testata tramite un post-test dopo una settimana.

I risultati mostrano che le parole apprese con Quizlet vengono utilizzate maggiormente nel compito scritto rispetto alle parole studiate con ChatGPT. Il post-test indica inoltre che i partecipanti ricordano meglio le parole apprese con Quizlet. Discutiamo questi risultati ponendoci la questione di quanto le due condizioni di apprendimento possano aver favorito la creazione e l'approfondimento delle tracce mnemoniche, facendo riferimento ai concetti di profondità di elaborazione (depth of processing), involvement load, e a teorie che considerano le risorse limitate nell'apprendimento del vocabolario.

[traduit par ChatGPT]

Tema

Raphael Berthele | Université de Fribourg



Raphael Berthele est professeur de plurilinguisme à l'université de Fribourg. Il a étudié et travaillé aux universités de Fribourg, Tübingen, Berkeley et Berne. Il a cofondé l'Institut de plurilinguisme fribourgeois en 2008. Il est responsable de deux programmes de master en sciences du plurilinguisme et en didactique des langues étrangères. Ses recherches portent sur divers aspects du plurilinguisme, du plan cognitif au plan social. Ces dernières années, il s'est consacré à l'étude empirique du plurilinguisme réceptif, à l'acquisition de la littératie plurilingue, à l'expression des relations informationnelles et spatiales chez les bi- et plurilingues, et aux différences individuelles dans l'apprentissage des langues.

Contexte et questions de recherche

Les enquêtes menées auprès des enseignant-e-s de langues vivantes révèlent souvent que le développement des ressources lexicales en langue cible demeure un défi majeur. En comparaison de l'apprentissage incident du vocabulaire qui repose, par exemple, sur la lecture ou le visionnage de matériel audiovisuel, les études empiriques démontrent la plus grande efficacité de l'apprentissage délibéré et ciblé de mots, notamment par l'emploi de fiches (parfois appelées *fiches bristol*, angl. *flashcards*; cf. Webb et al. 2020). Dans le projet de recherche présenté ici, notre but était de comparer l'efficacité de l'apprentissage par fiches à une méthode alternative qui implique l'interaction des apprenant-e-s avec un système comme ChatGPT, désormais largement accessible depuis que les agents conversationnels se sont généralisés. Concrètement, nos deux questions étaient les suivantes :

Est-ce que l'interaction avec ChatGPT en tant que 'coach lexical' pour des mots difficiles...

- (1) ... aide les apprenant-e-s à utiliser ces mots dans une tâche écrite?
- (2) ... permet d'apprendre ces mots de manière efficace?

Notre hypothèse était que le travail avec un agent conversationnel permettrait aux apprenant-e-s de développer et d'approfondir leurs connaissances lexicales avec davantage d'engagement. En raison de la nouveauté de l'outil et de son caractère ludique, nous partions du principe qu'une activité interactive avec ChatGPT entraînerait un degré d'engagement plus élevé dans la tâche. Cette intuition se voyait renforcée par le fait que le traitement cognitif de l'output généré par cet outil devait naturellement inciter l'apprenant-e à évaluer la justesse et la pertinence du texte produit, ce qui conduirait à une meilleure mémorisation des items lexicaux, selon la théorie

de la charge d'implication ('involvement load'; Yanagisawa & Webb, 2022). L'implication, telle qu'elle est conçue dans cette approche théorique, comporte trois dimensions qui chacune contribue à l'apprentissage de nouveaux mots : Le besoin d'exprimer une idée dans la langue cible ('need'), la recherche de moyens linguistiques permettant de l'exprimer ('search') et la vérification du résultat de cette recherche lors de la mise en mots ('evaluation'). Dans le cadre de notre projet, les deux dimensions recherche et évaluation semblent particulièrement pertinentes. De plus, l'interaction avec l'agent conversationnel devrait, selon la théorie de la profondeur du traitement (Craik & Tulving, 1975), également favoriser la mémorisation, puisque la tâche décrite ci-dessous met l'accent principalement sur la dimension sémantique des mots.

Tâche et déroulement

La langue cible de notre expérience didactique était le français langue étrangère. Les participant-e-s étaient des apprenants adultes majoritairement germanophones, tous étudiant-e-s au sein d'une haute école (pédagogique ou universitaire) en Suisse. Nous avons conçu une tâche écrite dans une orientation communicative. Cette tâche consistait à rédiger un courriel au rectorat de la haute école, afin de contester les prix trop élevés pratiqués par la Mensa. De plus, la consigne prévoyait que d'autres doléances doivent figurer dans cette lettre (environ 250 mots), notamment la mise à disposition de davantage d'espace de pique-nique et de fours à micro-ondes.

Avant de récolter les données, nous avons fait générer à l'IA ChatGPT (gpt-3.5-turbo) plusieurs lettres sur la base de cette consigne. Dans ces lettres, nous avons identifié des entités lexicales difficiles à apprendre, d'abord de manière introspective et intuitive. Nous nous sommes abstenus de pré-tester la connaissance de ces mots, car un tel pré-test aurait potentiellement déjà contribué à l'apprentissage des items. D'autre part, pour l'expérience finale, nous n'avons sélectionné que des mots qui, selon l'outil Flelex¹, sont enseignés à des niveaux avancés (C1, C2). Ceci afin de minimiser la probabilité que ces mots fassent déjà partie du vocabulaire productif des apprenants.

« Dans la condition ChatGPT, un éventuel effet positif dû à l'output idiomatique en langue cible illustrant l'emploi des mots en contexte serait donc contrecarré par une surcharge informationnelle. »

abordable / accroître / afin de / bondé / chronophage / défavoriser / en dépit de / en outre / impérativement / judicieux / l'agacement (m.) / l'aménagement / l'entrave (f.) / l'épanouissement / l'infléchissement (m.) / le gaspillage / notamment / pallier / soucieux / subvenir aux besoins de qqn

Au moyen de ces 20 items, nous avons composé quatre listes différentes qui ont été attribuées aux participant-e-s de manière aléatoire. Chaque participant-e a tenté de mémoriser la moitié des mots durant 15 minutes à l'aide des questions spécifiques posées au ChatGPT et l'autre moitié des items, pour la même durée, avec des fiches numériques (sur la plateforme Quizlet). Une moitié des participant-e-s a travaillé d'abord avec Quizlet et ensuite avec Chat GPT, alors que l'autre moitié des participants a suivi la procédure inverse.

Les apprenant-e-s étaient libres de travailler avec des fiches Quizlet en version français-allemand, français-italien ou français-anglais. La réalisation de la tâche avec ChatGPT était standardisée dans la mesure où les participant-e-s étaient priés d'utiliser pour chaque mot des consignes (*prompts*) préformulées. La première consigne consistait à demander à ChatGPT d'*expliquer le mot* «...» *de manière simple*. Puis, il fallait demander à l'outil d'*écrire 3 phrases avec le mot* «...» afin d'obtenir des exemples de son usage en contexte. Les réponses étaient sauvegardées au moyen d'un formulaire dans lequel les participant-e-s collaient l'output obtenu (voir l'encadré pour un exemple).

¹ <https://cental.uclouvain.be/cefrlex/flelex/search/>

(1) Explique le mot suivant en langage simple : **pallier**

Le mot «pallier» signifie trouver une solution temporaire à un problème ou une situation difficile, sans pour autant résoudre complètement le problème à long terme. C'est un peu comme mettre un pansement sur une blessure, cela aide à limiter les dégâts momentanément, mais cela ne guérit pas complètement la blessure.

(2) Fais trois phrases en français qui contiennent ce mot : **pallier**

1. Il faut trouver une solution pour pallier ce manque de personnel.

2. Les vitamines peuvent aider à pallier les carences alimentaires.

3. Les mesures d'urgence ont été mises en place pour pallier les conséquences de la catastrophe naturelle.

Les fiches Quizlet, rédigées de façon minimale, comportaient uniquement la traduction du mot français vers l'allemand, l'anglais ou l'italien. Dans l'élaboration de notre expérience, nous avons évoqué des alternatives, par exemple d'ajouter des phrases contenant le mot cible ou des images. Finalement, nous avons opté pour la solution la plus minimaliste afin de maximiser l'effet (hypothétique) de l'information beaucoup plus substantielle fournie par ChatGPT sur l'utilisation et l'apprentissage des mots.

Dans le cas de mots polysémiques, seule la traduction pertinente pour la tâche décrite ci-dessus était indiquée (p. ex. *pallier – ausgleichen/ovviare/compensate for*).

Afin de s'assurer que l'output généré par l'IA soit lu, les participant-e-s devaient répondre à de simples questions concernant les réponses fournies par ChatGPT (qualité de la description du sens ou pertinence des trois exemples). Les réponses des participant-e-s n'ont pas été analysées.

Dans la phase finale de la récolte, nous avons demandé aux participant-e-s d'écrire la lettre de réclamation concernant la mensa en les priant explicitement d'utiliser autant de mots que possible. Ces lettres manuscrites ont été transcrites par nos soins. L'analyse a ensuite principalement consisté à répertorier, dans chaque lettre, les emplois de mots appris. De manière plus approfondie, nous avons également observé si les emplois de ces mots avaient donné lieu à des erreurs et, le cas échéant, celles-ci ont été catégorisées en différents types (p. ex. orthographe, grammaire). Exactement une semaine après la rédaction de la lettre, les participant-e-s ont passé un post-test, afin de vérifier si les nouveaux mots français avaient bien été mémorisés. Ce test a opérationnalisé la profondeur des connaissances lexicales en demandant aux participants de traduire le mot cible en L1, et en leur donnant la possibilité de former une phrase dans la langue cible contenant ce mot.

Échantillon

Plusieurs groupes d'apprenant-e-s ont participé au projet de recherche. Les participant-e-s étaient des étudiant-e-s de divers programmes et facultés de l'Université de Fribourg ainsi que des HEP (Hautes Écoles pédagogiques) de Zurich et de Berne. Au total, 51 personnes ont pris part à l'étude. La moyenne d'âge était de 22,7 ans. Le niveau de français de ces apprenant-e-s se situait entre B1 et B2. Les participant-e-s avaient majoritairement l'allemand – seule ou combinée à une seconde langue – pour L1 (N =38), mais d'autres L1 étaient également représentées, à savoir : l'italien (N=12), l'albanais (N=2), l'anglais et le tamoul (N=1 respectivement). Cinq participant-e-s se considéraient comme bilingues avec l'allemand et une autre langue (auto-évaluation).

35 étudiant-e-s (groupe expérimental) ont suivi le protocole décrit ci-dessus, alors que 16 autres n'ont travaillé que

« Nous avons constaté que les apprenant-e-s qui travaillaient avec ChatGPT étaient confronté-e-s à une activité complexe nécessitant le traitement d'une grande quantité d'informations. »

10 mots de la liste, mais au moyen des deux plateformes (Quizlet et ChatGPT). La constitution de ce 'groupe contrôle' était en fait une erreur en matière de design expérimental. Celle-ci s'est cependant avérée utile puisque ce groupe nous permet de comparer l'utilisation et la mémorisation de mots non-travaillés ainsi que de mots ayant été appris avec les deux méthodes.

Mots employés dans les textes

Les analyses des lettres rédigées par les participant-e-s indique que les mots appris dans Quizlet (moyenne de 3.3 mots utilisés) sont davantage employés que ceux étudiés grâce à ChatGPT (moyenne de 1.3 mots). La figure 1 permet de comparer l'emploi de chacun des 20 mots dans les deux conditions. Seul un mot (*afin de*) apparaît plus fréquemment dans les textes lorsqu'il est appris selon la condition ChatGPT. Le groupe contrôle (données non présentées dans le graphique), qui a travaillé les mêmes 10 mots avec les deux outils, utilise en moyenne 5.9 mots dans les textes rédigés. Les mots absents de la liste du groupe contrôle ne sont employés que sporadiquement (moyenne de 0.20).

Bien que les résultats que nous donnent à voir la statistique descriptive soient d'ores et déjà éloquentes, nous avons voulu vérifier l'impact de la condition sur l'utilisation des mots appris lors de la rédaction des lettres au moyen d'un modèle logistique mixte. Ce modèle a pour fonction de mesurer la probabilité de l'emploi d'un mot spécifique des listes de vocabulaire mémorisé dans le texte écrit. Le facteur fixe correspond à la condition (Quizlet vs. ChatGPT), alors que les facteurs aléatoires sont l'item lexical (ordonnées à l'origine aléatoires), ainsi que le/la participant-e (ordonnées à l'origine aléatoires et pentes par condition aléatoires). Le résultat montre un effet significatif ($p < 0.001$) de la condition (les chances logarithmiques augmentent de 1.5 pour la condition Quizlet). L'inclusion des LI comme covariable ne contribue pas au modèle de manière notable.

Dans la plupart des cas, lorsqu'un mot est employé, les erreurs sont plus nombreuses dans la condition ChatGPT que dans la condition Quizlet. Nous pouvons donc conclure qu'au moment de

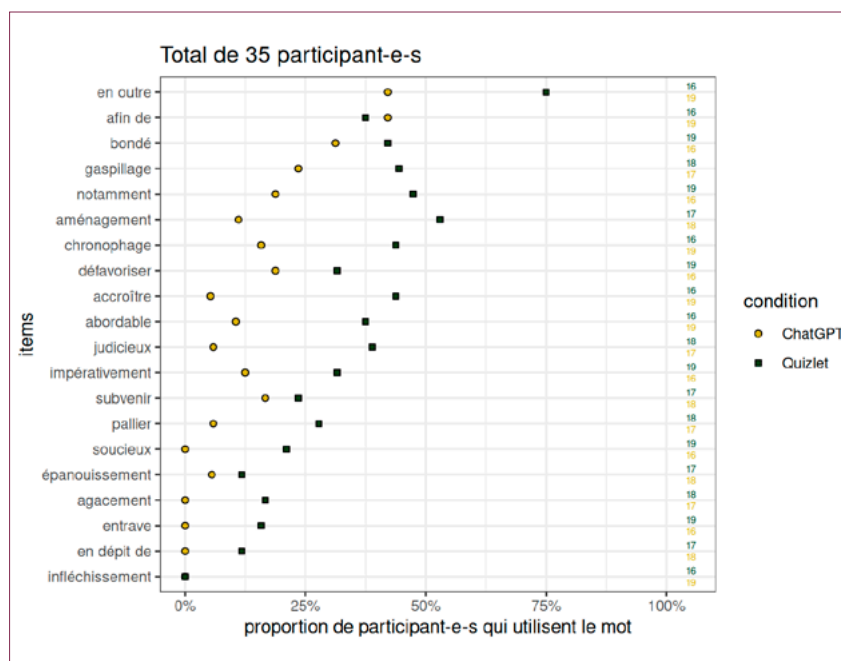


Figure 1

Proportion d'utilisation par mot et par condition (groupe contrôle pas compris)

la rédaction d'un texte, il est plus probable de se rappeler d'un mot appris avec Quizlet et que son emploi sera également plus correct.

Mots retenus après une semaine

Lors de la passation du post-test, après une semaine, chaque participant-e a dû répondre à des questions posées sur chaque mot étudié. En nous basant sur l'idée de la profondeur des connaissances lexicales (Wesche & Paribakht 1996), chaque mot (p. ex. <pallier>) a été présenté aux participant-e-s qui devaient indiquer / formuler une réponse selon les options suivantes :

- Je n'ai encore jamais vu ce mot.
- J'ai déjà rencontré ce mot mais je ne sais pas / plus ce qu'il signifie.
- J'ai déjà rencontré ce mot et il signifie ... (entrez un synonyme, une traduction, une brève explication dans le champ qui s'affichera)

Je sais comment utiliser ce mot correctement dans une phrase en français.

Lorsque la troisième option était choisie, un champ additionnel à compléter s'affichait. Ensuite, le dernier champ s'affichait afin que les participant-e-s puissent entrer une phrase (colonne 'bonus' de la Figure 2). Les réponses entrées dans ces deux derniers champs ont été évaluées manuellement par notre équipe.

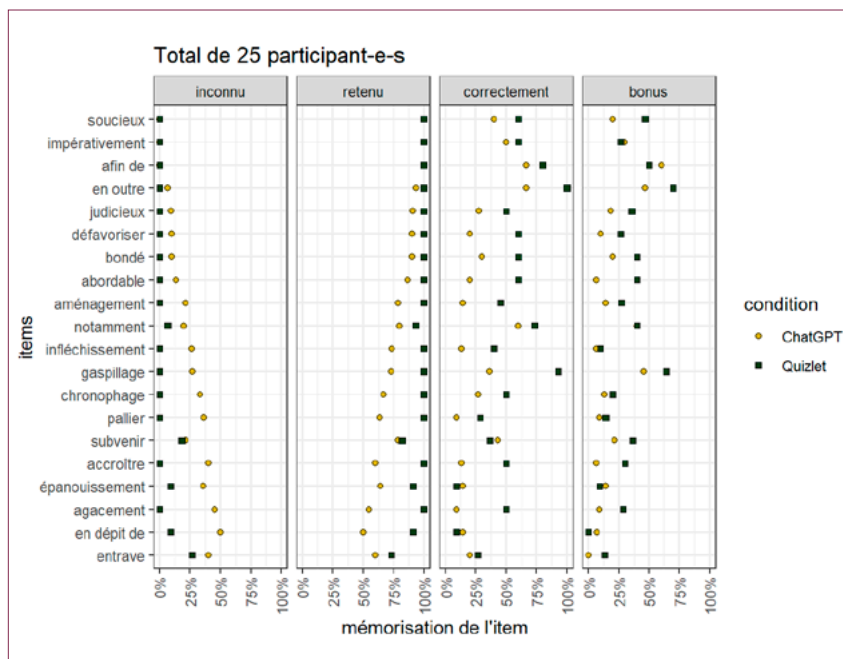


Figure 2
 Résultats du post-test après une semaine en fonction des conditions. Souvenir du sens des mots et utilisation correcte du mot dans une phrase (colonne 'bonus').

Sur la base de ces données, nous pouvons comparer le niveau de rétention en mémoire des items dans les deux conditions. En utilisant Quizlet, les participant-e-s se rappellent en moyenne d'environ 50 % des mots, pour seulement 30 % dans la condition ChatGPT. La figure 2 distingue plusieurs niveaux de connaissances lexicales, selon que les participant-e-s se rappellent du mot (mais peut-être seulement de manière incorrecte); se rappellent du mot de manière correcte (les participant-e-s ont été capables de donner un synonyme, une traduction, etc.); savent utiliser le mot correctement en contexte.

Les résultats du groupe ayant utilisé les deux méthodes pour les mots ne sont pas donnés ici. Sans surprise, le taux de rétention est bien plus élevé dans cette condition par rapport aux deux conditions comparées dans la Figure 2.

À nouveau, la modélisation statistique, que ce soit pour le niveau de mémorisation le plus bas (les participant-e-s ne se rappellent pas vs. se rappellent du mot) ou pour un niveau plus élevé (aucun souvenir ou souvenir incorrect vs. souvenir correct du mot) tel que décrit ci-dessus, indique des résultats significativement meilleurs dans la condition Quizlet ($p < 0.05$, augmentation des chances logarithmiques de 4.18 pour le premier

niveau, i.e. aucun souvenir vs. souvenir; $p < 0.001$, augmentation des chances logarithmiques de 1.43 pour le deuxième niveau, i.e. aucun souvenir ou souvenir incorrect vs. souvenir correct du mot).

Discussion

Cette exploration de l'utilisation d'un agent conversationnel à des fins pédagogiques est une première tentative qui présente des limites importantes.

Lors de l'observation du travail individuel ainsi que lors du débriefing avec les participant-e-s, nous avons constaté que les apprenant-e-s qui travaillaient avec ChatGPT étaient confronté-e-s à une activité complexe nécessitant le traitement d'une grande quantité d'informations. Les apprenant-e-s qui disposaient d'un niveau de lecture intermédiaire en français passaient ainsi beaucoup de temps à lire et à essayer de comprendre l'output de l'outil. Notre paradigme expérimental qui prévoyait d'employer un nombre de mots pour une passation estimée à une heure, a probablement surestimé la quantité d'informations que ces apprenant-e-s étaient en mesure de traiter en langue étrangère. La plus grande richesse informationnelle qui pourrait potentiellement améliorer l'apprentissage selon les théories de la charge d'implication ('involvement load') et de la profondeur de traitement a été en quelque sorte noyée dans la surcharge d'informations et dans la complexité de la condition ChatGPT.

Les meilleurs résultats observés dans la condition Quizlet (bien moins riche en informations) confirme le principe sous-jacent à certaines théories de l'apprentissage lexical qui postulent qu'une possible compétition entre différentes dimensions informationnelles péjorerait l'apprentissage (Barcroft 2015). L'effet potentiellement délétère d'un surplus d'informations expliquerait aussi d'autres observations que nous avons réalisées dans une étude récente à propos de l'emploi contre-productif de fiches répondant à une modalité plurilingue (plutôt que bilingue) pour l'apprentissage de nouveaux mots (Deuber & Berthele 2024). Dans la condition ChatGPT, un éventuel effet positif dû à l'output idiomatique en langue cible illustrant l'emploi des mots en contexte serait donc contrecarré par une surcharge informationnelle.

« L'emploi d'agents conversationnels dans une perspective de développement et d'approfondissement des connaissances lexicales présente un potentiel didactique certain. »

Nous restons néanmoins convaincus que l'emploi d'agents conversationnels dans une perspective de développement et d'approfondissement des connaissances lexicales présente un potentiel didactique certain, ce qu'ont confirmé certain-e-s participant-e-s à l'étude. Le protocole expérimental employé lors de cette dernière était cependant trop rigide et la quantité d'information à traiter dans un laps de temps imposé et relativement court était trop importante.

Plusieurs pistes pourraient être poursuivies pour de futures explorations. Afin d'augmenter la validité écologique de la tâche, une plus grande liberté de choix pourrait être donnée aux apprenant-e-s en ce qui concerne les prompts qu'ils entrent dans l'agent conversationnel. Ceci rendra bien évidemment l'analyse plus compliquée puisque le comportement des apprenant-e-s ne sera très certainement pas uniforme. Cependant, des enquêtes plus qualitatives et observationnelles sur l'utilisation d'agents conversationnels pourraient mettre en exergue l'emploi par les apprenant-e-s de stratégies interactives fructueuses et innovantes – dont on pourrait, dans un deuxième temps, à nouveau tester les apports en suivant un paradigme expérimental plus contrôlé.

Les agents conversationnels pourraient bien évidemment être mis à profit dans

une perspective acquisitionnelle au travers d'autres types d'activités. L'apprenant-e peut, par exemple, se servir de l'output de ChatGPT pour le comparer à ses propres productions, s'en servir à des fins de 'correction' ou en vue de la réécriture d'un précédent brouillon. Par une telle procédure, l'apprenant-e renouvellerait ainsi la comparaison entre la production rédigée de manière autonome et la version corrigée par l'IA.

Ces scénarios (et bien d'autres) partent évidemment du principe que les apprenant-e-s ont véritablement envie d'apprendre à mieux formuler et pas simplement à produire rapidement du texte dans la langue cible. Dans ce dernier cas, la disponibilité de ChatGPT, DeepL et autre GoogleTranslate rendrait l'apprentissage de compétences rédactionnelles dans une langue étrangère superflu.

Références bibliographiques

Barcroft, J. (2015). Specificity in Type of Processing and Learning: The TOPRA Model. In *Lexical Input Processing and Vocabulary Learning* (pp. 57–68). John Benjamins Publishing Company. <https://benjamins.com/catalog/lllt.43.05ch5>

Craik, F., & Tulving, E. (1975). Depth of Processing and the Retention of Words in Episodic Memory. *Journal of Experimental Psychology: General*, 104, 268–294. <https://doi.org/10.1037/0096-3445.104.3.268>

Deuber, F., & Berthele, R. (2024). Vocabulary card learning: Testing multilingual translations and increased semantic processing. *Language Teaching Research*, 13621688241254311. <https://doi.org/10.1177/13621688241254311>

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How Effective Are Intentional Vocabulary-Learning Activities? A Meta-Analysis. *The Modern Language Journal*, 104(4), 715–738. <https://doi.org/10.1111/modl.12671>

Wesche, M., & Paribakht, S. T. (1996). Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth. *Canadian Modern Language Review* 53, 13–40.

Yanagisawa, A., & Webb, S. (2022). Involvement load hypothesis plus: Creating an improved predictive model of incidental vocabulary learning. *Studies in Second Language Acquisition*, 44(5), 1279–1308. <https://doi.org/10.1017/S0272263121000577>

Remerciements: cette étude a été conçue et menée lors d'un séminaire que j'offre dans le cadre du Master Didactique des Langues Étrangères de l'Université de Fribourg. Ce projet collaboratif a été mené à bien par notre groupe au cours du semestre de printemps 2023. Je tiens à remercier Benjamin Abt, Laura Berger, Jordy Lorenzo, Livia Rose, Eva Roski, Noémie Schalbetter et Lisa Schneider pour leur collaboration. Merci aux collègues de différentes haute écoles (Oriana Sassone PHZH, Elisabeth Reiser Bello-Zago UNIFR, Simone Ganguillet PHBE) et leurs étudiant-e-s d'avoir participé à la récolte de données. Un grand merci aussi à Nicolas Félix pour sa relecture critique d'une première version de cet article.