



HAL
open science

Outils pour la correction de l'écrit dans le secondaire : à la découverte du corpus

Laura-May Simard

► **To cite this version:**

Laura-May Simard. Outils pour la correction de l'écrit dans le secondaire : à la découverte du corpus. Education. 2018. dumas-01977897

HAL Id: dumas-01977897

<https://dumas.ccsd.cnrs.fr/dumas-01977897>

Submitted on 11 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année universitaire 2017-2018

Master MEEF

Mention 2nd degré – parcours « anglais »

2^{ème} année

Outils pour la correction de l'écrit dans le secondaire : à la découverte du corpus.

**Mots Clefs : didactique des langues, enseignement secondaire, correction de l'écrit,
apprentissage sur corpus (ASC), techniques d'information et de communication pour l'enseignement
(TICE)**

Présenté par : Laura-May Simard

Encadré par : Cédric Sarré

SOMMAIRE

1. Introduction.....	1
1.1. Contexte	1
1.2. Principe de l'étude et questions de recherche	1
2. Etat de l'art.....	3
2.1. L'erreur et sa correction	3
2.1.1. Définitions et typologies variables	3
2.1.2. Utilisation des erreurs pour l'apprentissage	4
2.2. Le corpus comme aide à la rédaction et à la correction	4
2.2.1. Une ressource qui fait ses preuves.....	4
2.2.2. Une ressource à la portée de tous	5
2.2.3. Le web comme corpus ?	5
2.3. Recueillir des données d'utilisation d'un corpus	6
3. Méthodologie	8
3.1. Mise en place de l'étude.....	8
3.1.1. Public.....	8
3.1.2. Principe et chronologie.....	8
3.1.3. Introduction au corpus et à l'outil concordancier	9
3.2. Traitement des données.....	11
3.2.1. Recueil des données.....	11
3.2.2. Analyse des données.....	11
4. Résultats et discussion	13
4.1. Résultats des questionnaires de début d'étude	13
4.2. Utilisation des différents outils pour la correction	15
4.2.1. Ressources en lignes utilisées.....	15
4.2.2. Types d'utilisation du corpus	16
4.2.3. Utilisations inefficaces du corpus.....	17
4.2.4. Utilisations efficaces du corpus.....	18
4.2.5. Résultats des questionnaires de fin d'étude.....	19
4.3. Efficacité par type d'erreurs, comparaison des outils	21
4.4. La correction post-corpus.....	22
5. Conclusion.....	25

6. Références bibliographiques :	27
6.1. Ouvrages.....	27
6.2. Articles dans une revue	28
Liste des tableaux	30
Annexe 1 : Codage d'erreurs	31
Annexe 2 : Numéros et profils des élèves	32
Annexe 3 : Questionnaire de début d'étude	33
Annexe 4 : Questionnaire de fin d'étude	35
Annexe 5 : Utilisation détaillée du corpus par élève	37

1. Introduction

Le point de départ de ce mémoire est un constat : un certain nombre de mes élèves de lycée refont systématiquement les mêmes erreurs à l'écrit, et ce même après une correction. Il n'y a donc pas de vrai apprentissage par l'erreur dans ces cas-là. La correction des élèves est habituellement faite à partir de ressources en ligne : « des dictionnaires électroniques et des traducteurs automatiques qui représentent souvent un premier réflexe pour ces apprenants. » (Landure et Boulton 2010, 12). Je me suis donc demandé quels bénéfices les élèves pourraient tirer de l'introduction à un autre type d'outil pour la correction : le corpus.

Je me réfère ici à l'explication et la traduction de Landure et Boulton pour une définition du terme :

Par « corpus » nous n'entendons pas n'importe quel recueil de documents ; nous suivons la définition de McEnery *et al.* (2006, 5) selon laquelle un corpus est un grand ensemble de textes authentiques représentatif d'une variété donnée quelle qu'elle soit et exploitable sous forme électronique. (2010, 12)

1.1. Contexte

L'apprentissage par corpus (ASC) (Boulton et Tyne 2013), également nommé *data-driven learning* (DDL) par un de ses fervents utilisateurs, Johns (1990), a fait ses preuves auprès des chercheurs dans un contexte universitaire (Gaskell et Cobb 2004; O'Sullivan 2007; Chambers 2010; Kennedy et Miceli 2017; Boulton et Cobb 2017) mais également dans le secondaire. (Schaeffer-Lacroix 2009, 2010) Si cet outil sert pour l'apprentissage, il est également utilisé pour la rédaction, et de fait, la correction. Le but étant de sensibiliser les apprenants à la langue étudiée en les mettant dans une position de chercheurs (Bernardini 2000; Kennedy et Miceli 2017).

Notons que dans le contexte actuel, les élèves ont souvent le réflexe de vérifier des mots ou des segments de phrases en interrogeant un moteur de recherche, comme le souligne Ciekanski (2014) citant Chambers (2009) : « Ces millions de textes disponibles sur Internet forment en effet une "ressource" facile d'accès que la plupart des apprenants ont déjà l'habitude de consulter. » L'utilisation d'un corpus en ligne est donc sûrement moins étrange que pour des élèves d'anciennes générations pour qui l'utilisation de ressources en ligne n'était pas un réflexe.

1.2. Principe de l'étude et questions de recherche

La présente étude a pour but de déterminer si l'utilisation d'un nouvel outil pour la correction, le corpus, permettra aux élèves d'améliorer leurs corrections et leur réflexion sur

la langue. J'ai donc décidé de mener une étude comparative de la correction avec un corpus et avec des ressources en ligne¹ « traditionnelles ».

Je tâcherai donc de répondre aux questions suivantes :

- Les élèves arrivent-ils à s'appropriier un nouvel outil (le corpus) avec une introduction très succincte ?
- Le corpus, quand il est bien employé, permet-il une meilleure correction de la langue ? (si cela est valable, pour quels types d'erreurs ?)
- La correction avec un corpus permet-elle une plus grande réflexion sur la langue ?

Je commencerai par poser des jalons théoriques au sujet de l'erreur et sa correction et de l'ASC (2). J'exposerai ensuite ma méthodologie et parlerai plus en détail de mon public (3). Je finirai par présenter et analyser les résultats de l'étude (4).

¹ J'utilise également l'appellation « ressources » pour les ressources en ligne. L'appellation « outil » sert à désigner aussi bien les ressources que le corpus.

2. Etat de l'art

Le sujet choisi est à la jonction de nombreux concepts en enseignement : l'erreur, la correction de l'écrit, le travail à la maison, l'apprentissage par l'erreur, la réflexion linguistique... Dans le cadre de cette étude, j'ai fait le choix de parler principalement de deux grandes questions : l'erreur et sa correction (2.1) et le corpus comme aide à la correction et à l'apprentissage (2.2). Je finirai par des remarques concernant un aspect méthodologique : le recueil de données de correction. (2.3).

2.1. L'erreur et sa correction

2.1.1. Définitions et typologies variables

Bien qu'étant un concept quotidiennement utilisé par les enseignants et les apprenants, l'erreur ne fait pas l'objet d'une définition consensuelle. A titre d'exemple, Corder (1980) distingue les erreurs de performance des erreurs systématiques :

Aussi sera-t-il commode désormais d'appeler « fautes » les erreurs de performance, en réservant le terme d'« erreur » aux erreurs systématiques des apprenants, celles qui nous permettent de reconstruire leur connaissance temporaire de la langue, c'est-à-dire leur compétence transitoire. (Corder 1980, 13)

Cette définition pose des problèmes d'identification : comment est-ce que l'enseignant peut savoir si le passage en question est dû à un problème ponctuel ou régulier ?

Hamilton (2015), dans sa thèse centrée sur l'erreur, affirme que la distinction de sous-types d'erreurs varie selon l'objet d'étude :

Selon que l'on s'identifie à l'enseignant en classe de langue ou au chercheur en linguistique - les perspectives peuvent être diamétralement opposées, tout en étant complémentaires. (2015, 35).

En ce qui concerne la présente étude, je ne fais pas de distinction entre les origines de l'erreur. J'ai, par ailleurs, adopté une typologie afin d'aider les élèves à repérer leurs erreurs et à se corriger tout seuls : le segment erroné est souligné et associé à une lettre qui renvoie à un type d'erreur (voir Annexe 1 pour la liste des lettres et leur signification). Je parlerai des retours des élèves à propos de ce système en 4.1 mais il est intéressant de noter que celui-ci ne semble pas faire ses preuves partout :

Intriguingly, whether the teacher provides detailed indications (such as coding Prep for preposition, Tr for transitivity, etc.) or simply underlines the error does not seem to have much impact on the learners' ability to correct their errors.
(Boulton et Landure 2016)

Landure et Boulton soulignent le fait qu'un codage de l'erreur est supposé avoir un impact sur la capacité de l'élève à se corriger (même si cela ne semble pas toujours être le

cas). Ceci me mène à rattacher l'erreur et son traitement au but premier de l'enseignement : l'apprentissage.

2.1.2. Utilisation des erreurs pour l'apprentissage

Dès le début du XXème siècle, Palmer parle de la « nature correctrice » du cours de langue vivante (1917). La place primordiale des erreurs au sein du processus d'apprentissage d'une langue continue d'être réaffirmée (Hamilton 2015, 27).

L'erreur peut être identifiée de différentes manières, comme nous l'avons très brièvement évoqué, mais ce n'est pas là le plus grand problème selon Gaskell et Cobb : « *An error on a page is an important opportunity in acquisition. But apparently it is an opportunity largely unrealized.* » (2004, 304) C'est donc le traitement de l'erreur par l'apprenant qui doit être repensé. Parmi les différentes pistes envisageables, j'ai décidé de privilégier celle de l'utilisation des corpus.

En effet, de nombreux chercheurs et enseignants parlent des bénéfices de l'utilisation des corpus pour que l'étude de l'erreur résulte en un apprentissage durable (Boulton et Cobb 2017). Selon Boulton et Landure, on ne peut considérer que la consultation (d'un dictionnaire par exemple) entraîne la rétention (2016). Il faut donc que la correction soit associée à une réflexion qui, elle, sera source d'apprentissage. Cette réflexion linguistique est un des bénéfices prônés du *data-driven learning* :

De nombreux avantages ont été attribués au DDL ; notamment, la découverte interactive d'un corpus devrait aider les apprenants à devenir de « meilleurs apprenants » (Johns 1990, 31) puisqu'elle promeut une plus grande sensibilité à la langue et aux processus cognitifs afférents (O'Sullivan 2007; Tyne 2009) ainsi qu'une meilleure mémorisation (Cobb 1999). (Landure et Boulton 2010, 13)

C'est donc vers le corpus comme ressource pour la correction, mais également l'apprentissage, que je vais maintenant me tourner.

2.2. Le corpus comme aide à la rédaction et à la correction

2.2.1. Une ressource qui fait ses preuves

Le corpus a été étudié et validé comme source d'apprentissage d'une langue (Johns 1991; Tribble et Jones 1997) mais également, comme une ressource pour l'expression écrite (Kennedy et Miceli 2001; Boulton et Landure 2016). Il n'y a alors qu'un pas vers l'utilisation du corpus pour la correction (Gaskell et Cobb 2004; Chambers 2010; Boulton et Landure

2016). Boulton et Landure présentent deux cas concrets d'utilisation d'un corpus (le COCA²) pour aider des étudiants à se corriger. Ils mettent en avant certaines difficultés rencontrées par les étudiants, telle qu'une prise en main difficile pour des apprenants qui n'ont pas l'habitude d'interpréter ce type de résultats et des problèmes liés à l'interface. C'est également ce que soulignent Mueller et Jacobsen (2016, 8) ou encore Kennedy et Miceli (2017, 93).

2.2.2. Une ressource à la portée de tous

Malgré les problèmes rencontrés par de nouveaux utilisateurs, beaucoup de chercheurs mettent en avant une utilisation à la portée de tous, niveaux de langue et manières d'apprentissage confondus.

En effet, il semblerait que l'ASC puisse correspondre à différents types d'apprenants : « *guided DDL-type induction may be beneficial for both deductive and inductive learners irrespective of their learning styles* » (Mizumoto et Chujo 2016, 55). De plus, de nombreuses études mettent en avant la plus-value d'un corpus, et ce même avec un faible niveau en langue (Schaeffer-Lacroix 2009; Landure et Boulton 2010; Mueller et Jacobsen 2016; Simard 2018). Eva Schaeffer-Lacroix, dans une étude avec 14 élèves adolescents de niveau faible à peu avancé, montre que le corpus permet à tous les élèves de complexifier leurs textes, à des degrés variables (2010). Kennedy et Miceli ont également une expérience positive avec des étudiants de niveau intermédiaire : « *in that the seed appears to have been sown for many of them to make ongoing effective use of CWIC and other corpora.* » (2017, 11). Cette « aisance » s'explique peut être par le fait que de nombreux apprenants ont déjà l'habitude d'interroger le web pour répondre à leurs questions linguistiques : « une familiarité avec les moteurs de recherche sur Internet peut servir de tremplin à l'exploitation d'un corpus proprement dit. » (Landure et Boulton 2010, 13)

2.2.3. Le web comme corpus ?

Si les élèves ne sont pas accoutumés aux corpus, ils manient les TIC de manière très fréquente :

Les technologies de l'information et de la communication (TIC), et surtout Internet, figurent parmi les ressources les plus utilisées par les apprenants de nos jours (Littlemore et Oakey 2004; Kenning 2007). (Landure et Boulton 2010, 12)

² *Corpus of Contemporary American English* (<http://corpus.byu.edu/coca/>, consulté le 30/03/2018)

Or le web est considéré comme un corpus par certains. Sinclair, qui n'est pas de cet avis, concède toutefois « *the web itself... [is a] huge source of language that is available in the classroom or the study at home* » (Sinclair 2004, 297).

Il est intéressant de noter que si on retient le web comme corpus, alors on peut voir un moteur de recherche, tel que Google, comme un concordancier (offrant en effet peu d'outils d'analyse). Pour reprendre l'expression de Robb (2003) : « *Google as a quick 'n' dirty corpus tool* ». En regardant les pratiques des apprenants, Kilgarriff et Grefenstette remarquent qu'ils se servent du web comme d'un corpus et « l'interrogent » pour vérifier l'orthographe d'un mot par exemple :

Is it *speculater* or *speculator*? Google gives 67 for the former (usefully suggesting the latter might have been intended) and 82,000 for the latter. Question answered. (2003, 333)

Les arguments mis en avant par ceux qui considèrent que le web n'est pas un corpus sont nombreux. Parmi les plus récurrents on trouve l'argument de la « fiabilité » : le web n'est pas contrôlé, comment se fier à ce qu'on y trouve sur le plan linguistique ? Kilgarriff et Grefenstette (2003), qui pensent que le web est un corpus, contestent cette vision. Une version erronée sera toujours en quantité bien inférieure à une version correcte d'après leurs travaux (2003, 9), le tableau suivant (les recherches sont effectuées sur Google) en atteste :

I believe	*I beleave	*I beleive
4 million+	3910	70900

Tableau 1 : Nombre d'occurrences de *believe* et de ses versions erronées

Quelque soit la position adoptée dans ce débat, on ne peut nier que les apprenants « interrogent » le web de manière récurrente. Reste à voir si cette pratique est comparable à un « vrai » corpus sur le plan de l'apprentissage et de la réflexion.

2.3. Recueillir des données d'utilisation d'un corpus

Une étude ayant pour but d'analyser l'utilisation d'un corpus par des apprenants se heurte toujours au même problème : comment recueillir ces précieuses données d'utilisation ? « *The concern for keeping and examining a detailed record of students' actions while completing computer-based tasks is gaining importance in the research agenda.* » (Pérez-Paredes et al. 2011, 234). Pérez-Paredes et al. (2011, 235) recensent deux manières de collecter des données quantitatives : via un enregistrement automatique des recherches sur l'ordinateur (Gaskell et Cobb 2004; Johns 1997; Yoon 2008) ou alors avec des registres

manuels (Chambers et O'Sullivan 2004; O'Sullivan et Chambers 2006; Gaskell et Cobb 2004). Si un enregistrement automatique permet d'avoir des données fiables du point de vue du nombre de recherches et des manipulations effectuées, les registres manuels donnent cependant accès aux remarques des apprenants, ce qui laisse entrevoir leur processus cognitif, la quête de tout enseignant. C'est pourquoi j'ai décidé de mettre en place un protocole avec des registres manuels et des questionnaires, comme je l'explique en détail dans la partie qui suit.

3. Méthodologie

3.1. Mise en place de l'étude

3.1.1. Public

J'ai décidé de mettre en place cette étude avec une classe de Seconde comportant 38 élèves (seuls 36 d'entre eux ont fait les deux corrections donc le groupe étudié pour la correction est de 36 élèves). Ces élèves ont cours en demi-groupe une fois par semaine pendant une heure et demie. Je profite de cette séparation de la classe pour effectuer des comparaisons. Cette étude concerne donc deux groupes de la même classe de Seconde, désormais Groupe 1 (17 élèves) et Groupe 2 (19 élèves). Etant donné que ces élèves sont tous issus d'un processus de sélection rigoureux, leur niveau scolaire est supérieur à la moyenne française. Il en est de même pour leur niveau d'anglais, tous les élèves ont au moins un niveau B1 et certains ont un niveau B2, voire C1. J'ai donné un numéro d'anonymat à chaque élève en fonction de son numéro de groupe³.

3.1.2. Principe et chronologie

Afin de pouvoir comparer l'efficacité de différentes ressources pour la correction, j'ai fait une étude en deux temps. Les élèves ont dû composer deux expressions écrites, puis les corriger. Depuis le début de l'année, j'utilise un code de lettres pour corriger les élèves : les segments incorrects sont soulignés et associés à une lettre qui renvoie au type d'erreur (voir l'Annexe 1 pour une explicitation du codage). C'est à partir de ces travaux annotés que les élèves doivent se corriger en utilisant des ressources différentes selon le moment de l'étude, que je vais détailler à présent.

3.1.2.1. Première expression écrite et correction

Les élèves ont eu une première tâche d'expression écrite à effectuer (EE1 dorénavant) fin décembre. Le groupe 1 bénéficie d'une introduction au corpus et à l'outil concordancier une semaine avant que les EE1 ne soient rendues (détail en 3.1.3). Chaque groupe corrige ensuite son EE1 en salle informatique pendant une heure. Le groupe 1 n'a accès qu'au corpus, le groupe 2 peut utiliser toutes les ressources disponibles en ligne (le corpus n'est pas encore connu).

3.1.2.2. Deuxième expression écrite et correction

Les élèves ont eu une deuxième tâche d'expression écrite à effectuer (EE2 dorénavant) fin janvier. Le groupe 2 bénéficie d'une introduction au corpus et à l'outil concordancier une

³ La liste des groupes et du niveau de chaque élève est en Annexe 2.

semaine avant que les EE2 ne soient rendues. Chaque groupe corrige ensuite son EE2 en salle informatique pendant une heure. Le groupe 2 n'a accès qu'au corpus, le groupe 1 peut utiliser toutes les ressources disponibles en ligne (y compris le corpus, qu'ils connaissent).

3.1.2.3. Chronologie détaillée

Le tableau qui suit regroupe toutes les informations qui viennent d'être détaillées.

Date	Groupe 1	Groupe 2
16/12/17	Rédaction de l'EE1 en classe	
09/01/18	Introduction au corpus	X
16/01/18	EE1G1 : Correction de l'EE1 avec corpus uniquement (EE1G1)	EE1G2 : Correction de l'EE1 avec ressources en ligne diverses (corpus inconnu)
27/01/18	Rédaction de l'EE2 en classe	
30/01/18	X	Introduction au corpus
06/02/18	EE2G1 : Correction de l'EE2 avec ressources en ligne diverses (corpus connu et utilisable)	EE2G2 : Correction de l'EE2 avec corpus uniquement

Tableau 2 : Chronologie détaillée de l'étude

3.1.3. Introduction au corpus et à l'outil concordancier

References to the need for guidance are constant in empirical studies involving students searching and analyzing corpus data. Assistance is most often offered in the form of lessons or seminars on the basics of Corpus Linguistics or concordancing prior to hands-on work with the tools (Pérez-Paredes et al. 2011, 236)

Cette citation met en lumière le besoin d'accompagner les apprenants dans leur utilisation du corpus, outil dont ils ne connaissent souvent rien et qui n'est pas si facile à appréhender. Etant donné le niveau d'études des élèves, je n'ai pas jugé nécessaire de leur faire un cours de linguistique de corpus mais ai préféré leur inculquer quelques rudiments et puis rentrer dans le vif du sujet, comme je vais maintenant l'expliquer.

Les élèves ont rempli un premier questionnaire⁴ (dont je parle plus en 3.2.1) en début de séance d'introduction au corpus, à la fin duquel je leur demande si ils savent ce qu'est un corpus. La totalité des élèves ont répondu de manière négative, ce qui permet une parfaite entrée en matière pour la suite de la séance. De plus, le caractère curieux de la plupart des élèves, chez qui je viens de créer un déficit de connaissance, les rend très attentifs aux explications à venir. Je précise que j'ai fait toutes les explications en français, l'introduction

⁴ Voir Annexe 3

de nouveaux mots tels que « corpus », « concordancier » ou encore « concordance » étant déjà source de bien des froncements de sourcils. Mes explications de ces termes étaient succinctes mais dans le cadre de cette étude, je précise que j'adhère aux définitions suivantes pour les termes « concordance » et « collocation » :

Je définis le terme de "concordance" comme suit : une concordance est une mise en forme particulière des données d'un corpus. Le texte n'y est pas présenté en continu, mais il est matériellement réorganisé suite à une recherche de toutes les occurrences d'un mot-clé ou d'un segment dans le corpus. Les occurrences sont regroupées sous forme de liste. Les éléments de cette liste apparaissent sous forme de trois colonnes, à savoir le contexte de gauche du mot-clé, le mot-clé et le contexte de droite du mot-clé. [...] Les lignes de concordance aident à observer les mots-clés dans leur contexte immédiat. (Schaeffer-Lacroix 2010)

Les collocations [...] se composent d'un groupe d'au moins deux mots qui se suivent ou qui sont séparés par d'autres mots. Elles forment une unité de sens, et elles sont soumises à certaines restrictions syntaxiques, sémantiques et pragmatiques. Ces unités linguistiques se caractérisent par un certain degré de figement et de récurrence au sein d'un corpus. (Schaeffer-Lacroix 2010)

J'ai beaucoup réfléchi à la manière dont j'allais introduire le corpus, le but étant de donner assez d'informations aux élèves, sans pour autant les effrayer, pour les préparer à l'exercice à venir :

Le véritable défi pour l'enseignante est de les faire réfléchir sur la langue, son utilisation et son usage, ensuite de les amener à se poser les « bonnes » questions afin d'utiliser la ressource sciemment et efficacement et ainsi de pouvoir mieux appréhender et interpréter les résultats. (Landure 2014, 159)

Il convient ici de brièvement parler de mon choix de concordancier. Il fallait un concordancier en ligne pour des raisons pratiques (difficulté à installer des logiciels en salle informatique). J'ai fait le choix d'une interface proposée par l'université de Leeds⁵ qui permet d'interroger le *British National Corpus* (BNC). Cette interface est facilement accessible, gratuite, et ne propose pas trop de fonctionnalités, qui perdraient les élèves. Les deux fonctionnalités principales, que j'ai donc présentées aux élèves, sont la recherche de concordances et de collocations, avec possibilité d'utiliser l'outil *Part of Speech* afin de faire une recherche plus fine (chercher les verbes qui entre en collocation avec un certain nom par exemple). J'ai commencé par brièvement définir « corpus », afficher la page de concordancier que nous allons utiliser et expliquer le concept de « concordance » et « collocations », manipulations à l'appui. Comme le suggèrent Gaskell et Cobb, « *a teacher should set up the concordance outputs for the learners, at least initially, and should offer a period of training in their interpretation.* » (2004, 303), j'ai posé des questions aux élèves comme « quelle

⁵ <http://corpus.leeds.ac.uk/protected/query.html> (consulté le 12 avril 2018)

préposition utilise-t-on après *tell* ? » et leur ai montré comment interroger le corpus, tout en leur demandant d'interpréter eux-mêmes les résultats. Je leur ai ensuite proposé un exercice d'entraînement avec un texte (un extrait d'un devoir anonyme d'un élève d'un autre établissement) dans lequel j'avais mis en gras les erreurs que je leur suggérais d'essayer de corriger. J'ai circulé dans les rangs pendant ce quart d'heure d'entraînement et me suis assurée qu'ils avaient tous compris comment manier les deux types de moyens d'interroger le corpus.

3.2. Traitement des données

3.2.1. Recueil des données

Afin d'avoir un aperçu le plus complet possible du rapport à la correction et au corpus, j'ai utilisé plusieurs sources de données. Les élèves ont d'abord rempli un questionnaire anonyme de début d'étude portant sur leurs pratiques de correction (Annexe 3). Lors des deux sessions de correction, je leur ai fait remplir des tableaux m'informant de leur choix de correction sur le modèle du tableau 3, j'ai donc opté pour des registres manuels, comme je l'expliquais en 2.3.

Erreur commise	Code	Ressource	Utilisation de la ressource	Proposition de correction
----------------	------	-----------	-----------------------------	---------------------------

Tableau 3 : Rubriques à remplir par les élèves lors des phases de correction

Il faut noter ici que les tableaux n'étaient pas tous remplis avec le même degré de précision. Ce type de recueil de données fait appel au sérieux des élèves et, même si l'ensemble du groupe a effectué les tâches avec enthousiasme et sérieux, certains élèves n'ont parfois pas assez détaillé leurs tableaux. La case « Utilisation de la ressource » a très souvent été laissée vide, ce qui est dommage car c'est un moyen d'essayer de comprendre le processus de réflexion de l'apprenant.

Enfin, après leur session de correction avec le corpus, je leur ai demandé de remplir un deuxième questionnaire les invitant à réfléchir sur l'utilisation du corpus (Annexe 4).

3.2.2. Analyse des données

Les différentes données que je viens de présenter ont ensuite été analysées. Les questionnaires incluent des questions fermées (type échelle de Likert ou OUI/NON) mais également des questions ouvertes invitant les élèves à produire une réponse plus ou moins longue. Ils permettent donc d'avoir deux types de données : des données quantitatives et qualitatives.

Les tableaux remplis par les élèves lors des deux corrections m'ont permis de compiler d'autres tableaux de données quantitatives, à l'échelle d'un demi-groupe ou de la classe, sur le type d'utilisation du corpus ou encore le choix et l'efficacité de l'outil par type d'erreur. Il faut noter que les erreurs traitées ne sont que les erreurs qui ont fait l'objet d'une correction avec un outil. Je n'ai donc pas inclus les erreurs corrigées par les élèves eux-mêmes (ou leurs camarades) ni les erreurs qui n'avaient pas du tout été corrigées (par manque de temps) dans les chiffres. Il aurait été intéressant de comparer l'utilisation d'un outil à la correction par l'élève ou un camarade mais cela aurait généré beaucoup de données supplémentaires, que je n'aurais pas eu le temps de traiter. J'ai donc préféré m'en tenir aux données de correction avec un outil et les analyser en détail.

Toutes ces données font l'objet d'une présentation et analyse détaillées (Partie 4) et sont utilisées pour essayer de répondre à mes questions de recherche :

- Les élèves arrivent-ils à s'appropriier un nouvel outil (le corpus) avec une introduction très succincte ?
- Le corpus, quand il est bien employé, permet-il une meilleure correction de la langue ? (si cela est valable, pour quels types d'erreurs ?)
- La correction avec un corpus permet-elle une plus grande réflexion sur la langue ?

4. Résultats et discussion

J'ai décidé d'organiser mes données de manière chronologique et thématique. Je vais commencer par présenter les résultats du premier questionnaire portant sur la correction, me permettant de faire un état des lieux du rapport des élèves à cette dernière (4.1). Je vais ensuite détailler la manière dont les élèves se servent des outils de correction (ressources en ligne et corpus, avec des exemples d'utilisations efficaces et inefficaces du corpus). (4.2) Ceci me mènera à comparer les taux d'efficacité de chaque outil. (4.3) Je finirai par parler de la « correction post-corpus », l'étape où les élèves du G1 ont le choix entre le corpus et les ressources en ligne, ceci me permet de comparer le taux d'utilisation des deux ressources et leur efficacité pour un même groupe. (4.4)

4.1. Résultats des questionnaires de début d'étude

1. What resources do you use when you correct a WE at home?	
Paper dictionary all in English	2,63%
Paper dictionary English <-> French	47,37%
Online dictionary all in English	26,32%
Online dictionary English <-> French	84,21%
Online translator	63,16%
I just correct it by myself	39,47%
I ask someone (friend, family) to help out	31,58%
2. Do you always correct your WEs? (1=always; 4=never)	
1	18,42%
2	39,47%
3	34,21%
4	7,89%
3. What is the longest amount of time you could spend correcting a WE?	
Less than 10 minutes	15,79%
10 minutes	15,79%
20 minutes	26,32%
30 minutes	2,63%
As long as it takes	39,47%
4. Do you find the letter code useful?	
Yes	86,84%
No	13,16%
5. Do you sometimes struggle when correcting WE?	
Yes	68,42%
No	31,58%
6. If you answered YES, estimate how often (1=almost never; 4=very often)	
1	28,95%
2	36,84%
3	10,53%
4	2,63%

Tableau 4 : Résultats de questionnaire de début d'étude

Ce questionnaire comporte plusieurs types de questions (des questions fermées, des questions à choix multiples etc) dont des questions ouvertes. Le tableau 4 recense les réponses de toute la classe aux questions fermées alors que le tableau 5 présente des extraits de réponses aux deux questions ouvertes. Si le questionnaire a été rédigé en anglais, les élèves avaient le choix de répondre en anglais ou en français.

Les réponses du groupe entier permettent de dresser un portrait de la classe face à la correction. Il est d'abord intéressant de noter que les ressources de correction papiers résistent, 47,37% des élèves disent se servir d'un dictionnaire papier anglais-français. Néanmoins, on remarque, sans surprise, que le même type de dictionnaire en ligne est utilisé par 84,21% de la classe. De plus, les élèves préfèrent les traducteurs aux dictionnaires monolingues. Le processus de correction n'est pas le même pour tous, on voit que certains passent jusqu'à trois fois plus temps à se corriger (même si il est possible que cette différence de temps soit liée au nombre d'erreurs à corriger et non à la bonne volonté de l'élève). Enfin, si les deux tiers de la classe disent parfois rencontrer des difficultés lors de la correction, la plupart indiquent que ces difficultés ne sont pas fréquentes.

Les élèves sont majoritaires à trouver utile la typologie d'erreurs mise en place. Ceci va à l'encontre de la citation de Boulton et Landure en 2.1.1, selon laquelle un codage d'erreurs n'a pas d'influence sur la correction des élèves. Bien évidemment, ce n'est pas parce que les élèves disent que la typologie est utile qu'elle l'est réellement, et je n'ai pas de données à avancer pour illustrer cette utilité, cela pourrait faire l'objet d'une tout autre étude.

1. Do you think correcting written expressions helps you learn? Why?
1.1 <i>Yes, sometimes it helps me to understand my mistakes. But when I don't know how to correct one, it doesn't help me at all.</i>
1.2 <i>Yes, always. I can learn my mistakes, and don't do it again.</i>
1.3 <i>Yes, because when you correct them, you sometimes notice mistakes you used to do all the time but didn't realize you did.</i>
1.4 <i>Oui, parce que comme ça on peut comprendre, par nous même, comment éviter nos erreurs afin que l'on puisse s'améliorer, et éviter de les répéter.</i>
2. Do you think there is a way you could struggle less [when correcting written expressions]?
2.1 <i>Something like Linguee which explains me, shows me some sentences.</i>
2.2 <i>I don't struggle when I correct my WEs. But sometimes, it would be better if I was sure to what I found online. Indeed things can be fake.</i>
2.3 <i>I think I could struggle less if I spend more time on it or maybe if I find a better online translator.</i>
2.4 <i>Je pense que les corriger en classe m'oblige à les corriger parce que chez moi je ne le fais pas</i>

Tableau 5 : Extraits de réponses au questionnaire de début d'étude

Si l'on se tourne maintenant vers les extraits de réponses des élèves, on peut noter plusieurs choses. Les élèves pensent majoritairement que la correction aide à l'apprentissage et qu'une faute corrigée ne sera pas répétée. Or, j'ai pu voir au début de l'année que des fautes corrigées individuellement et en classe étaient parfois systématiquement répétées (sans parler du sempiternel oubli du S de la 3^{ème} personne que chaque élève a eu à corriger au moins une fois mais qui est toujours bien trop récurrent). Les élèves semblent donc avoir une vision quelque peu romancée des bienfaits de la correction. La réponse 1.4 pointe du doigt un phénomène important : le travail de réflexion qui se met en place lors de la correction. En effet, c'est alors aux élèves de comprendre « par eux-mêmes » comment se corriger. La question 2 a produit des réponses intéressantes. Les réponses 2.1 et 2.2 peuvent être vues comme un appel au corpus (du moins par la convertie à l'ASC que je suis) : Linguee fonctionne après tout comme un corpus parallèle et quoi de mieux qu'un corpus de langue authentique pour être sûr de ce que l'on trouve ? On remarque néanmoins que les élèves sont toujours proches de l'idée de traduction (Lingue en 2.1, *online translator* en 2.3). L'utilisation du corpus pour la correction doit donc concurrencer le traducteur en ligne si cher aux élèves. Commençons d'abord par analyser la manière dont les élèves se corrigent avec les deux outils proposés.

4.2. Utilisation des différents outils pour la correction

Je propose de commencer par regarder comment les élèves se servent des outils à leur disposition. Je vais développer l'utilisation du corpus plus en détail étant donné que c'est un outil nouveau pour les élèves et qu'un des buts de cette étude est de voir comment ils s'en servent.

4.2.1. Ressources en lignes utilisées

J'utilise régulièrement WordReference en cours dans l'espoir de convaincre les élèves que Google Traduction n'est pas la réponse à tout (voire même à quoi que ce soit). Ceci explique probablement en partie le fait que ce soit la ressource utilisée par le plus grand nombre dans le tableau 6. Ce tableau répertorie le nombre d'élèves utilisant chaque ressource, un élève peut donc être compté plusieurs fois si il utilise plusieurs ressources. On peut noter que les ressources de traduction sont les plus présentes, comme le questionnaire de début de cours le laissait penser. Il faut également souligner que, malgré mes mises en garde au sujet

de Google Traduction, presque un quart des élèves s'en servent (je leur ai bien dit de se sentir libres dans leur choix de ressources).

Ressource en ligne	Nombre d'élèves utilisant la ressource (n = 36)
Word Reference	13
Reverso	10
Google Traduction	8
Linguee	7
Google	3
Anglais facile	3
Larousse en ligne	2
Sites non spécifiés	2
Traducteur non spécifié	1
Correcteur Libre Office	1

Tableau 6 : Liste des ressources en lignes par popularité

4.2.2. Types d'utilisation du corpus⁶

Comme je le précise en 3.1.3, les élèves ont appris ce qu'était la recherche de concordance et de collocations avec le concordancier. J'ai répertorié leur type d'utilisation du corpus dans le tableau 7. Quelques élèves se sont apparemment assez bien approprié l'outil, jusqu'à en utiliser ses fonctionnalités de manière combinée (3 élèves font des recherches de concordance et de collocation pour une même requête). Ensuite on voit qu'environ 45% des élèves se sont servi du corpus pour faire des vérifications, donc pour confirmer une intuition linguistique. Je parlerai plus en détail de ce phénomène en 4.4, quand il est combiné à des recherches avec ressources. Enfin, on peut noter la popularité des recherches de type concordance. Je pense que l'on peut expliquer ce phénomène par la relative « plus grande accessibilité » de cet outil pour des apprenants non linguistes. En effet, les élèves ne savaient pas ce qu'était une collocation avant le début du cours et ont eu plus de mal à s'approprier ce type de recherche et à en comprendre les enjeux. La concordance donne des résultats d'apparence « plus directe » : l'élève tape une recherche et voit tout de suite apparaître (ou non) des lignes d'occurrences qu'il lui « suffit » de scruter pour trouver sa réponse. Ce résultat va par ailleurs à l'encontre de l'étude de Koo (2006) dans laquelle « *Collocations were the most common reason (69.0%) that they [the students] consulted the corpus* » (Mueller et Jacobsen 2016, 4).

	Concordance	Collocation	Collocation et concordance	Concordance Vérification	Collocation Vérification
Nombre d'élèves (n = 36)	31	14	3	15	1

⁶ Voir Annexe 5 pour l'utilisation détaillée du corpus pour chaque élève.

Nombres d'occurrences (n = 132)	82	24	3	22	1
---------------------------------	----	----	---	----	---

Tableau 7 : Types d'utilisation du corpus par nombre d'élèves et nombre d'occurrences

4.2.3. Utilisations inefficaces du corpus

Je vais détailler quelques utilisations inefficaces du corpus qui me paraissaient particulièrement pertinentes dans le cadre de l'étude car elles donnent un aperçu du type d'erreurs que fait le public en question.

Elève	Code Erreur	Ressource	Utilisation	VI ⁷	VC ⁸
216	S	corpus	concordance	english essay	essay of english
216	L/G	corpus	concordance	I did wrong (j'ai fait une erreur/ J'ai fait quelque chose de mal)	it went wrong
203	L	corpus	concordance	testimony (testament)	?
203	T	corpus	collocation	you just came back from the garden (tu venais de revenir...)	you just come back

Tableau 8 : Exemples d'utilisations inefficaces du corpus

Elève 216

La première erreur de cet élève est très commune, malgré mon insistance à chaque fois que je vois ce type d'erreur. Le code erreur (S pour *Spelling*) aurait dû mettre l'élève sur la voie car a priori il sait écrire *English* et *essay*. Cependant, il semble ne pas avoir compris le code et a fait des recherches de concordance. Les recherches *English essay*, **english essay*, *essay of English* et **essay of english* ne donnent aucun résultat dans le BNC. Je ne sais donc pas comment l'élève est parvenu à sa proposition de version corrigée avec le corpus. Je considère ce cas comme un exemple typique de réflexion d'apprenant qui semble incohérente au professeur et de situation où il manque de l'information pour savoir comment l'apprenant en est arrivé au résultat avancé – il n'a pas rempli la case « utilisation de la ressource ». On peut même se demander si l'élève a rempli son document honnêtement.

La deuxième erreur donne un exemple de cas où la proposition de correction est un segment linguistiquement correct mais qui ne marche pas dans le contexte de la copie de l'élève.

Elève 203

Cet élève cherchait le mot anglais pour « testament ». Cette recherche aurait été extrêmement rapide, et probablement très efficace, avec un traducteur en ligne. Or le corpus

⁷ VI : Version Initiale

⁸ VC : Version Corrigée

n'a pas permis à l'élève de trouver une réponse satisfaisante car ce dernier n'avait pas la moindre idée de la bonne version. On voit ici une des limites du corpus, qui permet de confirmer que *testimony* n'a pas le sens voulu mais pas de lui trouver une alternative.

La deuxième erreur concerne les temps. Les élèves ont globalement du mal à corriger ce type d'erreur (voir tableau 13), surtout quand ce qu'ils proposaient était correct de manière isolée mais pas dans le contexte de leur texte. C'est ce qu'on peut voir ici : l'élève propose un prétérit qui est correctement écrit mais pas ce qu'il faut dans le contexte (j'ai mis en français ce qu'on attend). Or le corpus ne va pas lui dire qu'il faut un *past perfect*. L'élève explique son utilisation de la manière suivante : « J'ai choisi *come back* dans collocations et j'ai demandé quels adverbess allaient devant ». Si elle avait fait suivre cette première recherche d'une recherche de concordances de *just come back*, elle aurait pu voir que cette forme est majoritairement utilisée en combinaison avec l'auxiliaire *have* pour former un *present* ou *past perfect*. Mais pourquoi faire une telle recherche supplémentaire si on pense avoir trouvé le bon résultat ?

Ces différents exemples permettent de comprendre pourquoi une recherche peut être inefficace : l'élève ne fait pas attention à ce qu'il cherche/doit chercher (216), l'élève prend une proposition qui ne fonctionne pas dans son contexte (216 et 203), l'élève cherche une traduction (203), l'élève arrête ses recherches trop rapidement car il pense avoir trouvé la bonne réponse (203).

4.2.4. Utilisations efficaces du corpus

Elève	Code Erreur	Ressource	Utilisation	VI	VC
210	P	corpus	concordance	get closer of them	closer to them
216	S	corpus	concordance	sincerly	sincerely
101	A	corpus	concordance vérification	many app	many apps
101	G	corpus	concordance	that great (c'est super)	that's great
103	L	corpus	collocation	spend a good time	have a good time
104	P	corpus	collocation vérification	reading in the sofa	reading on the sofa

Tableau 9 : Exemples d'utilisations efficaces du corpus

Le tableau 9 présente des exemples d'utilisations efficaces. Je ne vais pas commenter ceux-ci en détail mais juste faire quelques remarques d'ordre général. Le corpus permet de corriger plusieurs types d'erreurs efficacement (comme en témoigne le tableau 13). On note également que l'outil collocation peut être utilisé de manière tout à fait satisfaisante : l'élève

103 a pu trouver un verbe qui fonctionne avec *a good time* pour rendre l'expression « passer un bon moment », qui ne peut être traduite littéralement.

On a ici des exemples montrant que les apprenants arrivent à des résultats très encourageants après très peu d'initiation au corpus. Ceci va dans le sens des recherches de Landure et Boulton (2010) mais également Frankeberg-Garcia (2005) :

Frankenberg-Garcia (2005) observe que les étudiants dans son enquête consultaient les corpus, récemment découverts, aussi facilement que les dictionnaires dont ils se servaient depuis longtemps. Elle en conclut que la formation nécessaire pour consulter un corpus n'est pas plus importante que celle requise pour se servir convenablement d'un dictionnaire – au contraire. (Landure et Boulton 2010, 13)

4.2.5. Résultats des questionnaires de fin d'étude

Après avoir parlé de la manière dont les élèves se servent du corpus, il convient d'analyser leur propre ressenti. Le tableau 10 recense les réponses fermées au questionnaire de fin d'étude (voir Annexe 4).

1. Mon avis sur le tutoriel d'introduction au corpus (1=je n'ai rien compris; 4=très clair)	
1	0,00%
2	22,22%
3	41,67%
4	36,11%
2. Pour corriger, le corpus a été utile (1=pas du tout d'accord; 4=tout à fait d'accord)	
1	0,00%
2	44,44%
3	52,78%
4	2,78%
3. Le corpus était facile à utiliser (1=pas du tout d'accord; 4=tout à fait d'accord)	
1	11,11%
2	44,44%
3	41,67%
4	2,78%
4. Selon vous, est-ce que le corpus peut être préférable à un autre outil pour corriger une erreur ?	
Oui	66,67%
Non	33,33%
5. Le corpus me fait plus réfléchir qu'un autre outil quand je corrige (1=pas du tout d'accord; 2=tout à fait d'accord)	
1	2,78%
2	16,67%
3	36,11%
4	44,44%
6. Le corpus m'a appris des choses (1=pas du tout d'accord; 4=tout à fait d'accord)	
1	8,33%
2	25,00%
3	41,67%
4	25,00%
7. Est-ce que vous pensez vous resservir du corpus pour corriger vos expressions écrites si vous êtes en salle info ?	
Oui	30,56%

Non	13,89%
Peut être	55,56%

Tableau 10 : Résultats du questionnaire de fin d'étude

Le tutoriel d'introduction au corpus semble avoir été relativement réussi, près de 80% des élèves l'ont considéré clair ou très clair. Par ailleurs, à la question « avez-vous des suggestions pour l'améliorer ? », les élèves ne font que suggérer plus d'exemples. Pour ce qui est de la prise en main, les réponses sont plus mitigées. A peine plus de 50% des élèves trouvent que le corpus a été utile ou très utile pour la correction et seulement près de 45% pensent que le corpus a été facile à utiliser. Ces résultats sont moins positifs que ceux de Landure et Boulton dans une étude assez similaire :

Pourtant, selon les résultats du questionnaire, 95 % d'entre eux estiment avoir eu assez d'informations et de conseils pour utiliser le corpus, et 70% trouvent même la consultation du corpus « plutôt facile » ou « très facile ». Par ailleurs, 79 % d'entre eux considèrent que la consultation de corpus est « plutôt utile » ou « très utile » (2010, 11)

1.Qu'est-ce qui a posé problème spécifiquement ?
1.1 Par exemple quand on ne connaît pas un mot de vocabulaire pour le trouver et qu'on n'a pas d'idée de mots qui le précède souvent.
1.2 Il faut parfois réfléchir assez longuement pour savoir comment formuler la "requête", et c'est parfois difficile de savoir comment s'en sortir, là où d'autres outils pourraient être plus efficaces...
1.3 J'avais un peu de mal à trouver exactement ce que je cherchais, c'est à dire par exemple que je devais passer pas plusieurs étapes pour arriver à ce que cherchais, et ca aurait été peut être plus rapide avec un autre outil.
1.4 Il y a certains mots dont on voudrait juste avoir la traduction.
1.5 .J'ai eu parfois du mal a bien utiliser le concordancier.
2.Dans quel cas est-ce que le corpus peut être préférable à autre chose ?
2.1 Lorsqu'on a des fautes de formulations, de prépositions, de déterminants... c'est vraiment très pratique. Par contre pour les fautes de vocabulaire, ou de temps des verbes je préfère utiliser un autre outil.
2.2 Surtout pour les correspondances verbes-prépositions.
2.3 Le corpus peut permettre de corriger rapidement des erreurs de déterminant.
2.4 Dans le cas où un mot est correct mais que nous l'avons mal utiliser, notamment à cause du contexte.

Tableau 11 : Extraits de réponses au questionnaire de fin d'étude

On voit dans le tableau 11 des extraits de réponses aux questions ouvertes. Les remarques des élèves pour la question 1 rejoignent celles faites dans d'autres études. En effet, les apprenants ont régulièrement des problèmes avec l'interface, « *First, many participants experienced difficulty when using the corpus interface.* » (Mueller et Jacobsen 2016, 8), mais

également à se positionner comme des chercheurs et à formuler des requêtes et analyser les résultats : « *the difficulties that learners have in behaving like linguistic researchers when using corpus data* » (Kennedy et Miceli 2017, 93). Mueller et Jacobsen résument les difficultés rencontrées dans le paragraphe suivant :

One area of interest in the study was whether the participants experienced difficulty using the corpus and what particular aspects of using the corpus they found difficult. Most participants reported considerable difficulty. Among participants who elaborated on the causes of this, six mentioned general problems in using the computer (two participants mentioned that they do not own a computer); one specifically mentioned difficulty negotiating the interface, as the webpage elements all appeared in English. Other difficulties included trouble understanding how to determine the appropriate span when searching for collocates (two responses), difficulty using the chart function, and trouble choosing the best word for a query. This latter response is understandable as searches require some intuitions regarding the strength of association between two words. (Mueller et Jacobsen 2016, 14)

On note par ailleurs qu'une majorité d'étudiants pensent que le corpus peut être la meilleure solution pour certains types d'erreurs : les prépositions reviennent très souvent dans les réponses libres. C'est également ce qu'affirment Chambers et O'Sullivan (2004, 167-68). Enfin les questions 5 et 6 ont des réponses majoritairement positives (80% et 66%), les élèves ont donc entrevu le potentiel d'apprentissage du corpus. Pour conclure, les réponses à la question 7 sont relativement satisfaisantes, mais il faudrait évidemment mettre les élèves dans une telle situation afin de pouvoir confirmer une réutilisation, ou non, du corpus. Cela a été le cas pour le groupe 1 en EE2, j'en parlerai en 4.4.

4.3. Efficacité par type d'erreurs, comparaison des outils

Je vais maintenant m'intéresser de plus près aux données chiffrées de correction, et notamment à l'efficacité de chaque outil pour chaque type d'erreurs. Le tableau 12 indique le nombre d'erreurs de chaque type (je rappelle que seules sont comptées les erreurs corrigées par les élèves à l'aide d'un outil) et le nombre de ces erreurs ayant été corrigées avec succès. On peut voir que les types d'erreurs les plus fréquents sont les erreurs lexicales ; les erreurs d'accord, de détermination et d'ordre étant les moins fréquentes.

Code Erreur	Nombre d'erreurs	Nombre de corrections avec succès
A	14	12
D	7	5
G	38	24
L	69	44
O	10	8
P	46	37
S	42	30

T	45	27
---	----	----

Tableau 12 : Données générales liées aux erreurs et à leur correction

Le tableau 13 répertorie les taux d'efficacité des ressources (colonne 2) et du corpus (colonne 3). La ligne en bas du tableau indique les moyennes pour chaque colonne.

Code Erreur	Efficacité des ressources en ligne	Efficacité du corpus
A	80,00%	88,89%
D	100,00%	100,00%
G	55,56%	82,61%
L	91,67%	66,67%
O	100,00%	100,00%
P	70,00%	93,75%
S	100,00%	94,12%
T	93,33%	72,22%
	86,32%	87,28%

Tableau 13 : Fréquences d'utilisation et d'efficacité pour chaque type de correction et d'erreur

Si on commence par regarder les moyennes, on peut avancer que les outils sont globalement efficaces et que les taux d'efficacité sont similaires. Il faut donc s'intéresser aux différents types d'erreurs pour trouver d'éventuelles disparités.

Le corpus est plus efficace pour corriger les erreurs de d'accords, grammaire et de prépositions alors que les ressources ont l'ascendant pour les erreurs de lexique, d'orthographe et de temps. Ces résultats rejoignent les remarques des élèves mais également les intuitions qu'on peut avoir à propos de ces différents outils : un traducteur en ligne semble a priori plus efficace pour trouver le bon mot, du moins quand on ne sait pas assez bien se servir du corpus pour en exploiter les fonctions de collocations et de contextualisation. Ces résultats viennent confirmer ce que de nombreux chercheurs avancent : ces outils sont tous les deux pertinents mais leur efficacité varie selon le type d'erreur (Chambers et O'Sullivan 2004; Landure et Boulton 2010; Mueller et Jacobsen 2016).

4.4. La correction post-corpus

L'EE2G1 a permis de voir quels outils les élèves utilisent si on leur laisse le choix. Le tableau 14 permet d'avoir un aperçu du nombre d'élèves utilisant chaque outil pour chaque correction. Il y a quatre types d'outils : le corpus, une ressource, une combinaison de ressources (Google Traduction puis Linguee pour une même correction par exemple) et enfin

une combinaison corpus et ressource (collocation avec le corpus puis un traducteur en ligne par exemple, voir tableau 15). On peut voir que si quasiment tous les élèves se sont servi des ressources, moins de la moitié se sont servi du corpus seul. La combinaison d'outils est intéressante car je n'en ai pas parlé en classe, c'est donc une intuition de 4 élèves. Koo (2006) trouve que c'est ce qu'il y a de plus efficace :

Koo's qualitative study is notable as it involved an examination of both corpus-consultation and the use of conventional reference tools. [...] They [the students] often combined tools, and when they did so, were more successful in finding solutions to their writing problems. (Mueller et Jacobsen 2016, 4)

	EE1G1 (n = 17)	EE1G2 (n = 19)	EE2G1 (n = 17)	EE2G2 (n = 19)
Corpus	16		7	17
Ressource		16	15	
Combinaison de ressources			2	
Corpus + ressource			4	

Tableau 14 : Nombre d'élèves utilisant les différentes ressources par groupe et exercice

Les élèves ayant combiné corpus et ressource n'ont malheureusement pas détaillé leur processus d'utilisation. L'élève 102 a juste indiqué avoir utilisé le corpus en premier. On peut imaginer que le traducteur en ligne a servi pour trouver le nom entrant en collocation avec « *of brownie* ».

Elève	Code Erreur	Ressource	Utilisation	VI	VC
102	L	corpus + traducteur en ligne	collocation (pour la préposition après "dream")	I dream of a big part of brownie	I dream about a piece of brownie

Tableau 15 : Exemple d'utilisation combinée du corpus et d'une ressource

Le tableau 16 permet de comparer l'utilisation et l'efficacité des deux outils pour la correction d'un même travail. On y retrouve globalement les mêmes tendances que dans le tableau 13 si ce n'est que le corpus est légèrement plus efficace pour la correction des erreurs de lexique, mais cette différence est potentiellement due au fait que le corpus est moins utilisé, et donc moins source d'erreurs potentielles, que les ressources pour ce type d'erreur.

		Corpus				Ressources en ligne			
Code Erreur	Nombre d'erreurs	Nombre de corrections	Utilisation	Avec succès	Efficacité	Nombre de correction	Utilisation	Avec succès	Efficacité
A	1	1	100,00%	1	100,00%	1	100,00%	1	100,00%
D	2	1	50,00%	1	100,00%	1	50,00%	1	100,00%
G	6	3	50,00%	3	100,00%	3	50,00%	2	66,67%
L	15	4	26,67%	4	100,00%	14	93,33%	13	92,86%
O	2	0	0,00%			2	100,00%	2	100,00%
P	5	4	80,00%	3	75,00%	2	40,00%	1	50,00%

S	11	1	9,09%	1	100,00%	10	90,91%	10	100,00%
T	12	1	8,33%	1	100,00%	11	91,67%	10	90,91%

Tableau 16 : Comparaison de l'utilisation et de l'efficacité du corpus et des ressources en ligne par type d'erreur dans l'EE2G1

On peut conclure l'analyse de l'EE2G1 en rappelant que, dans le questionnaire de fin d'étude, 30% des élèves avaient dit qu'ils se resserraient d'un corpus pour la correction, 55% avaient dit que ce serait peut être le cas et 14% avaient dit que non. Or 40% des élèves se sont servi du corpus pendant l'EE2G1, sans aucune demande de ma part. Ce résultat est encourageant mais il faut rappeler que cette correction a eu lieu dans un contexte scolaire, il serait intéressant de savoir si les étudiants se resserraient du corpus pour faire des corrections chez eux.

5. Conclusion

J'ai débuté cette étude dans l'optique de présenter un nouvel outil de correction à mes élèves afin de générer une plus grande réflexion sur la langue. Je vais reprendre mes questions de recherche afin de présenter mes conclusions.

- **Les élèves arrivent-ils à s'appropriier un nouvel outil (le corpus) avec une introduction très succincte ?**

D'après mes résultats, il semble que la majorité des étudiants a réussi à se servir du corpus. Le taux d'efficacité de celui-ci étant similaire à celui des ressources, on est tenté de dire que l'appropriation est satisfaisante. Cependant, les retours des élèves (mais également la temporalité très réduite de l'étude) suggèrent que l'appropriation est très perfectible. On pourrait alors émettre l'hypothèse que l'efficacité augmenterait et surpasserait plus largement celle des ressources, comme le notent Mueller et Jacobsen :

In sum, qualitative research suggests that learners who receive adequate training can make use of online corpora to effectively remedy errors when revising their writing. Moreover, some recent quantitative research indicates that corpus data as a reference tool can be more effective than conventional resources such as dictionaries or usage manuals. (2016, 5)

De plus, si on regarde le nombre d'élèves s'étant servi du corpus, on peut parler d'appropriation réussie car la grande majorité d'entre eux a utilisé cet outil lors de la première phase de correction avec ce dernier (EE1G1 et EE2G2) et une portion non-négligeable d'entre eux lors de la deuxième phase alors qu'ils avaient accès à tout (EE2G1).

- **Le corpus, quand il est bien employé, permet-il une meilleure correction de la langue ? (si cela est valable, pour quels types d'erreurs ?)**

Si les taux d'efficacité ne permettent pas d'avancer que le corpus est meilleur pour la correction, on peut néanmoins souligner la supériorité avérée du corpus pour la correction d'erreurs de prépositions et de grammaire.

- **La correction avec un corpus permet-elle une plus grande réflexion sur la langue ?**

Les retours des élèves sont encourageants de ce point de vue, ils sont une majorité à penser que le corpus les fait davantage réfléchir. L'attitude des élèves pendant les sessions de correction m'amène également à confirmer cette tendance, ils se posaient beaucoup de questions et certains insistaient pour chercher seuls alors que leurs camarades proposaient de leur donner la bonne réponse. Je ne peux cependant pas me prononcer quant à l'influence du

corpus sur la réflexion linguistique des élèves sur le long terme : est-ce que les élèves qui ont utilisé un corpus vont maintenant être davantage sensibilisés à la langue quand ils écrivent ?

Cette étude me semble avoir été bénéfique pour les élèves, ne serait-ce que du point de vue de leur ouverture linguistique : ils ont appris ce qu'était un corpus et ont pu mener leur « enquête linguistique » le temps des sessions de correction. Ces dernières ont également créé des échanges intéressants entre les élèves. Ceux qui avaient fini leur correction avant les autres s'installaient naturellement avec un élève plus en difficulté et essayaient de l'aider dans ses recherches. J'ai également pu changer de rôle et ne plus agir comme une ressource ou une référence pour les élèves, mais comme un guide, les aidant à aborder les complexités de la langue anglaise :

Toutefois, le fait que l'apprenant ait un accès direct aux corpus inviterait à modifier le rôle de l'enseignant, comme suggéré par Kettemann et Marko : « *the teacher consequently acts as a guide and as an advisor, not as a mediator of knowledge nor a judge of students' performance.* » (2011, 20) (Ciekanski 2014)

6. Références bibliographiques :

6.1. Ouvrages

- Bernardini, S. (2000). Systemizing Serendipity : Proposals for Concordancing Large Corpora with Learners. In *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora*, édité par Burnard, L. et T. McEnery. Hamburg : Peter Lang. 183-190.
- Hamilton, C. (2015). *Cartographie des erreurs en anglais L2 : vers une typologie intégrant système et texte*. Thèse de doctorat, Université Sorbonne Nouvelle-Paris 3. <https://hal.archives-ouvertes.fr/tel-01378302/>.
- Johns, T. (1997). Contexts: The background, development, and trialling of a concordance-based CALL program. In *Teaching and Language Corpora*, édité par A Wichmann et & al, 100–115. London-New York : Longman.
- Kenning, M.-M. (2007). *ICT and Language Learning: From Print to the Mobile Phone*. New York : Palgrave Macmillan.
- Kettemann, B, et G Marko. (2011). Data-driving critical discourse analysis: learning about language and ideology by autonomously exploring a corpus (of creationist literature). In *Corpora, language, teaching, and resources: from theory to practice*, édité par Natalie Kübler, 20–48. Berne : Peter Lang.
- Koo, K. (2006). *Effects of using corpora and online reference tools on foreign language writing: a study of Korean learners of English as a second language*. Thèse de doctorat, University of Iowa. <http://ir.uiowa.edu/etd/65/>.
- Landure, C. (2014). *Étude Comparative de l'Exploitation Directe de Corpus Générique et Spécifique L2 par des Apprenants LANSAD*. Thèse de doctorat, Université Paris 7 – Paris Diderot.
- Littlemore, J, et D Oakey. (2004). Communication with a purpose: exploiting the internet to promote language learning. In *ICT and Language Learning: Integrating Pedagogy and Practice*, édité par A Chambers, J Conacher, et J Littlemore, 95–119. Birmingham, GB : University of Birmingham Press.
- McEnery, T., R. Xiao et Y. Tono. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. Londres : Routledge.
- Palmer, H. E. (1917). *The Scientific Study and Teaching of Languages*. London : Oxford University Press.
- Schaeffer-Lacroix, E. (2009). *Corpus numériques et production écrite en langue étrangère. Une recherche avec des apprenants d'allemand*. Thèse de doctorat, Université de la Sorbonne nouvelle - Paris III. <https://tel.archives-ouvertes.fr/tel-00439095/document>.
- Sinclair, J. M. (2004). New evidence, new priorities, new attitudes. In *How to use corpora in language teaching*, édité par John M Sinclair, 271–99. Amsterdam : John Benjamins.

Tribble, C, et G Jones. (1997). *Concordances in the classroom*. 2ème éd. Houston : Athelstan.

6.2. Articles dans une revue

Boulton, A, et T Cobb. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language Learning*, 67, (2), 348-93.

Boulton, A, et C Landure. (2010). Corpus et autocorrection pour l'apprentissage des langues. *ASp*, 57, 11-30.

Boulton, A, et C Landure. (2016). Using Corpora in Language Teaching, Learning and Use. *Recherche et Pratiques Pédagogiques En Langues de Spécialité*. [En ligne] 35, (2). Mis en ligne le 30 juin 2016, consulté le 13 avril 2018. URL : <http://journals.openedition.org/apliut/5433> ; DOI : 10.4000/apliut.5433

Chambers, A. (2010). L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé. *Revue française de linguistique appliquée*, 15, (2), 9-20.

Chambers, A, et Í O'Sullivan. (2004). Corpus Consultation and Advanced Learners' Writing Skills in French . *ReCALL*, 16, (1), 158-72.

Ciekanski, M. (2014). Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ?. *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, [En ligne] 11, (1). Mis en ligne le 07 janvier 2014, consulté le 13 avril 2018. URL : <http://journals.openedition.org/rdlc/1710> ; DOI : 10.4000/rdlc.1710

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning*, 12, (4), 345-60.

Corder, S.P. (1980). Que signifient les erreurs des apprenants ? *Langages*, 57, 17-28.

Frankenberg-Garcia, A. (2005). A peek into what today's language learners as researchers actually do. *International Journal of Lexicography*, 18, (3), 335-355.

Gaskell, D, et T Cobb. (2004). Can learners use concordance feedback for writing errors? *System*, 32, (3), 301-19.

Johns, T. (1990). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10, 14-34.

———. (1991). Should you be persuaded: Two examples of data-driven learning. *English Language Research Journal*, 4, 1-16.

Kennedy, C, et T Miceli. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5, (3), 77-90.

———. (2017). Cultivating Effective Corpus Use by Language Learners. *Computer Assisted Language Learning*, 30, (1-2), 91-114.

- Kilgarriff, A, et G Grefenstette. (2003). Introduction to the special issue on web as corpus. *Computational Linguistics*, 29, (3), 333-47.
- Lee, D, et J Swales. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25, (1), 56-75.
- Mizumoto, A, et K Chujo. (2016). Who Is Data-Driven Learning for? Challenging the Monolithic View of Its Relationship with Learning Styles. *System*, 61, 55-64.
- Mueller, C M., et N D. Jacobsen. (2016). A Comparison of the Effectiveness of EFL Students' Use of Dictionaries and an Online Corpus for the Enhancement of Revision Skills. *ReCALL*, 28, (1), 3-21.
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19, (3), 269-286.
- O'Sullivan, Í, et A Chambers. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15, (1), 49-68.
- Pérez-Paredes, P, M Sánchez Tornel, J Alcaraz Calero, et P Aguada Jiménez. (2011). Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, 24, (3), 233-53.
- Robb, T. (2003). Google as a quick 'n' dirty corpus tool. *TESL-EJ* [En ligne], 7, (2). Consulté le 13 avril 2018. URL : <http://www.cc.kyoto-su.ac.jp/information/tesl-ej/ej26/int.html>
- Schaeffer-Lacroix, E. (2010). Évolution du système linguistique d'apprenants d'allemand à l'aide de corpus numériques. L'exemple de l'adjectif. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, [En ligne] 13, (13). Mis en ligne le 01 décembre 2010, Consulté le 13 avril 2018. URL : <http://journals.openedition.org/alsic/1678> ; DOI : 10.4000/alsic.1678
- Simard, L-M. (2018). Le corpus comme aide à la rédaction de résumés scientifiques pour des étudiants LANSAD : une approche comparative. *ASp*, 73, 75-104.
- Tyne, H. (2009). Corpus oraux par et pour l'apprenant. *Mélanges CRAPEL*, 31, 91-111.
- Yoon, H. (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12, (2), 31-49.

Liste des tableaux

Tableau 1 :	Nombre d'occurrences de believe et de ses versions erronées _____	6
Tableau 2 :	Chronologie détaillée de l'étude _____	9
Tableau 3 :	Rubriques à remplir par les élèves lors des phases de correction _____	11
Tableau 4 :	Résultats de questionnaire de début d'étude _____	13
Tableau 5 :	Extraits de réponses au questionnaire de début d'étude _____	14
Tableau 6 :	Liste des ressources en lignes par popularité _____	16
Tableau 7 :	Types d'utilisation du corpus par nombre d'élèves et nombre d'occurrences _____	17
Tableau 8 :	Exemples d'utilisations inefficaces du corpus _____	17
Tableau 9 :	Exemples d'utilisations efficaces du corpus _____	18
Tableau 10 :	Résultats du questionnaire de fin d'étude _____	20
Tableau 11 :	Extraits de réponses au questionnaire de fin d'étude _____	20
Tableau 12 :	Données générales liées aux erreurs et à leur correction _____	22
Tableau 13 :	Fréquences d'utilisation et d'efficacité pour chaque type de correction et d'erreur _____	22
Tableau 14 :	Nombre d'élèves utilisant les différentes ressources par groupe et exercice _____	23
Tableau 15 :	Exemple d'utilisation combinée du corpus et d'une ressource _____	23
Tableau 16 :	Comparaison de l'utilisation et de l'efficacité du corpus et des ressources en ligne par type d'erreur dans l'EE2G1 _____	24
Tableau 17 :	Explication du codage des erreurs des élèves _____	31
Tableau 18 :	Répartition et niveau des élèves _____	32
Tableau 19 :	Détail du type d'utilisation du corpus pour chaque élève _____	37

Annexe 1 : Codage d'erreurs

Lettre	Signification
A	Accord
D	Détermination
G	Grammaire (tout ce qui relève de la grammaire et n'est pas déjà associé à une autre lettre)
L	Lexique
O	Ordre
P	Préposition
S	<i>Spelling</i>
T	Temps/Tense

Tableau 17 : Explication du codage des erreurs des élèves

Annexe 2 : Numéros et profils des élèves

Elève	Niveau	Elève	Niveau
101	B2	201	B2
102	B2	202	C1
103	B2	203	B2
104	C1	204	B2
105	B1	205	B2
106	B1	206	B2
107	B1	207	B2
108	B2	208	B1
109	C1	209	B2
110	B1	210	B2
111	B1	211	B2
112	B1	212	C1
113	B1	213	B1
114	B2	214	B1
115	C1	215	B1
116	B1	216	B1
117	B2	217	B1
118	B1	218	B1
119	B2	219	B1

Tableau 18 : Répartition et niveau des élèves⁹

⁹ Le premier chiffre correspond au numéro de groupe. Les élèves de la colonne de gauche font donc partie du Groupe 1 et ceux de la colonne de droite du Groupe 2.

Annexe 3 : Questionnaire de début d'étude

What resources do you use when you correct a WE at home? *

- Paper dictionary all in English
- Paper dictionary English <-> French
- Online dictionary all in English
- Online dictionary English <-> French
- Online translator
- I just correct it by myself
- Other:

Do you always correct your WEs? *

	1	2	3	4	
Always	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Never

What is the longest amount of time you could spend correcting a WE? *

- Less than 10 minutes
- 10 minutes
- 20 minutes
- 30 minutes
- As long as it takes for me to understand everything

Do you take notes (vocabulary, grammar points...) when you use resources to correct your WEs? *

- | | | | | | |
|-------|-----------------------|-----------------------|-----------------------|-----------------------|--------|
| | 1 | 2 | 3 | 4 | |
| Never | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Always |

Do you think correcting WEs helps you learn? Why? *

Your answer

Do you find the letter codes useful? *

- Yes
- No

Do you sometimes struggle when correcting WEs? *

- Yes
- No

If you answered YES before, estimate how often

- | | | | | | |
|--------------|-----------------------|-----------------------|-----------------------|-----------------------|------------|
| | 1 | 2 | 3 | 4 | |
| Almost never | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Very often |

Do you think there is a way you could struggle less? Develop. *

Your answer

Do you know what an online corpus is? *

- Yes
- No

If you answered YES before than explain what it is

Your answer

Annexe 4 : Questionnaire de fin d'étude

Mon avis sur le tutoriel d'introduction au corpus

	1	2	3	4	
Je n'ai rien compris	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Très clair

Avez-vous des suggestions pour améliorer ce tutoriel ?

Your answer

Pour corriger, le corpus a été utile (j'ai trouvé des réponses à mes questions)

	1	2	3	4	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

Le corpus était facile à utiliser

	1	2	3	4	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

Qu'est ce qui a posé problème spécifiquement ?

Your answer

Selon vous, est-ce que le corpus peut être préférable à un autre outil pour corriger une erreur ?

- Oui
- Non

Si oui, dans quel cas ?

Your answer

Le corpus me fait plus réfléchir qu'un autre outil quand je corrige

	1	2	3	4	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

Le corpus m'a appris des choses

	1	2	3	4	
Pas du tout d'accord	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Tout à fait d'accord

Est-ce que vous pensez vous resservir du corpus pour corriger vos expressions écrites si vous êtes en salle info ?

- Oui
- Non
- Peut-être

Annexe 5 : Utilisation détaillée du corpus par élève

	Concordance	Collocation	Collocation et concordance	Concordance Vérification	Collocation Vérification
101	1			2	
102	3	2			
103	1	1		1	
104	1	1			1
105	6	4	1	1	
106	1			1	
107					
108	2		1	2	
109				2	
110	4	1			
111					
112	2	1		3	
113	4	2			
114	2				
115					
116	3	3		3	
117	5	1	1		
118	5				
119	6				
Groupe 1	46	16	3	15	1
201					
202	1			1	
203	1	3			
204	1				
205	1			1	
206	1	2			
207	2				
208	4			1	
209	7	1			
210	2				
211	1				
212		1			
213	3			1	
214					
215	4			1	
216	4				
217	1			1	
218	2				
219	1	1		1	
Groupe 2	36	8	0	7	0
Total	82	24	3	22	1

Tableau 19 : Détail du type d'utilisation du corpus pour chaque élève