

Introduction des corpus numériques en classes de langues pour une utilisation effective du lexique dans les productions écrites chez de jeunes apprenants d'anglais non-natifs

Introduction of digital corpora in language classes for effective use of the lexicon in written productions by young non-native English learners

Dalia BELLIK
Université de Savoie Mont Blanc, Chambéry, France
Dalia.bellik@hotmail.com

Reçu le: 21/10/2020, **Accepté le:** 20/12/2020, **Publié le:** 05/03/2021

Résumé

L'ère du numérique, dans laquelle nous vivons, semble avoir apporté de nouvelles perspectives à travers les nombreux moyens technologiques dont nous jouissons aujourd'hui, ainsi l'exploration de ces outils en classes de langues est devenue inévitable pour améliorer l'enseignement de la langue seconde.

Les corpus numériques, appelés également concordanciers permettent aux usagers, enseignants comme apprenants, d'exploiter des données textuelles, à savoir le corpus linguistique stocké dans le logiciel. La particularité du concordancier réside dans sa capacité à fournir de multiples concordances, offrant ainsi plus de facilité dans la recherche des significations de mots et de leurs fonctions, où le mot est toujours exposé dans divers contextes.

La présente recherche va tenter de démontrer si le concordancier peut avoir un impact concret et positif sur les écrits académiques scolaires et s'il fournit un appui dans la recherche de solutions aux difficultés liées au lexique, rencontrées pendant une production écrite chez des apprenants d'anglais non-natifs de niveau novice, dans un milieu académique pré-universitaire, à savoir le collège.

Mots clés : Concordanciers - corpus linguistiques - rédaction académique - lexique - apprenants non-natifs.

Abstarct

The digital era, in which we live, seems to have brought new teaching and learning perspectives thanks to the technological means that surround us today, therefore, exploring these tools in language classes became something unavoidable in order to improve the second language teaching process.

Digital corpora, also called concordancers, allow users, teachers and learners alike, to exploit textual data, namely the linguistic corpus stored in the software. The specificity of the concordancer lies in its ability to provide multiple concordances, thus making it easier

to find the meanings of words and their functions, where the word is always exposed in various contexts.

This research will attempt to demonstrate whether concordancers can have a concrete and positive impact on learner's academic writings and if it provides support in the search for solutions to the lexical difficulties encountered during a written production of non-native, novice learners at a pre-university academic environment, middle school.

Keywords: Concordancers - linguistic corpus - academic writing - lexicon - non-native learners.

Introduction

L'apprentissage du lexique est un élément essentiel dans l'enseignement et l'apprentissage de la langue cible. La maîtrise d'un large lexique permet de découvrir de nouveaux concepts et favorise la construction et l'expression de nouvelles idées. Nombreux sont les chercheurs qui ont mis l'accent sur l'apport positif qu'entraîne une bonne maîtrise des connaissances lexicales sur l'acquisition de la langue cible en général et la production de l'écrit en particulier. Nation & Laufer (1998). expliquent qu'il est important pour les apprenants non-natifs d'avoir un large vocabulaire qui est considéré un élément majeur dans le processus rédactionnel. Une utilisation pertinente du lexique est d'autant plus importante car une bonne production écrite est souvent celle qui fait bon usage du vocabulaire. Ainsi, la qualité d'une production écrite est souvent liée à la richesse ainsi qu'à la pertinence du vocabulaire employé.

Les connaissances lexicales sont d'une part un moyen indispensable pour parvenir à une communication effective dans la langue cible (Alqahtani, 2015 : 22) et d'autre part, jouent un rôle crucial dans l'acquisition des compétences langagières, écrites et orales. La relation qui existe entre la connaissance lexicale et la langue est donc une relation complémentaire ; la première facilite l'utilisation de la langue alors que la seconde favorise la maîtrise du vocabulaire (Nation 2011, Alqahtani 2015 : 22). Le lexique est un domaine de connaissance de la langue parmi d'autres qui joue un rôle important dans l'acquisition de la langue chez les apprenants de la langue seconde (Ibid, 2015 : 22). C'est pourquoi, un enseignement effectif de la langue dépend des instruments, techniques et méthodes d'enseignement utilisés pour satisfaire au mieux les besoins d'apprentissages des apprenants en termes de lexique.

En revanche, l'enseignement des connaissances lexicales est en général difficile à réaliser dans la langue cible car leur acquisition nécessite beaucoup de temps et d'efforts d'attention en raison des différentes significations que peut avoir un mot. De plus, il existe des contextes divers et variés, parfois confus, dans lequel un mot peut être employé, si bien que plusieurs apprenants consacrent beaucoup de leur temps à mémoriser des listes de mots en s'appuyant sur des

dictionnaires, une méthode souvent considérée comme étant nécessaire dans l'acquisition de la langue seconde (Ibid, 2015 : 23).

Il est vrai que plusieurs chercheurs ont non seulement attiré l'attention sur les difficultés rencontrées face à l'apprentissage du lexique mais encore sur la courte période consacrée à cet apprentissage et à la langue anglaise en général. Cobb (1997) rapporte qu'apprendre un mot à partir d'une définition courte ou d'un seul contexte de phrase tend à produire une connaissance lexicale inerte qui ne facilite pas la compréhension de ce mot dans un contexte nouveau, tandis que, l'apprentissage d'un mot dans plusieurs contextes, avec ou sans définition, tend à produire des connaissances riches, transférables et réutilisables. Il expose dès lors deux stratégies distinctes qui visent à améliorer la maîtrise du lexique chez les apprenants d'anglais dans un milieu académique ; scolaire.

L'apprentissage du lexique à travers des listes (breadth learning)

Cobb explique que bien que cela soit le moyen le plus rapide pour apprendre le lexique en un court laps de temps, il n'en demeure pas moins qu'il y ait de nombreux désavantages. Apprendre le lexique à travers une liste crée une connaissance superficielle, où les apprenants se sentent contraints d'utiliser des dictionnaires afin de créer du sens, cela leur permet évidemment de constituer leurs propres glossaires mais ces derniers demeurent fragiles car ils sont basés sur de simples traductions de mot à mot avec une absence totale de contexte. Ceci ne permet donc pas aux apprenants d'améliorer leur compréhension et de lecture des textes même si les mots qu'ils retrouvent dans ces textes ont déjà été appris

L'apprentissage du lexique à travers la lecture approfondie (depth learning), cet apprentissage approfondi du lexique opère à travers des lectures intensives, où l'implicite prime sur l'explicite. Cela signifie que les apprenants sont confrontés à des contextes plus riches et doivent ainsi se munir de leur esprit analytique afin de repérer le sens et l'usage des mots. Ceci n'est pas toujours facile à faire car souvent ces lectures sont chronophages et les apprenants ne peuvent pas analyser tous les mots qu'ils repèrent. Le choix de cette stratégie pour l'enseignement du lexique représente un vrai défi aux enseignants de langues en raison du temps restreint qu'il leur ait accordé, un défi souvent difficile voire impossible à relever. Il se trouve aussi que l'apprentissage du vocabulaire qui prend place dans des contextes naturels tels que la lecture, est plus efficace car celui-ci apporte une compréhension complète du mot et de son usage (Cobb 1999). L'acquisition d'un lexique étendu, bien structuré et grandement inter relié avec des références croisées dans une langue seconde ne peut avoir lieu que si les mots sont vus dans divers contextes naturels, sur une longue période de temps, comme pendant l'enfance. (Ibid, 1999). Ceci nous mène donc à repenser la

manière d'enseigner le lexique pour une utilisation plus appropriée dans les productions écrites des apprenants d'anglais non-natifs.

L'exploration des outils concordanciers peut nous conduire vers de nouvelles alternatives d'enseignement du lexique et de la rédaction académique. Mais encore à mieux identifier leurs apports et leur efficacité. L'ère du numérique met aujourd'hui à notre disposition de multiples outils technologiques, appelés également les technologies de l'information et de la communication pour l'enseignement 'TICE' afin de développer de nouvelles stratégies et de techniques d'enseignement pour renforcer les compétences linguistiques des apprenants d'anglais non-natifs. Les concordanciers par exemple sont des outils qui ont déjà fait leurs preuves dans des milieux scolaires universitaires et ce depuis les années soixante, il serait donc le moment méritent de les davantage dans différents contextes d'enseignement, notamment dans l'enseignement secondaire avant de trancher sur la question de leur efficacité ou de leur défaillance. L'exploration des corpus concordanciers dans l'enseignement secondaire nous permet d'affirmer ou de contester l'apport positif de ces outils dans la maîtrise du lexique et la rédaction académique chez des apprenants d'anglais langue seconde.

Les corpus concordanciers sont alors l'axe majeur de cet article afin d'essayer de répondre aux mieux à certaines interrogations énoncées dans ce travail de recherche qui s'inscrit dans le cadre d'une thèse doctorale. Il est essentiel à ce stade d'expliquer les principaux concepts de ce travail de recherche, à savoir, ce qu'est un corpus linguistique ? ce qu'est un concordancier ? et quelle est la relation entre les deux ?

D'abord, un corpus linguistique comprend de larges collections de discours et d'écrits authentiques, ce dernier est construit autour d'un certain nombre de critères spécifiques et présentés sous format électronique. Wynne (2015) explique qu'un corpus linguistique est un ensemble de textes sélectionné et rassemblé pour que la langue puisse être étudiée sur ordinateur. Tout texte peut être considéré tel un corpus : les journaux, les publications Facebook, les recettes de cuisine, les romans, les discours, les scripts, les chats entre amis, les lettres, les livres, les magazines, les examens et les memos ...etc. Une multitude de corpus sont disponibles aujourd'hui, du corpus monolingue au corpus plurilingue avec des textes originaux et traduits. On retrouve d'une part de petits corpus de quelques dizaines de milliers de mots hautement spécialisés et d'autre part, de larges corpus de référence avec des centaines de millions voire billions de mots. Certains corpus contiennent des textes uniquement ; simples ou annotés. D'autres corpus multimodaux sont plus sophistiqués, équipés de sons et de vidéos. (Iénko-Szymanska & Boulton, 2015). Un corpus est constitué de plusieurs textes, extraits de différentes sources ; romans, magazines ou journaux ...etc. (Sinclair 2004, cité dans Wynne 2015) rapporte que les corpus n'ont pas toujours la même

orientation, ils sont conçus pour divers objectifs. Les corpus linguistiques destinés à la didactique des langues cherchent en général à repérer quels schémas sont associés aux éléments lexicaux et grammaticaux. Ils permettent de répondre à un nombre de questions liées à l'utilisation de la langue, à savoir, quels sont les mots et les phrases les plus fréquemment employés en anglais ? quelles sont les différences entre l'anglais écrit et l'anglais parlé ? Quels sont les temps les plus utilisés en anglais ? quelles prépositions suivent un verbe particulier ? comment utilise-t-on des mots comme *can*, *may* et *might* ? quels mots sont utilisés dans des situations formelles et ceux utilisés dans des situations informelles ? à quelle fréquence utilise-t-on les expressions idiomatiques ? et bien d'autres questions d'ordre langagier (Bennett, 2010). La linguistique de corpus est donc perçue comme étant une méthodologie dont le but n'est pas d'interpréter des occurrences linguistiques uniques, elle vise plutôt à faire des découvertes sur une langue en général (Teubert, 2009). Parmi les corpus les plus célèbres, il y a le British National Corpus (BNC) qui comprend 100 millions de mots, le Corpus of Contemporary American English (COCA), lancé en ligne en 2008 avec un corpus large et varié d'anglais américain contenant plus de 400 millions de mots collectés depuis 1990 jusqu'en 2010 (Davies, 2009) et regroupant une grande variété de genres de textes ; soit, l'anglais parlé 'spoken English', la fiction, les magazines populaires, les journaux et les revues académiques. Et encore l'American National Corpus (ANC) créé en 1990 avec 20 millions de mots (Ibid, 2009).

La linguistique de corpus s'est vue parallèlement évoluer avec le développement accru des nouvelles technologies. Il est tout à fait possible d'accéder aujourd'hui à des données linguistiques considérables, stockées électroniquement. L'accès à ces données est devenu une pratique ordinaire et commode pour faciliter l'enseignement des langues étrangères et ce grâce à des interfaces de logiciels ergonomiques. Teubert (2009) affirme que la linguistique de corpus a été développée pour extraire d'un corpus les connaissances linguistiques nécessaires à l'enseignement des langues. En outre, la linguistique de corpus est venue dans l'objectif d'améliorer l'enseignement de l'anglais comme langue étrangère à défaut des dictionnaires traditionnels, qui exposent les mots, détachés de leurs contextes, si bien que les apprenants se retrouvent incapables de les employer adéquatement : « *Les items lexicaux acquièrent un sens non pas de manière isolée mais seulement pris dans leur contexte* » (Ibid ; 2009). Ceci, signifie que le sens des mots dépend amplement des collocations et des unités lexicales et grammaticales qui les constituent. Les corpus offrent un enchaînement de mots dans divers contextes donnés afin d'aider les usagers à créer du sens. Boulton & Tyne (2014) affirment que le recours massif aux données en langue cible sous forme de documents authentiques s'est produit dans un climat de changement méthodologique avec la mise en place de l'approche communicative mais aussi un climat de développement technologique et médiatique. Les corpus laissent aux

apprenants l'opportunité d'avoir un contact direct avec le même langage authentique auxquels sont exposés les natifs de la langue cible. Ils expliquent aussi que le contexte actuel diffère de celui qui existait avant ; les apprenants ont aujourd'hui accès à une multitude de données via internet et y seront davantage exposés qu'il y a des décennies.

Par ailleurs, les concordanciers sont des logiciels informatisés, surnommés tantôt "concordanciers", tantôt "générateurs de concordances" et parfois "logiciels de concordances". Les concordanciers modernes sont des outils technologiques qui recueillent de larges collections de textes, pouvant être soit des outils spécialisés ou des outils généralisés. En d'autres mots, ils sont conçus sur mesure pour des fins purement pédagogiques ; d'enseignement linguistique ou pour répondre à un usage général (Colpaert, 2006).

John Sinclair, le père de la linguistique de corpus moderne avec le projet COBUILD fut le premier dictionnaire qui s'appuie entièrement sur une base de corpus. Il fut déterminant car il a réussi à démontrer le potentiel des logiciels concordanciers dans l'enseignement de la langue anglaise. Tim Johns est également l'un des pionniers à avoir incité la mise en pratique de ce type de logiciels dans l'apprentissage de la rédaction académique par le biais de la célèbre méthode "Data Driven Learning". Mike Scott est un autre pionnier qui a coécrit avec Tim Johns, il est le créateur du Micro Concord.

La 'Data Driven Learning' ou bien DDL (Johns, 1988) est une approche qui permet aux apprenants d'explorer les régularités et les structures de la langue à travers l'utilisation d'outils informatiques, à savoir, les concordanciers dans ce cas de figure. Les apprenants peuvent se focaliser sur tous les éléments communs dans les occurrences des mots de façon à tirer des généralisations. À titre d'exemple, il est tout à fait possible pour l'utilisateur ; l'apprenant, de réaliser des requêtes sur la machine et d'interroger la base de données qui, elle, permet de visualiser l'utilisation d'un mot dans différents énoncés et dans divers contextes. Ainsi, l'utilisateur obtient des informations sémantiques grâce à un traitement cognitif des structures linguistiques de la langue. En d'autres mots, ce sont les réflexions sur les contextes, fournies par les lignes de concordances qui permettent d'identifier des effets de sens. Nation appuie la démarche d'utilisation des concordanciers dans le processus d'apprentissage et lui accorde cette définition : "*Un concordancier est un programme où on saisit un mot, une phrase où plusieurs mots et le programme lance une recherche d'exemples de ces mots dans une collection de texte(s) ; le corpus*" (2014 : 15). Il souligne également qu'il existe deux types de concordanciers ; les plus simples à utiliser sont les concordanciers en ligne, où il suffit d'aller sur le site web pour utiliser le programme et le corpus proposé par ce site web. Et puis, il y a les concordanciers

téléchargeables sur ordinateurs et dont l'utilisation n'exige aucun accès direct à internet. En général, le concordancier propose une interface de concordance appelé KWIC "Key word in context". Le mot clé apparaît dans une suite d'énoncés, de haut en bas, le mot clé se situe toujours au milieu, entouré d'un contexte. "KWIC est la convention la plus utilisée dans les programmes de concordance, précisément, car ce format facilite l'analyse du contexte direct du mot clé" (Tribble & Jones 1997:5)

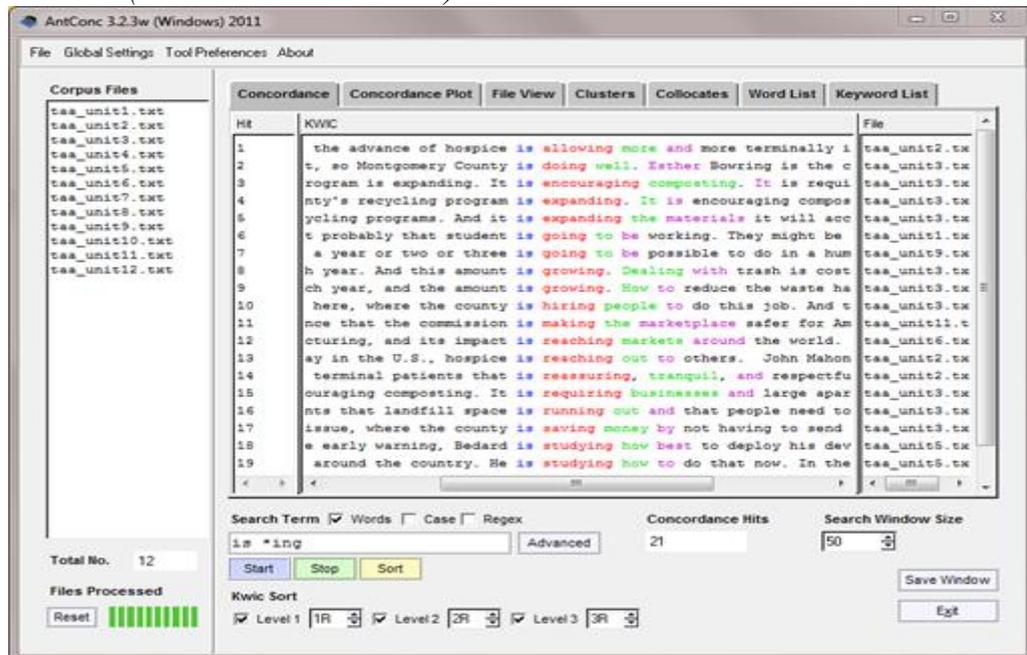


Figure 01 : une capture d'écran du concordancier AntConc de Laurence Anthony. <http://www.laurenceanthony.net/software/antconc/screenshots.html>

Par conséquent, un concordancier est un système pleinement opérationnel, pouvant répondre à des besoins éducatifs. Cobb (1997) affirme que ces outils ont la capacité de résoudre le paradoxe qui existe entre les deux stratégies d'enseignement du lexique « étendu du vocabulaire Vs profondeur du vocabulaire ». Ce qui paraissait être un problème insoluble avec les méthodes d'enseignement traditionnelles, semble désormais possible grâce aux technologies numériques. Le concordancier offre la possibilité de visualiser un mot dans plusieurs contextes, l'apprenant quant à lui analyse comment le mot est employé dans ces différents énoncés. L'apprentissage des mots dans leur contexte à travers de courtes lignes, prélevées d'un texte original a l'avantage de faciliter le repérage du mot cible initial ainsi que d'autres nouveaux mots rencontrés au sein de ce même texte. L'exposition des apprenants à plusieurs contextes est considérée comme une vraie opportunité pour rencontrer divers exemples de concordances,

ceci viendrait accroître la qualité d'apprentissage du lexique et de son utilisation. Ce qui n'est pas toujours réalisable avec les longues lectures de textes.

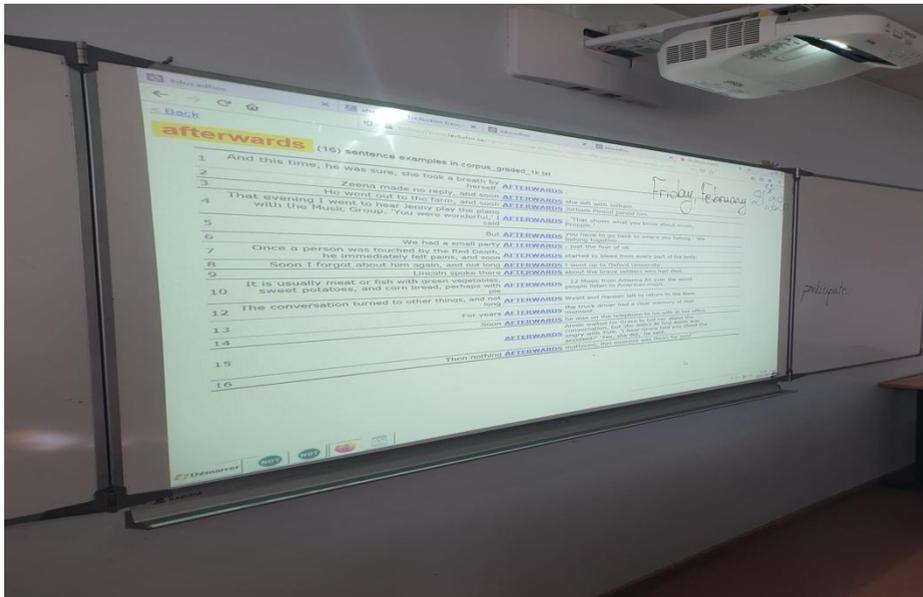


Figure 02 : exemple d'utilisation du concordancier en classe de 4^{ème} pour faire deviner la signification du mot 'afterwards'.

Comme le montre la figure 02, les apprenants ont eu plusieurs fois à interagir avec le concordancier 'Compleat Lexical Tutor' en classe. Dans cet exemple, les élèves ont rencontré pour la première fois le mot 'afterwards'. Au lieu de leur apporter directement la signification du mot, une liste de concordances leur a été projetée afin de susciter leur curiosité et les inciter à faire un travail réflexif et analytique pour trouver le sens du mot de manière tout à fait autonome. Les apprenants ont alors compris le mode de fonctionnement de l'outil concordancier, ce qui a préparé le terrain de l'expérimentation de cet outil par la suite.

Le projet de recherche exposé dans cet article, vise en premier lieu à mettre en pratique une approche d'enseignement sur corpus, communément connue sous le nom de 'DDL', Data Driven learning, avec un groupe d'apprenants d'anglais non-natifs d'un niveau scolaire secondaire ; plus précisément des collégiens. Cette approche s'appuie sur l'exploitation des corpus linguistiques numériques ; concordanciers, dans l'objectif d'améliorer le processus d'enseignement du lexique mais surtout d'améliorer la qualité rédactionnelle en

langue seconde. Une première tentative d'expérimentation a débuté en Janvier 2020 au collège Georges Charpak de Brindas. La collecte de données a été interrompue et le travail de recherche n'a malheureusement pas abouti à cause de la fermeture des écoles suite à la pandémie SARS CoV 2. Une nouvelle expérimentation a vu le jour depuis Septembre 2020 dans un nouvel établissement, le collège Pierre Brossette à Oullins, le même processus de collecte de données a été établi que lors de la première expérimentation.

Le nouveau corpus est composé de 3 classes de 5^{ème}, chaque classe compte entre 26 à 27 élèves. A la fin de chaque séquence pédagogique, les apprenants réalisent deux productions écrites. Les élèves travailleront sur trois séquences pédagogiques différentes, chaque séquence dure entre 8 à 9 séances d'une heure. Et réaliseront 6 travaux rédactionnels au total tout au long de la période d'expérimentation. Les travaux rédactionnels seront collectés et analysés ainsi : La première phase (Rédaction), celle-ci correspond au moment où les apprenants réalisent une production écrite sur le thème abordé en classe après avoir clôturer la séquence pédagogique qui porte sur la même thématique. Ensuite, la deuxième phase (Remédiation) représente le moment de la collecte des textes écrits par les apprenants. Ces textes sont analysés et vont permettre à l'enseignant de créer une typologie des erreurs lexicales identifiées, suivie d'une conception d'une série d'activités sur corpus qui visent à retravailler les erreurs lexicales récurrentes dans les productions écrites des apprenants. Seulement deux classes vont faire un travail de remédiation avec des exercices sur corpus afin d'aider au mieux les apprenants à repérer et à analyser des schémas linguistiques et surtout pour les encourager à trouver par eux-mêmes des réponses à leurs interrogations. Tandis qu'une troisième classe fera un travail de remédiation traditionnel ; les erreurs lexicales seront donc corrigées à l'oral. Il y aura évidemment une rotation dans le type de remédiation utilisé avec chaque classe. En dernier, vient la phase post-remédiation où l'enseignant va comparer la première et la seconde rédaction après le travail de remédiation. Il va également analyser si les rédactions qui ont suivies l'utilisation des activités sur corpus sont plus pertinentes que celles qui n'ont utilisé qu'une simple correction traditionnelle des erreurs lexicales.

Bien que les apprenants aient acquis un vocabulaire ciblé tout au long de la séquence pédagogique, ils ont souvent beaucoup de difficultés à les employer de manière adéquate lors des productions écrites, d'où le choix d'introduire l'outil concordancier « Compleat Lexical Tutor » dans le processus d'enseignement du lexique pour remédier aux lacunes et aux difficultés auxquelles ils sont confrontés mais surtout pour les aider à améliorer la qualité de leurs travaux rédactionnels. Le concordancier numérique « Compleat Lexical Tutor » a été choisi pour deux raisons principales. D'abord, son grand avantage est lié au fait qu'il contient

plusieurs corpus linguistiques, avec différents niveaux linguistiques et degrés de complexité. Ensuite, parce qu'il a de larges collections de textes authentiques, faciles à didactiser.

En effet, pour un public d'apprenants de niveau pré-intermédiaire, comme celui de cette étude de recherche, il est beaucoup plus commode d'explorer un petit corpus de quelques milliers de mots, qui propose une gamme de vocabulaire accessible, qui répond aux besoins d'apprentissage des apprenants et qui convient surtout à leur niveau de connaissances linguistiques. Le 'Graded Readers 1K corpus' et le 'Graded Readers 2K corpus' sont les deux corpus choisis parmi tous les corpus proposés par 'Compleat Lexical Tutor' afin de concevoir les activités sur corpus; le premier contient (530 K) et le second contient (920 K).

Le logiciel 'Compleat Lexical Tutor' de par son interface permet également de construire différentes activités ; des Quiz, des exercices à trous ... etc mais aussi de les imprimer. L'exploitation de cet outil par l'enseignant permet de réaliser des requêtes et de générer plusieurs exemples de concordances authentiques ; basés sur les erreurs lexicales des apprenants afin de créer des activités sur corpus. L'objectif de ces activités vise à aider les apprenants à porter une analyse approfondie sur différents contextes linguistiques, à renforcer la détection des erreurs lexicales ainsi qu'à déduire des conclusions sur l'utilisation adéquate d'un certain vocabulaire dans une production écrite, et ce de manière tout à fait autonome.

Pour la première collecte de données, les élèves ont travaillé sur la séquence « Life of a middle schooler » avec un lexique ciblé qui porte sur l'école, les règles de la vie, les sentiments et les goûts, la description physique, le métalangage de la littérature et de la bande dessinée, les couleurs et les vêtements. La tâche finale, viendra à la fin de la séquence pour démontrer comment les apprenants utilisent ce lexique dans une expression écrite. Ils sont encouragés à faire une production écrite en décrivant une collégienne, présentée dans un poster avec un certain nombre d'informations apporté en plus de l'image pour assurer une description riche et complète avec un maximum de vocabulaire possible. Après avoir analysé les travaux de rédaction des apprenants, une typologie d'erreurs est élaborée afin de mieux visualiser les lacunes et d'analyser la nature des erreurs lexicales des apprenants. Voici une typologie adaptée du travail de recherche de Ander and Yıldırım (2010), réalisée lors de la première expérimentation inaboutie pour faciliter le travail de conception des activités sur corpus :

Catégories d'erreurs lexicales	Fréquence	Exemples
--------------------------------	-----------	----------

<p>Erreurs du choix de mot.</p> <p>(Errors of wrong word choice).</p>	8	<p><u>Say</u> to him, au lieu de “tell”</p> <p>Canada <u>‘is’</u> different people, au lieu de “has”</p> <p>National parks <u>‘is’</u> two red chairs, au lieu de “has”</p> <p><u>‘There’</u> are bilingual, au lieu de “they”</p> <p>Going to Canada is so <u>‘pretty’</u>, au lieu de “amazing”</p> <p>You <u>‘have’</u> travel, au lieu de “can”</p> <p>You <u>‘want’</u> go to, au lieu de “can”.</p> <p>The Niagara Fall is a beautiful <u>‘landmark’</u>, au lieu de “landscape”</p>
<p>Erreurs de traduction littérale</p> <p>(Errors of literal translation)</p>	4	<p><u>‘Language official’</u>, au lieu de “official language”</p> <p><u>‘Savage animals’</u>, au lieu de “wild animals”</p> <p><u>‘Big sports’</u>, au lieu de “important sports”</p> <p><u>‘Animals special’</u>, au lieu de “special animals”.</p>
<p>Erreurs d’orthographe</p> <p>(Misspelling)</p>	41	<p><u>‘Sirop’</u>, au lieu de ‘syrup’ / <u>‘Meaple’</u>, au lieu de ‘maple’.</p> <p><u>‘Bufalow’</u>, au lieu de ‘Boffalo’ / <u>‘Otawa’</u>, au lieu de ‘Ottawa’</p> <p><u>‘whals’</u>, eu lieu de ‘wales’ / <u>‘tipical’</u>, au lieu de ‘typical’</p> <p><u>‘hall’</u>, au lieu de ‘all’ / <u>‘canadiens’</u>, au lieu de ‘Canadians’.</p>
<p>Errors of redundancy:</p>	2	<p>We <u>‘would can’</u>, visit au lieu de ‘can’.</p> <p><u>‘A maple tree leaf’</u>, au lieu de ‘a maple tree’ ou bien ‘a</p>

		maple leaf’.
Errors of word formation:	4	I am ‘ <u>interesting</u> ’, for au lieu de ‘interested’. Come with ‘ <u>my</u> ’, au lieu de ‘me’ The population is ‘ <u>friendship</u> ’, au lieu de ‘friendly’ ‘ <u>Popular</u> ’ food, au lieu de ‘popular’.

Tableau1 : Exemple de typologie des erreurs lexicales d’une production écrite.

Le même tableau présenté ci-dessus sera utilisé pour analyser les erreurs lexicales des Apprenants au cours de cette nouvelle expérimentation. Celui-ci donne un aperçu général des catégories d’erreurs lexicales Identifiées, leur fréquence mais bien plus important des exemples des erreurs lexicales faites par les apprenants. Ainsi, afin de remédier aux erreurs lexicales les plus fréquentes dans les productions écrites des apprenants, des activités sur corpus, inspirées des activités « Hands on concordancing » développées par Cobb (1997) seront conçues. Les apprenants sont encouragés à faire un travail réflexif sur la langue sans que l’enseignant n’apporte la moindre réponse ou explication directe quant aux erreurs effectuées, comme c’est le cas dans l’enseignement traditionnel.

Activité 7 :

1. Lis les phrases suivantes ensuite, choisis le sens approprié du mot « LANDMARK » :

Colorado's Grand Canyon, probably the most famous LANDMARK of the United States

National LANDMARK that receives our half a million recreational users and tourists each year.

It was an old and dirty wooden structure, sadly in need of repair but it was a LANDMARK

Deciding the fate of a 19th century building, a LANDMARK in the community.

Je déduis :

Landscape is :

1. The view that you can see when you look across an area of land.
2. It is an important building that symbolizes a given country.
3. It is the Financial capital city of Deutschland.

Figure 03 : Un exemple d'une activité sur corpus conçue pour la première expérimentation.

Activité 8 :

2. Lis les phrases suivantes ensuite, choisis le sens approprié du mot « **LANDSCAPE** » :

He loves nature and especially adores the English **LANDSCAPE**

Destroy the God made beauty of the **LANDSCAPE**

Beauty and serenity of the natural **LANDSCAPE**

Washington's **LANDSCAPE** is filling up with memorials.

Their house in the Scottish lowlands looks out over a calm, domestic **LANDSCAPE** of fields and gentle hills

Je déduis :

Landscape is :

1. The view that you can see when you look across an area of land.
2. It is an important building that symbolizes a given country.
3. It is the Financial capital city of Deutschland.

Figure 04 : Un exemple d'activité sur corpus conçue pour la première expérimentation.

Les figures 03 et 04 montrent des exemples d'activités sur corpus, conçues pour le travail de remédiation. Le but étant de montrer la forme et l'organisation de ces activités qui sont proposées aux apprenants pour revoir les erreurs lexicales de type « erreur du choix de mot » comme pour l'utilisation du mot 'landscape' où les élèves ont tendance à le confondre avec le mot 'landmark'. Cette activité vise donc à proposer plusieurs exemples de concordances pour aider les apprenants à se familiariser avec les différents contextes d'utilisation du mot 'landscape' pour se l'approprier et parvenir à les employer de manière effective dans d'autres productions écrites à l'avenir.

Pour conclure, il est impératif de préciser que cet article n'a pas pour objectif de présenter des résultats, car la présente recherche n'est toujours pas aboutie. La phase d'expérimentation est toujours en cours, une première collecte de données fut réalisée et les données textuelles sont actuellement en cours d'analyse. Deux autres collectes de données sont programmées d'ici début Avril, des résultats préliminaires seront donc prochainement disponibles. Toutefois, le

but principal but de cet article vise à présenter une approche d'enseignement, longtemps délaissée, en milieu scolaire pré-universitaire, d'évoquer les nombreux attributs qui sont données à la méthode d'enseignement sur corpus et à l'utilisation des concordanciers en général. Aussi, il paraissait important d'exposer des exemples d'exercices sur corpus facilement utilisables en classe de langue pour l'enseignement du lexique mais aussi de mettre en avant la méthodologie appliquée dans ce travail de recherche pour montrer le potentiel d'utilisation des outils concordanciers dans l'acquisition du lexique et l'amélioration des compétences rédactionnelles en anglais chez de jeunes apprenants d'anglais non-natifs.

Bibliographie :

AlqahtaniI, M. (2015). The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education*, Vol. III(3), pp. 21-34. , DOI: 10.20472/TE.2015.3.3.002

Bennett, G. R. (2010). *Using Corpora in the Language Learning Classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press.

Boulton, Alex et Henry Tyne. 2014. *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*. Paris : Didier Coll. Langues et didactique, 309 pages. ISBN : 978-2-278-07615-4.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25 (3), 301-315.

Cobb, T. (1999a). [Breadth and depth of vocabulary acquisition with hands-on concordancing](#). *Computer Assisted Language Learning* 12, 345 - 360.

Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal*, 23, 477-497. Retrieved from <https://calico.org/page.php?id=5>

Davies, Mark. (2008) The Corpus of Contemporary American English (COCA): 520 million words, 1990-present. Available online at <http://corpus.byu.edu/coca/>.

Johns, T. (1993), 'Data-driven learning: an update.' *TELL&CALL*, 2: 4-10

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307-322. doi:10.1093/applin/16.3.307

Leńko-Szymańska, A., & Boulton, A. (2015). Multiple Affordances of Language Corpora for Data-driven Learning. *International Journal of Corpus Linguistics*. 20. 560-569. 10.1075/ijcl.20.4.07gar.

Martin Wynne (ed.). Oxford: Oxbow Books for the Arts and Humanities Data Service, 2005

. ISBN: 1-84217-205-0. £ 15 (paperback). Online
: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora>

Nation, P. (2014). *What you need to know to learn a foreign language*. Available free from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>

Teubert, W. (2009). La linguistique de corpus : une alternative [version abrégée]. *Semen*. 185-211. 10.4000/semen.8914.

Tribble, C., & Jones, G. (1997). *Concordances in the classroom: Using corpora in language education*. Houston TX: Athelstan.

