HAL open science

Corpus et autocorrection pour l'apprentissage des langues.

Corinne Landure, Alex Boulton

▶ To cite this version:

Corinne Landure, Alex Boulton. Corpus et autocorrection pour l'apprentissage des langues.. ASp - La revue du GERAS, 2010, 57, pp.11-30. 10.4000/asp.931 . hal-00448970v2

HAL Id: hal-00448970

https://hal.science/hal-00448970v2

Submitted on 21 Oct 2010

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.

Corinne Landure & Alex Boulton. 2010. Corpus et autocorrection pour l'apprentissage des langues. *ASp*, 57, p. 11-30. [pre-print version]

Corpus et autocorrection pour l'apprentissage des langues

Corinne Landure et Alex Boulton
Université Nancy 2

MOTS CLÉS
Corpus, *data-driven learning*, auto-correction, traduction.

RESUME
La consultation de corpus n'est pas réservée aux seuls enseignants : elle peut également servir directement aux apprenants, avec des gains en autonomie et en sensibilité linguistique. Toutefois, cette approche, le « data-driven learning » (DDL), n'a connu qu'un succès limité dans des contextes conventionnels, en partie parce qu'elle demanderait des ressources importantes et un investissement considérable de la part des enseignants, et ne conviendrait qu'aux apprenants de niveau avancé après une formation spécifique lourde.

Cette étude met ces critiques à l'épreuve en décrivant une expérience qui initie, sans formation préalable, 43 étudiants en première année d'un Institut Universitaire de Technologie dont le niveau d'anglais est généralement assez modeste, à des activités DDL pour l'autocorrection. Cette étude suggère que l'utilisation des corpus en conjonction avec d'autres ressources (dictionnaires et traducteurs en ligne) permet aux apprenants d'améliorer leurs productions sans intervention de l'enseignant.

KEY WORDS
Corpus, data-driven learning, self-correction, translation.

ABSTRACT
Corpora in language learning are not the sole preserve of teachers: they can also be used directly by learners to promote autonomy and language awareness. However, this "data-driven learning" (DDL) approach has not been widely taken up in traditional teaching contexts, partly because it is thought to require substantial resources and considerable investment on the part of the teachers, as well as being appropriate only for the most advanced learners following specific training.

This paper addresses these issues in describing an experiment with lower-intermediate learners of English, 43 undergraduate students at a technical college in France. The study suggests that the use of a corpus for DDL-style consultation in conjunction with other resources (dictionaries and on-line translation engines) allows these learners to improve written work without help from a teacher.

pre-publication version

Corinne Landure & Alex Boulton. 2010. Corpus et autocorrection pour l'apprentissage des langues. *ASp*, 57, p. 11-30. [pre-print version]

1. Introduction

Le modèle traditionnel de l'enseignement s'inspire de la transmission des savoirs. En ce qui concerne l'enseignement des langues étrangères ou secondes (L2), l'enseignant dispose de savoirs linguistiques qu'il est censé transmettre aux apprenants (Brown *et al*. 2007) ; ceci se fait le plus généralement sous forme d'explications de règles à apprendre puis à appliquer pour démontrer que l'étudiant les a bien apprises, et pour automatiser leur application. Depuis l'apparition de l'approche communicative dans les années 1970, l'accent est également mis sur l'expression et la compréhension écrites et orales, ainsi que sur des savoir-faire et des compétences communicatives ; ces idées sont toujours d'actualité et largement diffusées aujourd'hui grâce notamment au *Cadre européen commun de référence pour les langues* ou CECR (Conseil de l'Europe 2001). L'attention accordée aux compétences communicatives est d'autant plus importante lorsque que nous avons affaire à des langues pour spécialistes d'autres disciplines (LANSAD) : certes, on peut souhaiter « une plus grande spécificité disciplinaire pour l'enseignant des langues sur objectifs spécifiques » (Dlaska 1999 : 402), mais l'enseignant n'est pas toujours en mesure de maîtriser parfaitement la spécialité de ses étudiants pour diverses raisons – parce qu'il intervient dans différentes sections ou dans divers instituts, ou parce que chaque groupe réunit des étudiants de diverses spécialités. Dans ce genre de situation, l'enseignant peut difficilement avoir recours à une « langue de spécialité » particulière : au mieux, il peut reprendre un lexique quasi professionnel commun à un certain nombre de métiers ou de disciplines (pour l'anglais, voir notamment l'*Academic Word List* de Coxhead 2000, 2002). Il peut également travailler des techniques d'apprentissage, ou encore sensibiliser les apprenants aux spécificités de différents types de genres, de registres ou de textes dont l'importance n'est pas à sous-estimer (Biber *et al*. 1998).

De la même façon, l'enseignant peut former les apprenants à mieux se servir des outils dont ils auront besoin ultérieurement. Cette approche est particulièrement importante en LANSAD où l'enseignant ne peut pas fournir tous les éléments de la langue nécessaires aux besoins futurs de ses apprenants, des besoins souvent imprévisibles et ponctuels. Pour ne prendre que des exemples classiques, il faut non seulement savoir écrire une lettre ou un rapport type, préparer une réunion ou un appel téléphonique général, mais aussi être capable de déployer ses savoirs et savoir-faire à une fin très précise lorsque la situation se présente. Pour ce faire, l'utilisateur aura besoin de ressources fiables qu'il saura exploiter de façon efficace et autonome : plutôt que de ne fournir que la langue, l'enseignant fournit également les moyens de s'approprier la langue le moment venu. C'est ainsi que bon nombre de manuels aujourd'hui comprennent des activités pour aider l'apprenant à se repérer dans un dictionnaire ; toutefois, il s'agit généralement de dictionnaires monolingues que l'apprenant utilise très rarement lorsqu'il dispose déjà d'un dictionnaire bilingue (Frankenberg-Garcia 2005), de sorte que les stratégies d'autonomie restent souvent lettre morte.

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 & Oakey 2004 ; Kenning 2007). Si les TIC font partie des sujets de prédilection des manuels, c'est rarement pour sensibiliser les apprenants à leur

pre-publication version

exploitation efficace. Pour bien s'en servir, les apprenants ont besoin de connaître les différents types de ressources disponibles en ligne, surtout (dans une perspective d'usage ultérieur) les ressources gratuites et faciles d'accès. Celles-ci comprennent des dictionnaires électroniques et des traducteurs automatiques qui représentent souvent un premier réflexe pour ces apprenants. Un autre type de ressource reste très peu connu du grand public enseignant comme apprenant : il s'agit des corpus linguistiques. Nous examinons ici les recherches menées à ce jour dans ce domaine, en particulier les études qui portent sur les cas où les apprenants de langue ont accès à des corpus, approche communément appelée le « *data-driven learning* » ou DDL (Johns 1991a, 1991b). Dans un premier temps, nous nous limitons aux applications des corpus comme aide à l'apprentissage, avant de passer à leur exploitation comme ressource disponible à tout moment. Ensuite, nous décrivons une expérience auprès d'un public LANSAD de niveau modeste, comparant leur utilisation d'un grand corpus électronique avec celle de ressources « traditionnelles » en production écrite et en autocorrection. Enfin, les résultats de cette expérience sont analysés ainsi que les réactions des participants.

2. Corpus et apprentissage de langues

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. L'influence des corpus en didactique des langues est difficile à sous-estimer : certains chercheurs parlent même d'une « révolution corpus » qui représente, selon McCarthy (2008 : 564), un véritable « *paradigm shift* » (« révolution scientifique ») en termes kuhniens (1966 [1962]).

Les corpus permettent de multiples applications en didactique des langues, de la description linguistique à la programmation des contenus en passant par les ressources pédagogiques (dictionnaires, grammaires, manuels d'usage), et plus rarement les manuels de cours (voir Boulton [à paraître b] pour un recensement et une discussion). Ils offrent de nouvelles possibilités aux enseignants qui peuvent s'en servir pour vérifier leurs intuitions, ou comme sources d'exemples pour l'enseignement, l'entraînement et l'évaluation de la langue. Les apprenants peuvent les exploiter à leur tour dans ce que Johns appelle le « *data-driven learning* » (DDL) qu'il définit comme « la tentative d'éliminer autant que faire se peut l'intermédiaire pour donner à l'apprenant un accès direct aux données [linguistiques] » (1991b : 30). Dans cette perspective, il ne s'agit plus de transmettre un savoir, mais de permettre aux apprenants de créer leurs propres savoirs de façon inductive. Plutôt que d'apprendre des « règles », ils explorent ainsi la langue à travers des corpus afin de détecter des « *patterns* » - des régularités ou des tendances générales, qu'elles soient linguistiques, pragmatiques, culturelles ou autres. Les corpus mettent ainsi en évidence l'usage de la langue dans des contextes authentiques (surtout les lignes de concordance) et révèlent des informations concernant les collocations, la fréquence et la distribution des items linguistiques, etc.

De nombreux avantages ont été attribués au DDL ; notamment, la découverte interactive d'un corpus devrait aider les apprenants à devenir de « meilleurs

pre-publication version

apprenants » (Johns 1991b : 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). Le processus est centré sur l'apprenant : explorer ses propres questions peut être motivant et mener à une plus grande autonomisation qui lui servira dans sa vie professionnelle ultérieure. Les concordances mettent en évidence les éléments de langue en contextes authentiques et permettent ainsi de réunir le fond (communicatif) et la forme (linguistique). Si les apprenants reconnaissent l'utilité de la grammaire, ils n'apprécient pas forcément son étude (Loewen *et al*. 2009), et cette approche réduit la distance entre leurs attentes et celles des enseignants (Brown 2009). Qui plus est, ce type d'approche pourrait se révéler plus « naturel » que la simple transmission des savoirs dans la mesure où le cerveau humain a évolué pour détecter des régularités dans son environnement, tandis que l'apprentissage de règles représente une activité cognitive intellectuellement contraignante et « artificielle » (Gaskell & Cobb 2004 : 304 ; Scott & Tribble 2006 : 6 ; Boulton 2009a). Il ne s'agit pas de tout abandonner en faveur d'une exploration constructiviste totalement libre (voir Kirschner *et al*. 2006 ; Mayer 2004), mais de fournir un échafaudage pour que les apprenants puissent construire leurs savoirs avec l'aide nécessaire.

Malgré ces avantages, Aijmer (2009 : 2) constate qu'encore aujourd'hui « l'exploitation directe des corpus dans la salle de classe […] reste inhabituelle ». Il est donc clair qu'il existe un certain nombre d'obstacles à l'intégration des corpus en classe (Chambers 2007 ; Farr 2008 ; Boulton 2009b). En particulier, comme le remarque Boulton (à paraître a), l'exploitation de corpus en apprentissage des langues reste très majoritairement aux confins universitaires, avec des apprenants motivés et de niveaux avancés bénéficiant de moyens considérables et des enseignants qui visent une production de recherche. De la même façon, Johns (1991a : 12) signale une objection fréquente de la part des enseignants : le DDL « peut convenir tout à fait à des étudiants aussi intelligents, sophistiqués et motivés que les nôtres [… mais] ne fonctionnerait pas avec des étudiants aussi peu intelligents, peu sophistiqués et peu motivés que les leurs ». Mais il ajoute : « Ce que je soupçonne toutefois c'est que la plupart des étudiants, si on leur donne la possibilité de montrer ce dont ils sont capables, pourraient se révéler (presque) aussi remarquables ». Cependant, les recherches manquent sur ce point : Boulton (2008a) recense une quarantaine d'études qui cherchent à évaluer le DDL sur le terrain ou en situation expérimentale : seule une poignée concerne des apprenants de niveau faible (Cobb 1997a, 1997b, 1999 ; Ciezielska-Ciupek 2001 ; St. John 2001 ; Hadley 2002). Ce déséquilibre dans les recherches consacrées au sujet peut contribuer à renforcer l'impression que le DDL ne convient qu'aux niveaux avancés (Granath 2009).

Il en est de même pour ce qui concerne les ressources utilisées : la plupart des études inventoriées par Boulton (2008a) utilisent des corpus spécialisés créés par les chercheurs eux-mêmes pour un public ciblé, souvent présentés à travers des concordanciers tout aussi spécialisés ou dans des suites d'outils EAO (enseignement assisté par ordinateur) conçus également par les chercheurs. Inévitablement, soit ces ressources sont inaccessibles à d'autres utilisateurs, soit elles coûtent très cher. Encore une fois, on peut en tirer la conclusion que seuls les corpus et les interfaces spécialisés sont efficaces, surtout lorsqu'il s'agit d'apprenants ayant des objectifs spécifiques (Gavioli 2005). Néanmoins, certains ont même utilisé les cent millions

pre-publication version

de mots du *British National Corpus* (BNC) avec succès auprès de leurs apprenants (Bernardini 2000, 2002 ; Cheng *et al*. 2003 ; Götz & Mukherjee 2006). Mais dans tous ces cas il s'agit d'étudiants d'un niveau avancé et spécialisés dans la L2 ; autrement, le BNC sert principalement de corpus de référence pour souligner les particularités de corpus spécialisés (Lee & Swales 2006). Toutefois, le BNC peut être consulté par type de texte et peut, par conséquent, se révéler utile même avec des objectifs assez spécifiques.

Dans ces conditions, il n'est guère surprenant que les enseignants doutent de leurs capacités à s'approprier les techniques et les technologies concernées et peinent à imaginer les usages potentiels auprès de leurs apprenants de niveau plus faible et avec des moyens réduits – grands groupes, manque de temps dans un emploi du temps déjà chargé, accès limité aux ressources informatiques, etc. (Gabrielatos 2005). Mauranen (2004 : 208) résume le problème : « Pour apporter une contribution sérieuse à l'enseignement des langues, les corpus doivent se faire accepter par des enseignants et des apprenants *ordinaires* dans des salles de classe *ordinaires* » (nos italiques). Pour ces raisons, le présent article décrit une expérience menée auprès d'apprenants LANSAD de niveau modeste en anglais et qui n'utilisent que des ressources disponibles gratuitement en ligne, ceci sans formation préalable.

3. Les corpus comme outils de référence : productions écrites, autocorrection, traduction

Cette étude vise à explorer la manière dont les étudiants peuvent se servir d'un corpus pour améliorer leurs productions écrites. En effet, c'est peut-être en tant qu'outils de référence que les corpus se révèlent les plus utiles, mais il faut souligner que s'en servir dans un tel but ne garantit en rien un *apprentissage*, pas plus que de se servir d'un dictionnaire. Cet argument est déployé avec vigueur par Kirschner *et al*. (2006 : 80) dans leur critique sévère des approches constructivistes dans les sciences dures.

Une démarche classique et individualisée est proposée par Johns (1997, 2002) qui, à la suite de longues années de pratique dans ce sens, ouvre ses séances individuelles aux « *kibbitzing* »[1] pour que d'autres puissent les commenter. En bref, les étudiants prennent rendez-vous pour travailler leurs productions écrites ; ils posent des questions, font part des commentaires de leurs enseignants et réfléchissent à l'amélioration de différents points. Johns refuse d'apporter des réponses directement ; en revanche, le corpus reste accessible pendant la séance pour qu'avec l'étudiant ils puissent l'explorer ensemble – formuler des hypothèses, les tester, les améliorer, et ainsi de suite.

Plusieurs recherches proposent des corpus comme aide à l'écriture en classe. Yoon (2008) décrit comment ses apprenants s'en servent parfois au moment de la rédaction, mais surtout pour l'amélioration de leurs productions écrites suite aux commentaires des enseignants, avec des progrès en confiance et en autonomie. En

[1] Ses kibbitzers sur 77 points individuels sont encore disponibles en ligne à : <http://www.eisu2.bham.ac.uk/johnstf/timeap3.htm> ; voir aussi les kibbitzers chez MICASE à <http://lw.lsa.umich.edu/eli/micase/kibbitzer.htm>.

pre-publication version

effet, la quasi-totalité de ces études se concentre sur la révision des écrits, principalement la correction des erreurs signalées par l'enseignant. Turnbull et Burston (1998) indiquent les points à travailler pour deux étudiants : l'un d'entre eux, d'un style cognitif « indépendant du champ » et d'une motivation bien précise, est plus favorable aux corpus et s'en sert plus souvent et de façon plus efficace que l'autre. Todd (2001) demande à ses étudiants de consulter un corpus pour corriger deux erreurs indiquées par l'enseignant, avec un certain succès. Gaskell et Cobb (2004), quant à eux, fournissent au départ des liens vers des concordances présélectionnées pour des erreurs identifiées ; dans un deuxième temps, les erreurs sont indiquées mais les étudiants doivent eux-mêmes chercher les informations nécessaires dans le corpus pour se corriger. Si un tiers d'entre eux deviennent ainsi relativement indépendant dans la consultation de corpus, dans l'ensemble ils préfèrent que les concordances leur soient fournies. De la même façon, Chambers et O'Sullivan (2004) et O'Sullivan et Chambers (2006) soulignent les erreurs à corriger, mais fournissent une analyse quantitative, rare dans le domaine. Elles trouvent des améliorations grâce aux corpus pour une grande variété d'éléments linguistiques, même pour des points aussi simples que le genre qui auraient pu être retrouvées facilement dans un dictionnaire.

Afin de préparer les étudiants à un travail en autonomie, Sun (2007) a développé ses *Scholarly Writing Templates* qui guident les utilisateurs lors de la rédaction d'un article de recherche, à la fois sur la structure discursive et sur la langue, et qui obtiennent des réactions généralement positives. Milton (2006) fait référence aussi à de nombreux outils informatiques (dont les corpus) pour aider les apprenants à produire et à réviser leurs écrits. Mais il semble peu probable que ce genre de logiciel soit à la portée de tous sans une formation approfondie, et convienne à ceux qui ont des besoins bien définis pour l'expression écrite. Les corpus ont comme avantage, une fois que l'on sait s'en servir, de ne pas être limités à une seule fin mais d'offrir de multiples possibilités d'exploitation (les « *affordances* » de Hafner et Candlin [2007]). Certes, une formation et de l'expérience sont indispensables pour en tirer le meilleur parti possible ; mais les difficultés sont souvent surestimées (Bernardini 2001 : 243) alors qu'en fait on peut commencer à s'en servir presque immédiatement (Sinclair 2004 : 288). Thomas (2008) rapporte que ses étudiants réussissent à en bénéficier sans formation préalable, s'appuyant sur les résultats de plusieurs expériences récentes (Boulton 2007, 2008b, 2009a).

Notre étude prend un autre point de départ afin de permettre aux apprenants de corriger leurs écrits sans intervention de la part de l'enseignant. Cette approche serait difficile avec les productions individuelles longues qu'on voit dans la plupart des études citées ci-dessus. Seul Sun (2003) propose un nombre limité de phrases courtes à ses étudiants qui ont la tâche d'identifier et de corriger les erreurs qu'elles contiennent en se servant de corpus. Toutefois, il propose toujours une aide considérable aux trois participants : en effet, son rôle de guide et de conseiller pour développer l'autonomie des étudiants est l'un des facteurs déterminants de cette expérience.

Comme Sun (2003), nous commençons avec un petit nombre de phrases à travailler ; mais notre démarche est quelque peu différente parce que nous avons recours à la traduction, autre élément souvent sous-exploité dans les études de corpus auprès du grand public apprenant. En effet, bien que de nombreuses recherches portent sur l'utilisation de corpus pour la traduction, ces travaux sont

pre-publication version

menés uniquement auprès de publics spécialisés dans la L2, quasi exclusivement de futurs traducteurs. Dans la plupart des cas, il s'agit de corpus bilingues ; à des fins de traduction de spécialité, Maia (1997) demande même à ses étudiants de créer leurs propres corpus – des corpus parallèles ou comparables ou encore des glossaires bilingues – pour les aider dans une tâche de traduction spécifique. Bernardini (2002) commence avec le BNC avant de passer aux corpus parallèles et obtient des appréciations positives de la part de ses apprenants. Bowker (1998, 1999) s'en sert pour la traduction vers la langue maternelle (L1) de ses étudiants, mais là encore elle trouve moins d'erreurs dans chaque catégorie analysée ainsi que des réactions plutôt positives. Toutefois, même ces futurs traducteurs ont tendance à faire davantage confiance aux dictionnaires qu'aux corpus. Frankenberg-Garcia (2005), elle aussi, constate que les étudiants préfèrent souvent les dictionnaires aux corpus, qui ne représentent que 9 % (corpus monolingues) et 2 % (bilingues) des consultations effectuées par ses apprentis traducteurs parmi l'ensemble des ressources disponibles. Il est toutefois intéressant de noter que les outils choisis en priorité (les dictionnaires) donnent les moins bons résultats parmi l'ensemble des ressources utilisées, tandis que les corpus sont utiles dans deux tiers des cas. Ceci pourrait refléter la tendance qu'ont des apprenants de niveau avancé dans n'importe quelle discipline à préférer une approche et des informations structurées, même si ce n'est pas ce qu'il y a de plus bénéfique pour eux (Kirschner *et al*. 2006 : 82).

Lorsque l'objectif spécifique n'est pas la traduction, le recours à la L1 reste controversé pour certains enseignants. Certes, il serait peu souhaitable d'imposer aux apprenants des liens systématiques entre la L2 et la L1 ; mais interdire tout usage de la L1, surtout aux niveaux plus modestes, priverait les apprenants de leur arme principale pour aborder la L2. Par ailleurs, une comparaison directe avec la L1 peut mettre en évidence certaines particularités de la L2 et ainsi mener à une plus grande sensibilité linguistique (Johns *et al*. 2008). En outre, la traduction représente un besoin réel dans de multiples contextes professionnels, comme le souligne une récente enquête menée par l'IFOP auprès de cadres d'entreprises privées en France. Lorsqu'ils ont affaire à une langue étrangère, les premières ressources pour 44 % d'entre eux sont les traducteurs automatiques – et pour 83 % leur « expérience de la traduction automatique [est] limitée aux portails de traduction gratuits sur Internet » (IFOP 2009 : 3). Plutôt que d'interdire tout recours à la traduction, il semble donc plus raisonnable de l'exploiter en cours afin de la canaliser pour éviter l'abus et un certain nombre de « pièges ». De toute façon, il semble clair que les apprenants continueront à traduire même face à une interdiction totale (Decoo 2001).

Enfin, il convient de rappeler que presque toutes les recherches citées ici ont été menées dans des contextes universitaires auprès d'étudiants en troisième cycle d'études et de niveaux relativement élevés, le plus souvent des spécialistes de la langue cible. Rares sont les études en DDL qui portent sur des apprenants de niveaux plus faibles, mais les résultats ne sont pas pour autant moins probants que pour les plus avancés (Tian 2005 ; Boulton 2007, 2008b, 2009a, à paraître a). Qui plus est, Yoon et Hirvela (2004) obtiennent de meilleurs résultats avec des apprenants d'un niveau moins élevé car ce sont eux qui ont tout à gagner d'une nouvelle approche. De plus Lee et Liou (2003) notent que l'écart entre les plus forts et les plus faibles s'estompe suite à un travail sur corpus qui, selon eux, constitue lui-même une sorte d'échafaudage. Comme le remarque Boulton (2009a : 38), « il se peut que le DDL soit plus approprié pour des apprenants ayant 'moins bien

pre-publication version

réussi' comme alternative aux méthodes classiques qui, par définition, n'ont pas été efficaces pour eux ».

4. Méthodologie

Cette étude vise donc à initier des étudiants LANSAD d'un niveau modeste à l'exploitation d'un corpus d'anglais avec un minimum de formation préalable. Le but principal est d'évaluer dans quelle mesure un corpus peut aider à l'autocorrection par le biais de la traduction, sans intervention d'un enseignant.

L'expérience s'est déroulée dans un Institut Universitaire de Technologie (IUT) en France avec la participation de 43 étudiants en première année d'études de logistique et maintenance, répartis en quatre groupes de onze apprenants indépendamment de leur niveau en anglais. Bien qu'ils aient quatre heures d'anglais par semaine, soit une centaine d'heures à l'année, ce public, principalement masculin (86 %) et âgé d'une vingtaine d'années, a le sentiment de ne pas progresser de façon satisfaisante, notamment à l'écrit. En effet, malgré plusieurs années d'anglais scolaire pour la plupart d'entre eux, leur niveau reste relativement faible comme l'indiquent les résultats à un test informel du TOEIC[2] effectué en début d'année. Leur score moyen atteint 380 points sur 990, ce qui correspond approximativement au niveau A2 du CECR, alors que le niveau requis en anglais par de nombreux employeurs dans ce domaine se situe entre 650 et 750 points. Comme nous l'avons vu ci-dessus, la consultation de corpus peut être utile même à des niveaux modestes ; comme pour tout usage de document authentique, il convient de sélectionner un support approprié et de proposer des tâches en rapport avec le niveau linguistique des apprenants.

Pour cette expérience, chaque étudiant disposait d'un dictionnaire bilingue ainsi que d'un ordinateur équipé d'un accès Internet pour les outils de traduction et les corpus en ligne. Le dictionnaire utilisé était le *Harrap's poche* qui compte 75 000 entrées et 110 000 traductions ; celui-ci avait été choisi car on disposait d'un exemplaire pour chaque étudiant et par conséquent il leur était familier[3]. Les traducteurs proposés comprenaient les sites de *Reverso*, *Google*, et *BabelFish*, tous gratuits en ligne et faciles à manipuler[4]. Le corpus de référence était le *British National Corpus*, soit 100 millions de mots d'anglais britannique parlé et écrit du début des années 1990, le plus grand projet de son époque représentant une collaboration entre plusieurs universités britanniques et des maisons d'édition privées. De nos jours, il reste certainement le corpus le plus soigneusement conçu et réalisé de cette ampleur (Leech 2007). Le BNC est disponible gratuitement en ligne grâce à l'interface créée par Mark Davies ; celle-ci offre de nombreuses fonctionnalités, mais les requêtes de bases permettent une utilisation immédiate[5]. Il s'agit bien d'un corpus d'anglais général, mais l'interface permet de consulter des

---

[2] *Test of English for International Communication* : <http://www.toeic.eu/>. Bien connu par les employeurs en France, une réussite à ce test est un atout sur un CV et est fortement recommandé, voire obligatoire, dans de nombreuses institutions supérieures.

[3] *Harrap's anglais : dictionnaire poche anglais-français / français-anglais*. 2008. Paris : Chambers.

[4] <http://www.reverso.net/text_translation.asp?lang=FR>, <http://translate.google.fr>, <http://babelfish.yahoo.com/>.

[5] *BYU-BNC : British National Corpus*. <http://corpus.byu.edu/bnc/x.asp>.

pre-publication version

sous-corpus en forme de registres « macro » (par ex. ACADEMIC) ou « micro » (par ex. WRITTEN ACADEMIC TECHNICAL ENGLISH). Par ailleurs, de nombreuses recherches indiquent que ce type de grand corpus général peut servir d'entraînement avant de s'engager dans un travail sur corpus plus spécialisés (notamment Gavioli 2005 ; Aston 1997 ; Bernardini 2000, 2005). Puisqu'il ne s'agit pas d'apprenants expérimentés ou de niveau avancé, les activités et les points de langue à étudier doivent rester très simples. La tâche consiste à traduire en anglais dix phrases portant sur des sujets étudiés en cours quelques mois auparavant. Celles-ci ne représentent pas un « anglais de spécialité » dans la mesure où elles ne sont pas liées à un domaine d'études particulier, mais reflètent un besoin spécifique pour ces étudiants puisqu'elles sont représentatives du TOEIC dans leur contextualisation et le type de langage.

Chaque étudiant doit consulter les mêmes outils et suivre la même procédure : une première traduction sans ressources ; une deuxième avec un dictionnaire bilingue ; une troisième avec un traducteur en ligne. Ces différentes étapes permettent aux étudiants de comparer les résultats des traductions réalisées à l'aide des différentes ressources, de combler quelques lacunes et de corriger certaines erreurs. Mais lorsque le dictionnaire et les traducteurs divergent dans leurs réponses, les étudiants n'ont bien souvent d'autre alternative qu'un choix arbitraire ou intuitif. C'est alors qu'ils sont invités à utiliser un corpus pour résoudre ces problèmes, après en avoir reçu une brève présentation. Cette approche laisse à l'initiative des étudiants le choix des items à consulter ; le travail sur corpus permet ainsi une véritable autocorrection. En effet, l'une des conséquences de la consultation de corpus est de modifier les rôles traditionnels. Dans un premier temps, « l'étudiant accepte plus de responsabilité pour son apprentissage, l'enseignant assume le rôle d'un directeur de recherche ou collaborateur plutôt qu'un transmetteur de savoir » (Johns 1988 : 14). Si l'apprenant peut trouver des réponses aux questions qu'il se pose dès maintenant, il y a alors toutes les raisons de penser qu'il pourra continuer à s'en servir plus tard dans sa vie professionnelle lorsque des besoins concrets se présenteront en situations réelles où il ne pourra compter que sur lui-même et non plus sur l'enseignant pour y répondre.

Le tableau 1 rend compte de la production de l'un des étudiants pour l'une des phrases à traduire : « Je vais reporter mon rendez-vous, la réunion commence dans une minute ». La première traduction à partir de ses seuls savoirs actuels révèle un certain nombre de lacunes (« reporter » et « réunion »), ainsi que quelques problèmes de vocabulaire (« rendez-vous »). La consultation du dictionnaire (étape 2) puis l'utilisation du traducteur automatique (étape 3) permettent de combler ces lacunes ; toutefois, les résultats provenant de ces ressources ne sont pas toujours identiques, par conséquent certains doutes peuvent subsister. L'étudiant choisit alors d'utiliser le corpus pour vérifier la traduction de ces différents éléments.

pre-publication version

Tableau 1. Exemple de production d'un étudiant

	Je vais reporter	mon rendez-vous	la réunion commence dans une minute
1 (savoir)	I go ***	my date	the *** begin in one minute
2 (dictionnaire)	I go report	my appointment	the meeting begin in one minute
3 (traducteur)	I'm going to postpone	my meeting	the meeting begins in one minute
CORPUS	**report**	**appointment**	**meeting**
4 (final)	I'm going to **report**	my **appointment**	the **meeting** begins in one minute

L'étude des items sélectionnés pour consultation dans le corpus et l'analyse des résultats de la traduction finale permettent de repérer les items que les apprenants ont consultés et les changements qui en ont résulté. L'étude est complétée par un questionnaire composé de treize questions ouvertes et fermées. Celui-ci nous fournit un aperçu de l'appréciation des étudiants concernant la facilité, l'utilité et l'apport de la consultation de corpus. L'ensemble de ces données nous renseigne sur l'efficacité linguistique de la consultation, les compétences développées et les stratégies mises en place par les étudiants au fur et à mesure de l'utilisation du corpus.

5. Résultats

Pendant la phase de découverte, une certaine déstabilisation a été observée chez tous les étudiants, en partie à cause de la nouveauté de l'outil et en partie à cause du comportement de l'enseignante qui n'est jamais intervenue spontanément et qui s'est abstenue de fournir des réponses linguistiques complètes aux étudiants pendant l'expérience. Son rôle a plutôt été de les guider, de leur conseiller de se poser des questions pertinentes et exploitables, de les inciter à trouver les solutions par eux-mêmes et de les encourager à partager leurs problèmes et découvertes au sein du groupe. Devant la nouveauté de cette approche, les étudiants se sont trouvés tous pour une fois au même niveau : la répartition en « bons » et en « mauvais » étudiants des cours traditionnels avait disparu. D'ailleurs, cela n'a pas échappé à certains : les habituellement « faibles » ont fait preuve de beaucoup d'enthousiasme et se sont révélés être des preneurs de risques alors que les étudiants habituellement « bons » ont manifesté une plus grande réserve.

Au fil de leurs découvertes, les étudiants ont commencé à se communiquer leurs résultats et à échanger astuces, conseils et pièges à éviter, ce qui a accru leur motivation. Ils ont par la suite collaboré activement à l'élaboration de stratégies appropriées à ces activités, ce qui les a conduits à mieux prendre conscience du fonctionnement de la langue. En général, ces observations notées pendant l'expérience révèlent une bonne implication des étudiants en général et une fierté dans le partage d'informations et de conseils.

5.1. Résultats linguistiques

Deux heures ont été consacrées à la consultation du corpus et 420 consultations ont été comptabilisées pour les 43 étudiants, ce qui donne 9,8 items consultés par étudiant et environ 12 minutes par item consulté. Ces consultations ont été

pre-publication version

analysées en fonction de la nature des changements opérés dans leurs traductions, comme nous pouvons le voir dans le tableau 2 ci-dessous :

- des modifications « positives » qui, grâce à la consultation du corpus, ont abouti à une traduction adéquate alors que l'étudiant avait trouvé une solution erronée dans l'une des autres ressources utilisées (dictionnaire ou traducteur automatique) : par ex. *appointment* ;
- des modifications « négatives » qui, malgré la consultation du corpus, ont abouti à une traduction erronée alors que l'étudiant avait trouvé une solution adéquate dans l'une des autres ressources utilisées (dictionnaire ou traducteur automatique) : par ex. *report* ;
- des modifications « zéro » où, suite à la consultation de corpus, l'étudiant a conservé les réponses communes fournies par les autres ressources ; par ex. *meeting*.

Tableau 2. Exemple de comptabilisation des modifications

dictionnaire →	traducteur →	corpus	modification
appointment	meeting	appointment	positive
report	postpone	report	négative
meeting	meeting	meeting	zéro

Sur ces 420 consultations, 230 n'ont pas entraîné de modification, 121 ont enregistré une modification positive et 69 une modification négative. Le taux considérablement élevé de modifications zéro (environ 55 %) pourrait indiquer que les étudiants n'ont pas su cibler leurs recherches ou encore qu'ils ont eu besoin d'une confirmation supplémentaire par rapport aux résultats des autres ressources et que, dans ce cas, la consultation du corpus les a confortés dans leurs décisions. Les 30 % de modifications positives représentent un résultat encourageant, portant principalement sur les expressions idiomatiques et les unités lexicales composées qui sont difficilement repérables dans les dictionnaires ou traducteurs en ligne et que les étudiants ont souvent tendance à traduire littéralement. À l'origine des 16 % de modifications négatives, certaines sont probablement non intentionnelles, mais, pour la plupart, elles résultent de l'emploi d'expressions erronées ou inadaptées au contexte. Ceci pourrait signifier que les étudiants ont des difficultés de compréhension et de maîtrise des expressions idiomatiques et que, si la consultation d'un corpus peut aider à mieux les appréhender, certaines restent apparemment plus difficiles à localiser.

Afin de vérifier si le niveau des étudiants a une incidence sur leur pratique de la consultation du corpus, nous les avons répartis en deux groupes selon leur score au TOEIC, supérieur ou inférieur à la moyenne de 380 points. Le tableau 3 ne révèle pas de différences majeures dans le nombre moyen d'items consultés par étudiant ni le temps moyen passé pour chaque item consulté. Afin de corroborer ces résultats, nous avons comparé nos données avec celles de Chambers et O'Sullivan (2004). Leurs étudiants avaient non seulement un niveau avancé mais étaient aussi spécialistes de la langue ; par ailleurs, ils consultaient un corpus spécialisé après une formation approfondie en vue d'améliorer leur production textuelle dont les imperfections avaient préalablement été mises en évidence par leurs enseignants. Malgré ces différences, les résultats sont encore une fois comparables. On peut

pre-publication version

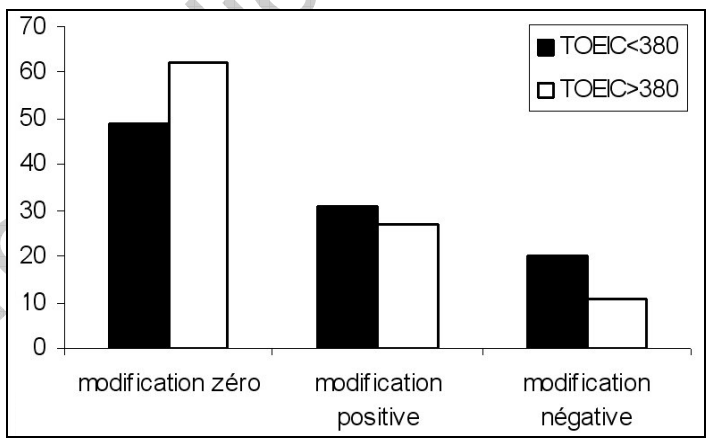
alors avancer que, dans ce cas précis, le niveau des étudiants n'influe ni sur le temps moyen passé par item consulté ni sur le nombre moyen d'items consultés par étudiant.

Tableau 3. Items recherchés par niveau

	TOEIC <380	TOEIC >380	Chambers & O'Sullivan
moyenne	275	487	–
étudiants	25	18	8
durée	2h	2h	1h40
items	257	163	85
items / étudiant	10,3	9,06	10,6
minutes / item	11,7	13,2	9,4

Si les comportements sont relativement homogènes quel que soit le niveau, qu'en est-il des résultats de leurs recherches ? Il apparaît que les étudiants ayant un score supérieur à 380 ont fait moins de modifications (positives ou négatives) mais ont effectué davantage de consultations sans modification (figure 1). Une explication pourrait être qu'ils ont plus de connaissances et par conséquent moins de modifications à apporter ; dans ce cas, la consultation du corpus leur permet simplement de confirmer leurs intuitions ou les suggestions des autres ressources. Dans tous les cas, le taux de modifications positives est encourageant et indique que même des étudiants de niveau modeste en anglais peuvent bénéficier de la consultation de corpus.

Figure 1. Modifications par niveau



5.2. Résultats du questionnaire

Dans le cadre de notre étude, les apprenants n'avaient jamais utilisé de corpus en tant que tel avant cette expérience et ils n'ont bénéficié que d'une brève présentation préalable. 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 ».

pre-publication version

Par ailleurs, 79 % d'entre eux considèrent que la consultation de corpus est « plutôt utile » ou « très utile » ; 53 % pensent même qu'elle est aussi utile que la traduction automatique, ce qui est particulièrement intéressant quand on sait que les traducteurs automatiques étaient la ressource préférée de 95 % de nos étudiants avant l'expérience. En outre, 93 % d'entre eux estiment que l'exploitation directe d'un corpus peut les aider dans la correction de leurs erreurs, ce qui met en évidence l'utilité d'un corpus en tant qu'outil de référence. Enfin, 93 % des participants souhaitent utiliser un corpus dans le cadre de leurs études ou carrières professionnelles futures.

Les réactions positives sont peut-être révélatrices d'un changement profond des attitudes. Néanmoins, nous devons faire preuve de circonspection face à cette majorité enthousiaste, puisqu'il reste 7 % de participants qui ont répondu de façon négative à l'ensemble du questionnaire, des étudiants qui ne voient aucun avantage dans la consultation de corpus et qui n'envisagent pas de l'utiliser plus tard.

6. Discussion

Cette étude a essayé de mettre en pratique les techniques du *data-driven learning* dans un contexte ordinaire : c'est-à-dire, l'exploitation d'un grand corpus gratuit directement en ligne, sans formation préalable, par des apprenants LANSAD d'un niveau modeste. Cette approche semble avoir de multiples applications et bénéfices même dans ces circonstances et elle souligne les avantages multiples, tant linguistiques qu'individuels et collaboratifs, de la consultation de corpus. La séquence d'activités adoptée est, de nécessité, quelque peu artificielle. Toutefois, elle montre que les apprenants peuvent se servir d'un corpus pour une autocorrection sans l'intervention de l'enseignant, une possibilité absente dans toutes les autres études portant sur l'exploitation d'un corpus comme outil de référence. De cette façon, les étudiants ne sont plus passifs face à leurs erreurs : ils s'entraînent activement à repérer les éléments à revoir pour ensuite essayer de les corriger grâce au corpus. Cette double implication personnelle de la part des étudiants, à la fois active et réfléchie, a entraîné des améliorations de leurs travaux à l'écrit. Le corpus leur permet ainsi un accès aux « exemples » si importants pour bon nombre d'apprenants (Loewen *et al.* 2009 : 100). Si le corpus représente alors un « informant » linguistique (Johns 1991a : 1), le rôle principal de l'enseignant consiste par conséquent à rendre l'étudiant capable de l'exploiter de façon autonome afin qu'il soit en mesure de répondre à ses futurs besoins professionnels, lorsqu'il aura quitté l'université et qu'il ne pourra plus recourir à l'aide de l'enseignant.

L'analyse de l'évolution des traductions dans cette expérience indique une amélioration sensible des compétences écrites des étudiants ainsi que de leur capacité à l'autocorrection, en particulier des expressions idiomatiques. De plus, leurs réponses à un questionnaire font ressortir une évaluation positive quasi unanime concernant la consultation de corpus, ce qui laisse supposer de possibles utilisations futures dans le cadre de leurs études ou carrière professionnelle, à condition bien entendu que l'expérience ne s'arrête pas là (Hafner & Candlin 2007). L'observation en classe et les commentaires des participants révèlent que les étudiants ont collaboré activement à l'élaboration de stratégies appropriées à ces activités, menant à une meilleure prise de conscience du fonctionnement de la

pre-publication version

langue et à des améliorations appréciables de leurs productions ; il semble alors que les corpus peuvent constituer des ressources de référence utiles. Cette étude a également mis en relief le rôle de l'individualisation et de l'autonomisation de cette approche, avec la possibilité pour les étudiants de cibler leurs recherches en fonction de leurs besoins individuels, en leur permettant de développer des compétences autonomes dans la gestion de leur apprentissage. Enfin, cette expérience souligne l'importance d'une collaboration active dans la mise en place de stratégies appropriées, favorisant ainsi la motivation des étudiants, facteur essentiel pour la réussite et la pérennité de cette approche.

Il semble évident que, sans formation, les étudiants ne tireront pas le meilleur parti possible des corpus ; avec un entraînement plus régulier, on peut supposer que les résultats seront plus probants. 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. En effet, l'utilisation d'un corpus n'est pas une activité complètement nouvelle : selon McCarthy (2008 : 56), « nous sommes tous [déjà] des utilisateurs de corpus, puisque nous utilisons Internet. Internet n'est rien d'autre qu'un immense corpus ». Si cette position reste controversée (Adolphs 2006 : 33), il semble néanmoins clair qu'une familiarité avec les moteurs de recherche sur Internet peut servir de tremplin à l'exploitation d'un corpus proprement dit. En effet, dans cette expérience nous constatons que le manque de formation spécifique n'empêche pas les étudiants « ordinaires » d'accéder aux corpus même s'ils s'y essaient pour la première fois.

Il est important également de rappeler que les participants à cette expérience ne sont pas des spécialistes de langue, n'ont pas une motivation particulière pour ce domaine, et ont un niveau relativement faible en anglais. Pour des publics de ce type, toute étude formelle peut être perçue comme « ennuyeuse » (Loewen *et al.* 2009 : 101), tandis que toute approche nouvelle peut augmenter la motivation, même si celle-ci risque de disparaître rapidement (en particulier, le simple fait de travailler sur ordinateur ne garantit en rien une motivation à long terme, Jarvis [2004]). Les étudiants ne trouveront la motivation de continuer que s'ils estiment l'approche intéressante et efficace, c'est-à-dire s'ils découvrent facilement les réponses aux questions qu'ils se posent. Pour ce faire, il sera essentiel de passer à un corpus spécialisé (Gavioli 2005). Pour leur premier contact avec les corpus dans cette étude, nous nous sommes limités à un grand corpus de langue générale qui offre plusieurs avantages. Tout d'abord, il répond tout simplement aux besoins et objectifs à court terme des étudiants concernés et, de par sa taille, il leur offre plus de chances de trouver des réponses à leurs questions. Ensuite, l'interface est simple d'utilisation et l'ensemble est disponible gratuitement en ligne. Ce ne sont pas des critères négligeables : non seulement un grand corpus général évite les difficultés inhérentes à la création d'un corpus spécifique, mais les apprenants pourront retrouver facilement le même corpus plus tard. Par ailleurs, l'exploitation d'un grand corpus peut servir d'introduction à l'utilisation ultérieure de corpus plus spécialisés (Aston 1996). Pour faire découvrir le *data-driven learning* à un public plus large (le rôle de « missionnaire » préconisé par Römer [2009]), il faut commencer par des outils faciles, gratuits et offrant de multiples « *affordances* ».

pre-publication version

Corinne Landure & Alex Boulton. 2010. Corpus et autocorrection pour l'apprentissage des langues. *ASp*, 57, p. 11-30. [pre-print version]

Références bibliographiques

ADOLPHS, S. 2006. *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*. Londres : Routledge.

AIJMER, K. (dir.) 2009. *Corpora and Language Teaching*. Amsterdam : John Benjamins.

ASTON, G. 1996. « The British National Corpus as a language learner resource ». In S. BOTLEY, J. GLASS, A. MCENERY et A. WILSON (dir.) *Proceedings of TALC 1996. UCREL Technical Papers* 9, 178-191.

ASTON, G. 1997. « Small and large corpora in language learning ». In B. LEWANDOWSKA-TOMASZCZYK et J. MELIA (dir.) *Practical Applications in Language Corpora*. Lodz : Lodz University Press, 51-62.

BERNARDINI, S. 2000. « Systematising serendipity: proposals for concordancing large corpora with language learners ». In L. BURNARD et T. MCENERY (dir.) *Rethinking Language Pedagogy from a Corpus Perspective*. Francfort : Peter Lang, 225-234.

BERNARDINI, S. 2001. « 'Spoilt for choice.' A learner explores general language corpora ». In G. ASTON (dir.) *Learning with Corpora*. Houston, TX : Athelstan, 220-249.

BERNARDINI, S. 2002. « Exploring new directions for discovery learning ». In B. KETTEMANN et G. MARKO (dir.) *Teaching and Learning by Doing Corpus Analysis*. Amsterdam : Rodopi, 165-182.

BIBER, D., S. CONRAD et R. REPPEN. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge : Cambridge University Press.

BOULTON, A. 2007. « DDL is in the details… and in the big themes ». *Proceedings of 4th Corpus Linguistics Conference*. <http://ucrel.lancs.ac.uk/publications/CL2007/>.

BOULTON, A. 2008a. « Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues ». *Texte et corpus* 3, 37-46.

BOULTON, A. 2008b. « Looking for empirical evidence of DDL at lower levels ». In B. LEWANDOWSKA-TOMASZCZYK (dir.) *Corpus Linguistics, Computer Tools, and Applications – State of the Art*. Francfort : Peter Lang, 581-598.

BOULTON, A. 2009a. « Testing the limits of data-driven learning: language proficiency and training ». *ReCALL* 21/11, 37-51.

BOULTON, A. 2009b. « Data-driven learning: reasonable fears and rational reassurance ». *CALL in Second Language Acquisition: New Approaches for Teaching and Testing. Indian Journal of Applied Linguistics* 35/1, 81-106.

BOULTON, A. À paraître a. « Data-driven learning: taking the computer out of the equation ». *Language Learning* 60/3.

BOULTON, A. À paraître b. « Data-driven learning: on paper, in practice ». In T. HARRIS et M. MORENO JAÉN (dir.) *Corpus Linguistics in Language Teaching*. Berne : Peter Lang.

BOWKER, L. 1998. « Using specialised monolingual native-language corpora as a translation resource: a pilot study ». *Meta* 43/4, 631-651.

BOWKER, L. 1999. « Exploring the potential of corpora for raising language awareness in student translators ». *Language Awareness* 8/3-4, 160-173.

BROWN, A. 2009. « Students' and teachers' perceptions of effective foreign language teaching: a comparison of ideals ». *Modern Language Journal* 93/1, 46-60.

BROWN, H., E. TARONE, M. SWAN, R. ELLIS, L. PRODROMOU, U. Jung, A. BRUTON, K. JOHNSON, D. NUNAN, R. OXFORD, C. GOH, A. WATERS et S. SAVIGNON. 2007. « Forty years of language teaching ». *Language Teaching* 40, 1-15.

CHAMBERS, A. 2007. « Popularising corpus consultation by language learners and teachers ». In E. HIDALGO, L. QUEREDA et J. SANTANA (dir.) *Corpora in the Foreign Language Classroom*. Amsterdam : Rodopi, 3-16.

pre-publication version

CHAMBERS, A. et I. O'SULLIVAN. 2004. « Corpus consultation and advanced learners' writing skills in French ». *ReCALL* 16/1, 158-172.

CHENG, W., M. WARREN et X. XUN-FENG. 2003. « The language learner as language researcher: putting corpus linguistics on the timetable ». *System* 31/2, 173-186.

CIEZIELSKA-CIUPEK, M. 2001. « Teaching with the internet and corpus materials: preparation of the ELT materials using the internet and corpus resources ». In B. LEWANDOWSKA-TOMASZCZYK (dir.) *PALC 2001: Practical Applications in Language Corpora*. Francfort : Peter Lang, 521-531.

COBB, T. 1997a. *From Concord to Lexicon: Development and Test of a Corpus-Based Lexical Tutor*. Thèse de doctorat. Montréal : Concordia University. <http://www.er.uqam.ca/nobel/r21270/webthesis/Thesis0.html>.

COBB, T. 1997b. « Is there any measurable learning from hands-on concordancing? » *System* 25/3, 301-315.

COBB, T. 1999. « Breadth and depth of lexical acquisition with hands-on concordancing ». *Computer Assisted Language Learning* 12/4, 345-360.

CONSEIL DE L'EUROPE, 2001. *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Paris : Didier.

COXHEAD, A. 2000. « A new academic word list ». *TESOL Quarterly* 34/2, 213-238.

COXHEAD, A. 2002. « The academic word list: a corpus-based word list for academic purposes ». In B. Kettemann et G. Marko (dir.) *Teaching and Learning by Doing Corpus Analysis*. Amsterdam : Rodopi, 73-89.

DECOO, W. 2001. « On the mortality of language learning methods ». *L. Barker Lecture*. Provo, UT : Brigham Young University, 8 novembre. <http://www.didascalia.be/mortality.htm>.

DLASKA, A. 1999. « Suggestions for a subject-specific approach in teaching foreign languages to engineering and science students ». *System* 27/3, 401-417.

FARR, F. 2008. « Evaluating the use of corpus-based instruction in a language teacher education context : perspectives from the users ». *Language Awareness* 17/1, 25-43.

FRANKENBERG-GARCIA, A. 2005. « A peek into what today's language learners as researchers actually do ». *International Journal of Lexicography* 18/3, 335-355.

GABRIELATOS, C. 2005. « Corpora and language teaching: just a fling or wedding bells? » *Teaching English as a Second Language – Electronic Journal* 8/4, 1-35. <http://tesl-ej.org/ej32/a1.html>.

GASKELL, D. et T. COBB. 2004. « Can learners use concordance feedback for writing errors? » *System* 32/3, 301-319.

GAVIOLI, L. 2005. *Exploring Corpora for ESP Learning*. Amsterdam : John Benjamins.

GÖTZ, S. et J. MUKHERJEE. 2006. « Evaluation of data-driven learning in university teaching: a project report ». In S. BRAUN, K. KOHN et J. MUKHERJEE (dir.) *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Francfort : Peter Lang, 49-67.

GRANATH, S. 2009. « Who benefits from learning how to use corpora ». In K. AIJMER (dir.) *Corpora and Language Teaching*. Amsterdam : John Benjamins, 47-65.

HADLEY, G. 2002. « Sensing the winds of change: an introduction to data-driven learning ». *RELC Journal* 33/2, 99-124.

HAFNER, C. et C. CANDLIN. 2007. « Corpus tools as an affordance to learning in professional legal education ». *Journal of English for Academic Purposes* 6/4, 303-318.

IFOP, Institut Français d'Opinion Publique (Département Opinion et Stratégies d'Entreprises, pour Systran). 2009. *Le multilinguisme dans l'entreprise*. <http://www.ifop.com/europe/docs/multilinguismeentreprise.pdf> et <http://www.ifop.com/europe/docs/multilinguismeentreprise_cp.pdf>.

pre-publication version

JARVIS, H. 2004. « Investigating the classroom applications of computers on EFL courses at higher education institutions in UK ». *Journal of English for Academic Purposes* 3, 111-137.

JOHNS, T. 1988. « Whence and whither classroom concordancing? » In T. BONGAERTS, P. de HAAN, S. LOBBE et H. WEKKER (dir.) *Computer Applications in Language Learning*. Dordrecht : Foris, 9-27.

JOHNS, T. 1991a. « Should you be persuaded: two examples of data-driven learning ». In T. Johns et P. King (dir.) *Classroom Concordancing. English Language Research Journal* 4, 1-16.

JOHNS, T. 1991b. « From printout to handout: grammar and vocabulary teaching in the context of data-driven learning ». In T. JOHNS et P. KING (dir.) *Classroom Concordancing. English Language Research Journal* 4, 27-45.

JOHNS, T. 1997. « Kibbitzing one-to-ones ». Web version. University of Reading : *BALEAP on Academic Writing*, 29 novembre. <http://www.eisu.bham.ac.uk/johnstf/pimnotes.htm>.

JOHNS, T. 2002. « Data-driven learning: the perpetual challenge ». In B. KETTEMANN et G. MARKO (dir.) *Teaching and Learning by Doing Corpus Analysis*. Amsterdam : Rodopi, 107-117.

JOHNS, T., L. HSINGCHIN et W. LIXUN. 2008. « Integrating corpus-based CALL programs and teaching English through children's literature ». *Computer Assisted Language Learning* 21/5, 483-506.

KENNING, M.-M. 2007. *ICT and Language Learning: From Print to the Mobile Phone*. New York : Palgrave Macmillan.

KIRSCHNER, P., J. SWELLER et R. CLARK. 2006. « Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching ». *Educational Psychologist* 41/2, 75-86. <http://www.cogtech.usc.edu/publications/kirschner_Sweller_Clark.pdf>.

KUHN, T. 1996 [1962]. *The Structure of Scientific Revolutions*. 3e édition. Chicago : University of Chicago Press.

LEE, C.-Y. et H.-C. LIOU. 2003. « A study of using web concordancing for English vocabulary learning in a Taiwanese high school context ». *English Teaching and Learning* 27/3, 35-56.

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, 56-75.

LEECH, G. 2007. « New resources, or just better old ones? The Holy Grail of representativeness ». In M. HUNDT, N. NESSELHAUF et C. BIEWER (dir.) *Corpus Linguistics and the Web*. Amsterdam : Rodopi, 133-149.

LITTLEMORE, J. et D. OAKEY. 2004. « Communication with a purpose : exploiting the internet to promote language learning ». In A. CHAMBERS, J. CONACHER et J. LITTLEMORE (dir.) *ICT and Language Learning: Integrating Pedagogy and Practice*. Birmingham, GB: University of Birmingham Press, 95-119.

LOEWEN, S., S. LI, F. FEI, A. THOMPSON, K. NAKATSUKASA, S. AHN et X. CHEN. 2009. « Second language learners' beliefs about grammar instruction and error-correction ». *Modern Language Journal* 93/1, 91-104.

MAIA, B. 1997. « Making corpora: a learning process ». In G. ASTON, L. GAVIOLI et F. ZANETTIN (dir.) *Proceedings of Corpus Use and Learning to Translate*. <http://www.sslmit.unibo.it/cultpaps/paps.htm>.

MAYER, R. 2004. « Should there be a three-strike rule against pure discovery learning? The case for guided methods of instruction ». *American Psychologist* 59/1, 14-19.

MAURANEN, A. 2004. « Speech corpora in the classroom ». In G. ASTON, S. BERNARDINI et D. STEWART (dir.) *Corpora and Language Learners*. Amsterdam : John Benjamins, 195-211.

pre-publication version

MCCARTHY, M. 2008. « Accessing and interpreting corpus information in the teacher education context ». *Language Teaching* 41/4, 563-574.

MCENERY, T., R. XIAO et Y. TONO. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. Londres : Routledge.

MILTON, J. 2006. « Resource-rich web-based feedback: helping learners become independent writers ». In K. HYLAND et F. HYLAND (dir.) *Feedback in Second Language Writing: Contexts and Issues*. Cambridge : Cambridge University Press, 123-137.

O'SULLIVAN, I. 2007. « Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy ». *ReCALL* 19/3, 269-286.

O'SULLIVAN, I. et A. CHAMBERS. 2006. « Learners' writing skills in French: corpus consultation and learner evaluation ». *Journal of Second Language Writing* 15/1, 49-68.

RÖMER, U. 2009. « Corpus research and practice: what help do teachers need and what can we offer? » In K. AIJMER (dir.) *Corpora and Language Teaching*. Amsterdam : John Benjamins, 83-98.

SCOTT, M. et C. TRIBBLE. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam : John Benjamins.

SINCLAIR, J. 2004. « New evidence, new priorities, new attitudes ». In J. Sinclair (dir.) *How to Use Corpora in Language Teaching*. Amsterdam : John Benjamins, 271-299.

ST. JOHN, E. 2001. « A case for using a parallel corpus and concordancer for beginners of a foreign language ». *Language Learning & Technology* 5/3, 185-203.

SUN, Y.-C. 2003. « Learning process, strategies and web-based concordancers: a case-study ». *British Journal of Educational Technology* 34/5, 601-613.

SUN, Y.-C. 2007. « Learner perceptions of a concordancing tool for academic writing ». *Computer Assisted Language Learning* 20/4, 323-343.

THOMAS, J. 2008. « Impatience is a virtue: students and teachers interact with corpus data – now ». In A. Frankenberg-Garcia (dir.) *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbonne : Instituto Superior de Línguas e Administração, 463-469.

TIAN, S. 2005. « The impact of learning tasks and learner proficiency on the effectiveness of data-driven learning ». *Journal of Pan-Pacific Association of Applied Linguistics* 9/2, 263-275.

TODD, R. 2001. « Induction from self-selected concordances and self-correction ». *System* 29/1, 91-102.

TURNBULL, J. et J. BURSTON. 1998. « Towards independent concordance work for students: lessons from a case study ». *ON-CALL* 12/2, 10-21.

TYNE, H. 2009. « Corpus oraux par et pour l'apprenant ». In A. Boulton (dir.) *Des documents authentiques oraux aux corpus : questions d'apprentissage en didactique des langues. 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.

YOON, H. et A. HIRVELA. 2004. « ESL student attitudes toward corpus use in L2 ». *Journal of Second Language Writing* 13/4, 257-283.

pre-publication version