

# **L'importance d'un programme lexical en didactique des langues étrangères**

Heather Hilton<sup>1</sup>, Michael Gauthier<sup>1</sup>, & Ronald Peereman<sup>2</sup>

1. CRTT, Université Lyon 2

2. LPNC, Université Grenoble Alpes

## **Introduction : lexique et didactique des langues**

La notion d'un programme lexical semble aussi ancienne que la didactique des langues étrangères, car il est évident que les mots constituent un élément central de la langue à apprendre, et que les apprenants ont besoin d'être guidés dans leur apprentissage de plusieurs milliers de mots. Quels mots apprendre ? Par où commencer ? Le grand didacticien tchèque Jan Komenski (Comenius) a construit ses manuels autour des « mots les plus usités » du latin, avec une phrase pour chaque nouveau mot, et des images pour illustrer le sens de 800 mots du niveau débutant (le célèbre *Orbus sensualium pictus* de 1658). Le manuel intermédiaire, *Janua linguarum reserata*, comporte 1200 nouveaux mots, présentés dans « douze centuries de sentences » (1643 : 4-5). Les mots ont été si centraux en didactique des langues étrangères au fil des siècles qu'il est difficile de comprendre comment, dans la didactique communicative qui domine en Europe depuis le milieu des années 1980, le lexique a glissé à l'arrière-plan de nos préoccupations méthodologiques, et le syllabus lexical est devenu un sujet quasiment tabou (Arrouays 1985).

Dans cet article, nous allons souligner d'abord l'importance du lexique dans l'utilisation communicative du langage, avant de considérer les travaux scientifiques dédiés au sujet de la fréquence lexicale : au milieu du XX<sup>e</sup> siècle, les « listes de fréquence » sont devenues une ressource incontournable pour

l'enseignement des langues vivantes. Nous présenterons ensuite une étude récente qui a analysé les contenus lexicaux de 16 manuels d'anglais langue étrangère (avec leurs cahiers), destinés aux collégiens français. Cette analyse illustre l'absence d'une programmation lexicale dans la didactique « actionnelle » qui est en vigueur dans nos écoles et programmes.

## **Lexique et compétence communicative**

Nous savons depuis longtemps qu'il y a une corrélation « robuste » (significative et régulièrement constatée) entre les connaissances lexicales d'un individu et sa capacité à utiliser, par exemple, sa langue maternelle (L1<sup>1</sup>). Cette corrélation est constatée pour la lecture en L1 (Lee 2010), mais également pour la compréhension de l'oral (Andringa *et al.* 2012). En langue étrangère ou seconde (L2), la recherche a observé des corrélations semblables pour les compétences réceptives (Zhang & Zhang 2020), ainsi que des corrélations élevées entre la richesse lexicale d'un discours oral, et la fluidité (ou aisance) de sa production<sup>2</sup>.

D'autres études ont montré que la compréhension aisée d'un discours oral en L2 nécessite la reconnaissance automatique<sup>3</sup> de 96 % des mots qui le composent, et de 98 % des mots dans un texte écrit (Nation 2006 ; van Zeeland & Schmitt 2013). Pour la lecture aisée d'un texte professionnel en anglais L2, Paul Nation et ses collègues ont calculé qu'il faut un lexique de quinze à vingt mille mots (Nation 2001 : 20), soit 70 % environ des connaissances lexicales d'un anglophone adulte lettré (estimé à trente ou trente-cinq mille mots). Une équipe néerlandaise confirme que dix mille mots serait un minimum pour fonctionner au niveau universitaire dans une langue étrangère (Hazenberg & Hulstijn 1996). Basé sur ces chiffres décrivant une compétence de niveau avancé (ou « expérimenté »), et sur les travaux de Paul Nation (2006), le Tableau 1 présente

---

1. Notre utilisation des abréviations « L1 » et « L2 » reflète les conventions scientifiques internationales et une chronologie acquisitionnelle ; aucun jugement de la valeur n'est exprimé par ces chiffres.

2. Dans (Référence Auteur 1), nous avons constaté une corrélation quasi-parfaite (.90) entre la diversité lexicale et un taux temporel de production (mots par minute) dans le corpus oral PAROLE.

3. Qui ne nécessite pas un effort métalinguistique conscient pour accéder au sens des mots.

notre estimation des connaissances lexicales qui pourraient sous-tendre les différents « niveaux de référence » décrits dans le *Cadre européen* (2001).<sup>4</sup>

**Tableau 1 – Connaissances lexicales et niveaux de référence européens**

niveau européen	lexique réceptif (estimé)	objectifs (LV1) de l'Education nationale française
C2	15 000+ mots	(lecture aisée de romans, textes spécialisés : 20 000 mots)
C1	10-12 000 mots	niveau visé, fin lycée filière « LLCER »
B2	8000 mots	niveau visé, fin lycée
B1	4500 mots	niveau « seuil », fin de la scolarité obligatoire
A2	2000 mots	niveau visé, fin 5 <sup>e</sup>
A1	800 mots	niveau visé, fin primaire

Une grande question de la didactique des langues – maternelle, seconde, étrangère – devient donc l'identification des mots à apprendre et leur programmation dans le temps.

### **Listes de fréquence et enseignement des langues**

Dès ses débuts, la psychologie éducative s'est préoccupée de la programmation des apprentissages lexicaux à l'école. Pendant les premières décennies du XX<sup>e</sup> siècle, le psychologue américain Edward Thorndike (Columbia University) a

---

4. Des estimations de ce type n'ont pas encore été entreprises par la Commission culturelle du Conseil de l'Europe. L'analyse menée par Suzanne Hazenberg et Jan Hulstijn en 1996 suggère que bon nombre d'estimations modestes (trois à cinq mille mots pour l'aisance en L2 ou le niveau B2) sont nettement en deçà des exigences de la communication en temps réel – que ce soit en modalité orale ou écrite.

développé une méthodologie pour identifier les mots à inclure dans un syllabus lexical gradué pour la langue anglaise (L1). Selon Thorndike, l'apprenti-lecteur doit apprendre les mots les plus fréquents d'abord (ceux qu'il rencontrera le plus souvent) avant d'aborder des mots de moins en moins fréquents.

Pour élaborer les *Teacher's Word Books*, Thorndike et ses collègues ont rassemblé un grand corpus, composé non seulement de textes scolaires et littéraires, mais aussi de catalogues commerciaux, de publicités et articles de journaux, de lettres administratives, etc. En 1921, le corpus comportait 4,5 millions de mots ; il fut constamment élargi, pour atteindre 23 millions de mots en 1944 – un corpus colossal dans un monde non-numérique. Le premier *Livre des mots* (1921) présente les 10 000 mots les plus fréquents de l'anglais « juvénile », qui deviennent la base d'un programme lexical pour les écoles primaires aux États-Unis. Le deuxième livre présente 10 000 mots de plus pour l'éducation secondaire (Thorndike & Lorge 1932), et l'édition de 1944 présente 10 000 mots additionnels d'anglais général adulte. Les mots sont organisés selon leur fréquence mais aussi leur pertinence pour chaque groupe d'âge. Qualifiées d'« une des ressources scientifiques les plus utiles jamais développée » (Goodenough 1950 : 296), les listes de Thorndike et ses collègues ont également servi au développement d'outils pour mesurer les connaissances lexicales et des niveaux de compétence écrite (lecture et écriture) en L1.

En collaboration avec Thorndike, trois linguistes enseignant l'anglais en Asie – Lawrence Faucett, Harold Palmer et Michael West – ont élargi et ajusté le corpus juvénile, le portant à cinq millions de mots (Faucett et al. 1936). Ils y ont identifié les mots les plus utiles pour l'apprentissage de l'anglais langue étrangère/ seconde, une liste de 2000 « familles de mots » (environ 4000 mots) les plus « serviables » (*of greatest general service*), selon des critères de pertinence et de fréquence (West 1953)<sup>5</sup>. La « GSL » (*General Service Word List*) a servi de base non seulement aux manuels d'anglais L2, mais également à la rédaction de livres de lecture gradués. Elle a été mise à jour en 2013 (à partir d'un corpus numérique de 273 millions de mots), et la *New GSL* comporte actuellement 2 800 familles de mots (Browne et al. 2013).

---

5. Les Britanniques West et Palmer enseignaient l'anglais en Inde et au Japon (respectivement), et l'Américain Faucett avait effectué une partie de ses études à Oxford University (Angleterre), avant d'enseigner l'anglais en Chine et au Japon (où il a rencontré Palmer). Faucett et un collègue japonais, Itsu Maki, avaient déjà élaboré une liste de 1500 mots pour la conception de manuels et de livres de lecture en anglais (Smith 2021).

Ces travaux en anglais L2 ont inspiré des démarches pour d'autres langues dans les années 1950 et 1960. Les linguistes de l'École normale supérieure de Saint-Cloud ont constitué un corpus de français oral, qu'ils ont utilisé pour identifier les mots et les structures grammaticales à incorporer aux manuels d'enseignement du français langue étrangère et seconde (Gougenheim *et al.* 1964). Mais lors de la « révolution » communicative des années 1970 et 1980 en Europe, les didacticiens ont abandonné la notion d'un programme lexical, associant la lexicologie quantitative à la linguistique structurale, qu'ils opposaient à leur nouvelle orientation sociolinguistique (Debyser 1985). L'attention didactique s'est focalisée sur les « activités langagières » (compréhension, expression, communication), et le programme s'organise autour des fonctions interactionnelles du langage (Coste *et al.* 1976). Cette focalisation sur la communication a déstructuré la programmation des éléments formels de la langue à apprendre (vocabulaire, morphologie, syntaxe, prononciation), les ingénieurs de formation intégrant les formes selon les besoins de telle situation interactionnelle, et non plus selon des critères linguistiques (comme la fréquence d'un mot ou d'une structure : Meara 1980 ; Swan 1985 ; Nordlund 2016). La dernière décennie en France a vu le retour des préoccupations lexicales dans les classes de français à l'école primaire (Dehaene 2011), mais ce renouveau d'intérêt pour le vocabulaire n'est pas encore reflété dans les textes qui régissent l'enseignement des langues vivantes en France<sup>6</sup>.

## **L'environnement lexical des classes de collège – le projet *apprend2***

Dans ce contexte, un groupe de chercheurs dans quatre universités françaises a élaboré un projet visant à analyser de plus près l'acquisition lexicale dans des classes de langue vivante en France, à l'école élémentaire et au collège<sup>7</sup>. Un volet du projet porte sur les contenus lexicaux auxquels les élèves de langue vivante sont confrontés, notamment en collège.

---

6. Le mot *vocabulaire* apparaît dans le *Socle commun* 80 fois dans les sections consacrées à l'enseignement du français, de l'histoire, des arts plastiques, etc. – et une seule fois dans les sections dédiées à l'enseignement des langues vivantes (Ministère de l'Éducation nationale 2015).

7. *APPRELD2, Learning and development of the L2 lexicon in a school learning context*, a obtenu un financement de l'Agence nationale de la recherche en 2016. Cet article résume une partie des travaux entrepris dans la première « tâche » du projet, qui en comporte cinq au total.

## **Un corpus de manuels**

L'équipe appreL2 a numérisé un corpus de 32 supports : 16 manuels d'anglais (quatre par année de collège), avec leurs 16 cahiers d'activité, pour un corpus de 657 940 mots au total<sup>8</sup>. L'*étiquetage grammatical* du corpus a été réalisé avec le logiciel *CLAWS 7* (Garside et Smith 1997), qui reconnaît et indique automatiquement si chaque mot du corpus est un nom, verbe, adjectif, pronom, etc. Le programme *Stanford NLP* (Toutanova *et al.* 2003) fut utilisé pour *lemmatiser* le corpus : il rassemble sous une seule base verbale ses différentes formes conjuguées, et sous un nom ou adjectif ses variantes selon le nombre ou le genre. Le corpus comporte 11 640 lemmes différents au total.

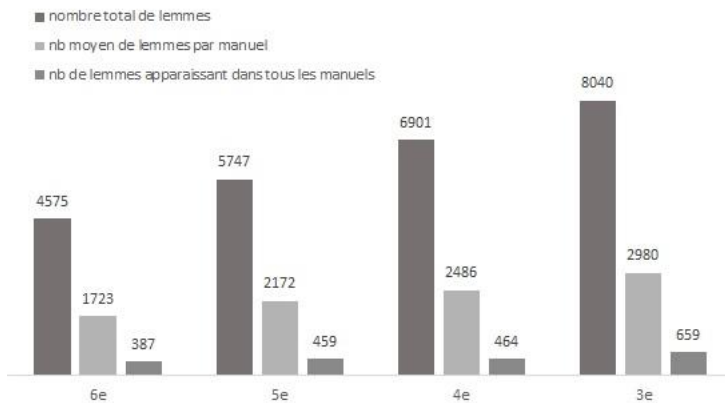
L'analyse de cette base de lemmes nous permet d'estimer le lexique qui est présenté aux élèves de chaque année de collège, et de comparer les supports d'enseignement. Nous pouvons également faire des comparaisons entre le contenu des supports et des outils didactiques comme la *New General Service List*.

---

8. Les noms propres, sigles et interjections sont exclus de ces analyses. Les manuels des différentes années viennent des quatre mêmes maisons d'édition.

## Analyses

Nos analyses montrent des supports dont les contenus lexicaux sont en net contraste avec l'idée d'une programmation structurée et progressive des mots d'une nouvelle langue. La Figure 1 présente le nombre de lemmes recensés (axe ordonné) par année scolaire (abscisse) dans les 32 supports analysés.



**Figure 1 – Le nombre de mots différents dans les manuels, par année scolaire**

Si nous regardons d'abord les barres foncées (représentant le nombre total de lemmes dans les manuels d'une même année), nous voyons que le nombre de mots augmente chaque année, de façon prévisible. De façon moins prévisible, il y a un nombre étonnamment élevé de mots différents figurant dans ces manuels de collège. Avec près de 4600 mots différents dans les manuels de 6<sup>e</sup>, près de 6000, 7000 et 8000 mots dans les manuels de 5<sup>e</sup>, 4<sup>e</sup> et 3<sup>e</sup> (respectivement), nous sommes bien au-delà des valeurs présentées dans le Tableau 1, pour les niveaux élémentaires et « seuil » du parcours collégien (A1, A2, B1-). Au total, près de 12 000 mots différents sont recensés dans ces supports, l'équivalent d'un lexique de niveau C1, bien trop étendu pour des apprenants dans les niveaux inférieurs de compétence langagière.

Les barres claires (au milieu des histogrammes) montrent que la diversité lexicale est plus restreinte dans les supports individuels, bien sûr : avec 1700 lemmes en moyenne dans un manuel de 6<sup>e</sup>, et près de 3000 lemmes pour un manuel de 3<sup>e</sup> (entre 2000 et 2500 en moyenne pour un manuel de 5<sup>e</sup> ou de 4<sup>e</sup>).

Les barres en gris moyen à droite des histogrammes montrent que la diversité est très forte entre les manuels d'une même année. Moins de 500 lemmes sont partagés entre tous les manuels de 6<sup>e</sup>, de 5<sup>e</sup> ou de 4<sup>e</sup> (avec seulement 387 mots partagés dans les manuels de 6<sup>e</sup>), et moins de 700 lemmes en commun dans les manuels de 3<sup>e</sup>.

Au regard du nombre élevé de lemmes pour chaque niveau scolaire, il est assez surprenant de constater que 90 lemmes de la *New General Service List* manquent à cet inventaire. Certains de ces mots sont effectivement peu pertinents pour les niveaux A2 et B1- au collège (*incentive, venture, strain*), mais on s'attendrait à en trouver d'autres, dans les textes de lecture en 3<sup>e</sup>, par exemple : *collapse, elsewhere, fragment, proceed*.

Nous avons également analysé le « recyclage » des lemmes d'une année scolaire à l'autre – un principe didactique important. Le Tableau 2 résume ces informations, pour les quatre maisons d'édition (colonnes A à D). Les chiffres réfèrent au nombre de lemmes d'une année qui sont repris dans les manuels de l'année suivante. La première ligne (en gris), présente le nombre de lemmes de 6<sup>e</sup> qui sont présents dans les supports des trois années suivantes pour une même maison d'édition : près de 900 lemmes recyclés au fil des ans pour l'éditeur A, entre 720 et 760 pour les éditeurs B, C et D.

**Tableau 2 – Le nombre de lemmes recyclés au niveau suivant, selon les éditeurs**

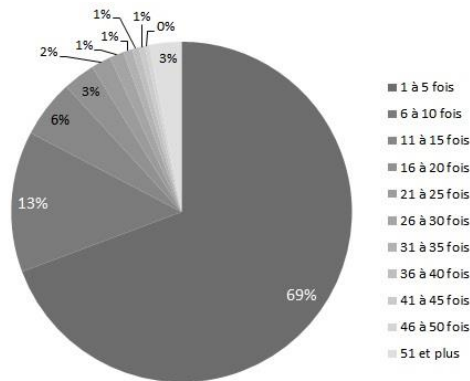
maison d'édition :	A	B	C	D
de 6e à 3e	895	728	746	759
de 6e à 5e	248	272	274	300
de 5e à 4e	520	351	372	266
de 4e à 3e	433	384	369	518

Les lignes suivantes montrent le nombre de lemmes en plus (de ces 700-900 mots en commun sur les quatre années) qui sont recyclés dans les supports de l'année suivante : de 248 à 300 lemmes en 5<sup>e</sup>, de 266 à 520 en 4<sup>e</sup> et de 369 à 518 en 3<sup>e</sup>. Ces taux varient d'un éditeur à l'autre et d'une année à l'autre, mais ils



illustrent un taux de recyclage limité. Au-delà des quelques 800 lemmes très fréquents de l'année de 6<sup>e</sup>, seulement 360 lemmes des supports d'une année (en moyenne) sont repris dans les supports de l'année qui suit.

Finalement, nous pouvons regarder la fréquence avec laquelle les lemmes apparaissent dans les supports d'un même niveau (livre + cahier). La Figure 2 présente les valeurs moyennes de reprise des mots dans le corpus, tous éditeurs confondus.



**Figure 2 – Nombre de fois que les mots apparaissent dans les supports d'un niveau, en moyenne (hors pronoms, articles et prépositions)**

Plus de deux tiers des lemmes dans les supports d'une même année sont repris moins de six fois, et 80% des lemmes apparaissent moins de 11 fois. A part un petit groupe de mots (3%) qui apparaissent très souvent (plus de 50 fois), seul 15% des mots dans les supports d'une même année sont repris plus de 10 fois.

### **Discussion et conclusions didactiques**

Ces analyses montrent que les collégiens et enseignants français travaillent avec des supports qui ne sont pas optimisés pour l'acquisition lexicale. La disparité

lexicale des manuels pose un réel problème didactique (Figure 1), car dans une classe de seconde, par exemple, seule une petite proportion des mots sera connue des élèves ayant utilisé différents manuels au collège. Ils auront du mal à communiquer entre eux, et les enseignants auront fort à faire pour mutualiser et harmoniser des connaissances lexicales si disparates.

Malgré un taux moyen de répétition des lemmes qui peut paraître élevé (Figure 2), la recherche sur l'acquisition lexicale en langue étrangère a montré qu'il faut de 7 à 20 rencontres actives avec un mot pour qu'il se fixe dans la mémoire à long terme (Webb & Nation 2017). Seul un petit tiers des lemmes d'un même ensemble de supports (livre + manuel) sera rencontré assez fréquemment pour être incorporé dans la mémoire à long terme des élèves. Nos analyses montrent également un recyclage insuffisant des mots d'une année sur l'autre, avec seulement 360 mots en moyenne recyclés d'une année sur l'autre, en plus des 800 mots qui figurent dans tous les manuels (Tableau 2). Ce chiffre suggère une progression très limitée de la compétence lexicale au fil des années de collège.

Le programme fonctionnelle et interactionnelle de la didactique européenne semble donc induire une atomisation des formes lexicales qui sont rencontrées le long du cursus en langue vivante (Nordlund 2016). En France, cet effet est probablement exacerbé par l'organisation des unités de travail autour d'« axes » thématiques abstraits (« voyages et migrations », « citoyenneté et mondes virtuels », « la confrontation et la différence », etc.), qui explique sans doute aussi le recyclage limité des lemmes.

Si le niveau B2 de compétence communicative repose sur la maîtrise réceptive et productive d'environ 8000 mots d'anglais, nous ne pouvons pas éviter la question didactique de la programmation de ces nombreux éléments linguistiques. Sur trente-cinq semaines annuelles, il y aurait plus de trente mots à apprendre chaque semaine dans la formation secondaire. Si le choix des mots et leur réintroduction au fil des leçons sont laissés au hasard, un tel volume de connaissances ne sera pas acquis. Un « programme lexical » permet non seulement d'identifier les mots à inclure à tous les manuels d'un même niveau, mais aussi leur reprise d'année en année, afin que le lexique correspondant au niveau de compétence visé soit réellement appris.

La programmation lexicale strictement limitée aux « bandes de fréquence »<sup>9</sup> peut s'avérer un peu stérile pour les concepteurs de supports (Milton 2022). Comme les didacticiens du milieu du XX<sup>e</sup> siècle (Michael West et Harold Palmer, notamment), il faut savoir combiner une systémativité de la programmation lexicale avec la créativité nécessaire pour produire des supports d'apprentissage intéressants et motivants.

Le programme lexical n'est pas incompatible avec une politique linguistique qui vise à promouvoir la compétence communicative des élèves – bien au contraire. Les connaissances lexicales sont la base même d'une communication efficace, et doivent à nouveau recevoir toute l'attention didactique qu'elles méritent. La mémorisation initiale des nouvelles associations entre les formes en L2 et leurs sens peut avoir lieu en dehors de la salle de classe, grâce à des outils numériques bien conçus. Les associations ainsi mémorisées peuvent ensuite être activement utilisées pendant les leçons – aussi bien en compréhension qu'en production et en interaction (Hilton 2019). Dans les classes de collège et lycée que nous observons actuellement en France, les enseignants attendent qu'une élève « ait besoin » d'un mot avant de le proposer comme cible d'apprentissage – mais cette approche n'est efficace ni du point de vue du temps de cours consommé (dans l'explication du sens) ou de la mémorisation de ces items, ni de leur « couverture » ou pertinence générale. Une approche aussi aléatoire ne permet pas à l'enseignant d'inclure avec certitude les mots les plus utiles dans son cursus (Milton & Benn 1932).

## **Références bibliographiques**

Andringa, Sible, et al. « Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach ». *Language Learning*, vol. 62, no 2, 2012, p. 49-78.

Arrouays, Michel. « Le Droit À l'Oubli ? » *Les Langues Modernes*, vol. 79, no 3, 1985, p. 86-89.

Browne, Charles, et al. *The New General Service List*. no 4, 2013, p. 13-16.

---

9. Les « bandes de fréquence » sont déterminées par des tranches de 500 ou 1000 mots. En lexicologie anglophone, on parle des mots de la première bande (« 1K words »), de la deuxième bande (« 2K »), etc. – c'est-à-dire, les mille premiers mots, les mille mots suivants, etc.

- Coste, Daniel, et al. *Un niveau-seuil: systèmes d'apprentissage des langues vivantes par les adultes*. Hatier, 1976.
- Debyser, Francis. « De l'imparfait du subjonctif aux méthodes communicatives: Où en est l'enseignement des langues vivantes? » *Teaching Modern Languages*, Routledge, 2002, p. 56-61.
- Dehaene, Stanislas. *Apprendre à lire: des sciences cognitives à la salle de classe*. Odile Jacob, 2011.
- Faucett, Lawrence William, et al. *Interim Report on Vocabulary Selection for the Teaching of English as a Foreign Language*. King and Son, 1936.
- Garside, Roger. « A hybrid grammatical tagger: CLAWS 4 ». *Corpus annotation: Linguistic information from computer text corpora*, 1997, p. 102-21.
- Goodenough, Florence L. « Edward Lee Thorndike: 1874-1949 ». *The American Journal of Psychology*, vol. 63, no 2, 1950, p. 291-301.
- Gougenheim, Georges. *L'élaboration du français fondamental (1er degré): étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Didier, 1964.
- Hazenberg, Suzanne, et Jan H. Hulstun. « Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation ». *Applied linguistics*, vol. 17, no 2, 1996, p. 145-63.
- Hilton, Heather E. « Oral fluency and spoken proficiency: ideas for testing and research ». *Measuring L2 Proficiency: perspectives from SLA*, édité par Pascale Leclercq et al., Multilingual Matters, 2014, p. 27-53.
- Hilton, Heather E. *Sciences cognitives et didactique des langues, Rapport d'expertise*. CNESCO, 2019.
- Komenski, Jan Amos Comenius. *Janua Linguarum Reserata*. 1631.
- Komenski, Jan Amos Comenius. *Orbus Sensualium Pictus*. Michael Endter, 1658.
- Lee, Joanne. « Size matters: Early vocabulary as a predictor of language and literacy competence ». *Applied Psycholinguistics*, vol. 32, no 1, 2011, p. 69-92.
- Meara, Paul. « Vocabulary acquisition: A neglected aspect of language learning ». *Language Teaching*, vol. 13, no 3-4, 1980, p. 221-46.

Milton, H., et Vincent T. Benn. « Study of the vocabulary of thirty first-year French courses ». *Modern Languages*, vol. 14, no 1-3, 1933, p. 11-17.

Milton, James. *Communication personnelle*. juin 2021.

Ministère de l'éducation nationale. *Programmes pour le cycle 2, 3, 4*. 2015.

Nation, Paul. « How large a vocabulary is needed for reading and listening? » *Canadian modern language review*, vol. 63, no 1, 2006, p. 59-82.

Nation, Paul. *Learning vocabulary in another language*. Cambridge University Press, 2001.

Nordlund, Marie. « EFL textbooks for young learners: A comparative analysis of vocabulary ». *Education Inquiry*, vol. 7, no 1, 2016, p. 47-68.

Smith, Richard C. *Teaching English as a Foreign Language, 1912-1936: Lawrence Faucett*. Taylor & Francis, 2003.

Swan, Michael. « A critical look at the communicative approach (1) ». *ELT journal*, vol. 39, no 1, 1985, p. 2-12.

Thorndike, Edward Lee. *A Teacher's Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People*. Teachers College Press, 1921.

Thorndike, Edward Lee, et Irving Lorge. *The teacher's word book of 30,000 words*. Bureau of Publications, Teachers College Press, 1944.

Toutanova, Kristina, et al. « Feature-rich part-of-speech tagging with a cyclic dependency network ». *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 2003, p. 252-59.

Van Zeeland, Hilde, et Norbert Schmitt. « Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? » *Applied Linguistics*, vol. 34, no 4, 2013, p. 457-79.

Webb, Stuart, et Paul Nation. *How vocabulary is learned*. Oxford University Press, 2017.

West, Michael. *A general service list of English words*. Longmans Green and Co. Ltd., 1965.

Zhang, Songshan, et Xian Zhang. « The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis ». *Language Teaching Research*, SAGE Publications Sage UK: Londres, 2020, p. 1-30.