



ASp
la revue du GERAS

66 | 2014
**Intersections - l'anglais de spécialité, creuset
multidomaine**

Alex Boulton, Henry Tyne, *Des documents
authentiques aux corpus : démarches pour
l'apprentissage des langues*

Paris : Didier, 2014

Geneviève Bordet



Édition électronique

URL : <http://journals.openedition.org/asp/4547>

DOI : [10.4000/asp.4547](https://doi.org/10.4000/asp.4547)

ISSN : 2108-6354

Éditeur

Groupe d'étude et de recherche en anglais de spécialité

Édition imprimée

Date de publication : 1 novembre 2014

Pagination : 159-163

ISSN : 1246-8185

Référence électronique

Geneviève Bordet, « Alex Boulton, Henry Tyne, *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues* », *ASp* [En ligne], 66 | 2014, mis en ligne le 27 octobre 2014, consulté le 03 novembre 2020. URL : <http://journals.openedition.org/asp/4547> ; DOI : <https://doi.org/10.4000/asp.4547>

Ce document a été généré automatiquement le 3 novembre 2020.

Tous droits réservés

Alex Boulton, Henry Tyne, *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*

Paris : Didier, 2014

Geneviève Bordet

RÉFÉRENCE

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

- 1 Alex Boulton et Henry Tyne, tous deux spécialistes reconnus de l'utilisation de corpus pour la didactique des langues étrangères, s'adressent ici aussi bien aux chercheurs qu'aux enseignants et formateurs en langues pour exposer les principes de cette démarche et en prouver l'intérêt pour l'apprentissage des langues. Ils s'appuient sur leur connaissance exhaustive du monde de la recherche en linguistique de corpus pour promouvoir une approche pédagogique innovante. Leur ouvrage constitue en effet un plaidoyer pour une évolution des approches de l'apprentissage des langues, s'appuyant sur l'exposition à des documents dits « authentiques ». C'est aussi un ouvrage ambitieux alliant prise en compte de l'état de la recherche et application didactique.
- 2 Dès l'introduction, le lectorat visé est clairement identifié : il s'agit avant tout des « acteurs de terrain », enseignants et apprenants de langues étrangères. L'objectif est de les sensibiliser à la démarche d'apprentissage sur corpus (ASC) pour leur permettre d'avancer dans leurs propres pratiques.
- 3 Le premier des sept chapitres est consacré à l'exposition à la langue cible comme condition des apprentissages. Le contact avec la langue cible sous la forme de documents authentiques nécessite en effet un accompagnement des apprenants et une familiarisation avec les outils technologiques qui permettent cette mise en contact directe, puisque l'apprentissage sur corpus s'appuie sur les technologies de l'information et de la communication (TIC). Un exposé historique met en évidence le passage, depuis les années 1970, de la transmission des savoirs à une dynamique d'apprentissage autonome, dans une perspective socioconstructiviste dont témoignent par exemple les exigences définies par le Cadre européen commun de références pour les langues (CECRL), formulées en termes de réalisation de tâches en langue cible. La démarche proposée par l'ASC s'appuie sur cette évolution pédagogique et dépasse donc la simple application de la linguistique de corpus. La construction de l'autonomie, son accompagnement, son intégration parfois dans une démarche plus conventionnelle a fait l'objet de nombreuses recherches (Porcher 1992) qui montrent la nécessité de prendre en compte les spécificités des apprenants, en termes d'âge et de motivation notamment. L'autonomie apparaît dans le monde de l'apprentissage des langues comme une condition indispensable à l'adaptation de l'apprenant à la variabilité des situations de communication langagière. Une autre caractéristique de l'évolution de l'apprentissage des langues est l'exposition directe à des documents authentiques, extraits de situations de communication réelles et non remaniés à des fins d'exemplification didactique. Les auteurs mettent en évidence le lien entre corpus et authenticité des données. L'accès direct à des textes écrits et oraux permet de proposer



à l'apprenant un « bouquet » choisi de données brutes sur lesquelles il fondera ses observations et la vérification d'hypothèses élaborées dans un contexte de communication. Cet accès aux corpus a été rendu possible par une longue chaîne d'inventions technologiques de l'imprimerie à l'ordinateur en passant par le magnétophone. De même, le XX^e siècle a vu le passage de l'enseignement assisté par ordinateur (EAO) à l'apprentissage des langues assisté par ordinateur (ALAO), marquant le recentrage sur l'apprenant et sur son autonomisation. Néanmoins, le progrès technologique est d'une telle rapidité qu'il est parfois difficile à combiner avec de véritables avancées pédagogiques. Par ailleurs, alors que tous les apprenants sont quotidiennement exposés aux langues étrangères grâce aux divers modes de communication qu'ils utilisent dans leur vie quotidienne, leur niveau de « littératie » ne leur permet pas toujours d'en tirer profit pour des apprentissages. Si les TIC fournissent un contexte favorable à l'autonomisation de l'apprentissage, elles ne le garantissent donc pas. L'ASC et l'utilisation des corpus nécessitent un accompagnement et la construction de méthodologies *ad hoc*, dont le point commun est qu'elles se fondent sur l'autonomie et sur la motivation de l'apprenant.

- 4 Le chapitre 2 s'attache à décrire les conditions et les conséquences de l'émergence des corpus pour le monde de la linguistique et la manière d'aborder les langues. La linguistique de corpus est d'abord définie par opposition à ce qui serait la « linguistique de fauteuil » (Fillmore 1992). Cette approche de la langue permet de dépasser l'intuition individuelle du linguiste en se fondant sur l'observation de données réelles. Le corpus est présenté comme un ensemble de textes collectés à des fins précises et interrogeables de manière automatisée. Les outils disponibles actuellement permettent également l'accès à des données orales, même si leur collecte est nettement en retard sur celle des données écrites. Qu'il s'agisse de données orales enregistrées ou de données orales transcrites, leur prise en compte ouvre ainsi de nouvelles perspectives pour la formation des apprenants aux écarts entre pratiques de la langue orale et de la langue écrite. Cependant, les modalités d'accès à ces données restent encore assez limitées et calquées sur l'étude de l'écrit. À l'oral comme à l'écrit, l'application des TIC aux corpus permet de fonder l'approche de la langue sur l'observation et l'analyse de données statistiques, telles que la fréquence des termes, leur distribution dans les textes et leurs modalités de combinaison, collocations et cooccurrences. Les applications de ces nouvelles approches du comportement de la langue en contexte pour l'élaboration de ressources destinées à l'apprentissage des langues sont de deux ordres. Indirectes, elles permettent l'élaboration d'outils de formation et d'apprentissage tels que manuels, dictionnaires et syllabus. Sont notamment cités le dictionnaire *COBUILD* (1987) et le *Longman Grammar of Spoken and Written English* (Biber 1999). Le corpus sur lequel se fondent ces ressources reste néanmoins le plus souvent invisible. Ce n'est pas le cas pour l'approche dite du « *data-driven learning* », où l'apprenant observe directement cette fois les données collectées grâce à des outils d'interrogation statistiques. Ces outils, initialement conçus par des chercheurs pour des chercheurs, donnent aux apprenants la possibilité de confronter leurs hypothèses et leurs interrogations avec des données textuelles provenant de situations de communication réelles. L'apprenant devient alors un « détective linguistique » (Johns 1997), dans une approche typiquement constructiviste.
- 5 Le chapitre suivant est consacré à une découverte pas à pas du concept de corpus et des perspectives ainsi que des difficultés qu'ouvre son utilisation à des fins didactiques. Une première étape met en évidence des utilisations de techniques d'interrogation

automatique de textes dans des applications quotidiennes de la bureautique et d'Internet. Puis sont présentées les modalités d'interrogation de corpus constitués tels que le *British National Corpus* (BNC) ou le *Corpus of Contemporary American English* (CoCA). Une dernière partie porte sur l'analyse d'un mot en contexte grâce à l'utilisation de concordanciers.

- 6 Dans le chapitre 4 sont abordées les questions méthodologiques liées à l'approche proposée par l'apprentissage sur corpus. La portée des apprentissages est très large puisqu'elle peut s'appliquer à l'acquisition de toutes les compétences grâce à une approche inductive qui procède par découverte et par vérification. Néanmoins, la mise au point de procédures d'apprentissage adaptées aux corpus et leur utilisation par les apprenants exigent un temps considérable. C'est la raison pour laquelle l'ASC peut n'être qu'une simple composante parmi d'autres du processus d'apprentissage, en fonction des besoins. Les auteurs de l'ouvrage plaident pour une intégration progressive, répondant à des besoins réels. De la même manière, l'approche préconisée en ce qui concerne la taille souhaitable des corpus est avant tout pragmatique. Qu'il s'agisse de corpus déjà constitués et de grande taille ou de petits corpus collectés pour une étude spécifique, ils doivent être considérés comme une ressource parmi d'autres et leur représentativité ne peut jamais être considérée comme absolue. Le corpus n'est pas le reflet parfait de la langue mais le moyen de vérifier une hypothèse et, surtout, d'interpréter l'écart entre le résultat escompté et le résultat obtenu. La question de la taille des corpus cède la place à celle de leur composition et de leur adaptation à un questionnement. La nature du corpus doit notamment permettre la prise en compte des phénomènes de variation de genre ou de registre par exemple et l'éventuelle identification et analyse des erreurs.
- 7 Le chapitre 5 procède au bilan et à l'analyse critique de l'ASC fondés sur les études scientifiques disponibles dans ce domaine. Au total, 129 études portant sur l'efficacité de l'ASC pour l'apprentissage des langues sont recensées et analysées. L'acceptation de l'ASC adoptée pour le choix de ces études est très large puisqu'elle inclut « toutes données tirées de l'utilisation explicite d'un corpus à des fins d'apprentissage ou d'aide à l'expression écrite » (p. 145). Un panorama exhaustif montre que la plupart ont été rédigées en anglais et ont été publiées depuis 1990 et plus particulièrement dans les années 2000. Les paramètres pris en compte incluent la nature des corpus et des outils, la variété des langues, les objectifs de formation, les pays concernés, le profil des apprenants et le niveau de formation préalable requis. Les résultats font apparaître que, malgré les difficultés liées au temps et à la formation préalable requis, l'utilisation des corpus pour l'apprentissage fait de ses utilisateurs de « meilleurs apprenants », plus autonomes et plus conscients des phénomènes linguistiques. Ces bénéfices demeurent néanmoins très difficiles à quantifier.
- 8 Le chapitre 6 présente un échantillon de requêtes possibles sur corpus, en prenant pour exemple les corpus disponibles sur le site de l'université américaine de Brigham Young (BYU) et notamment le corpus CoCA. Les requêtes de base sont commentées et illustrées. Est ainsi abordée l'étude de la distribution des occurrences d'un mot, des formes synonymes, des mots-clés en contexte (KWIC) et des collocations. En conclusion de ce chapitre, les auteurs attirent l'attention sur les limites de ces outils d'interrogation, auxquels un dictionnaire, dans certains cas, peut fournir un substitut moins coûteux en temps et en manipulations.

- 9 Le dernier chapitre évoque plusieurs pistes pour constituer ce que les auteurs appellent « corpus de fortune », « ad hoc », « éphémères » ou « jetables ». La première possibilité est l'utilisation, présentée comme aléatoire, du « web comme corpus » grâce à des outils tels que Webcorp¹. La deuxième, pour la collecte de textes sur le Web, est celle du recours à une ressource telle que Web Boot Cat² à partir de mots-clés associés à un domaine. Est ensuite abordée la question du travail sur des données orales dans un contexte d'analyse de corpus et de ses spécificités, notamment liées au passage par la transcription. Enfin, sont présentés un ensemble d'outils, comme le concordancier AntConc³ ou les ressources proposées sur le site Lex Tutor⁴, qui permettent de se composer des corpus sur mesure et de les interroger en fonction des objectifs didactiques.
- 10 Dans leur conclusion, les auteurs reviennent sur les possibilités ouvertes par l'utilisation de corpus rassemblant des données textuelles authentiques et leur interrogation grâce aux outils technologiques adaptés. Ils insistent notamment sur les vertus heuristiques de l'ASC qui permet de faire émerger des réalités discursives parfois inattendues. Enfin, ils plaident pour une « application ouverte » de cette méthode, loin de tout dogmatisme.
- 11 L'ouvrage comprend également une très abondante bibliographie et deux annexes, dont une liste de corpus et de ressources en ligne ainsi que la liste bibliographique des études sur l'ASC recensées.
- 12 Le grand mérite *Des documents authentiques aux corpus* est son accessibilité et son absence de dogmatisme, parfaitement conformes aux objectifs annoncés de sensibilisation à l'ASC des acteurs de terrain de la formation à la maîtrise des langues. Les exemples sont concrets, les ressources sont décrites de manière facilement compréhensible et bien illustrées. Néanmoins, cette volonté de simplification a parfois pour corollaire une certaine imprécision dans la formulation ou dans la définition des notions. On peut ainsi regretter qu'il faille attendre la page 43 et le deuxième chapitre pour que soit définie la notion de corpus, pourtant centrale. Par ailleurs, on pourrait discuter l'accent mis sur les corpus en ligne de grande taille, au détriment des corpus « de fortune » considérés comme des ressources accessoires et pour tout dire par défaut. Or, si les grands corpus généraux sont extrêmement utiles pour étudier les usages standards de la langue, la constitution de corpus *ad hoc* est indispensable pour l'apprentissage des langues spécialisées et pour leur traduction. Ce dernier aspect, pourtant mentionné à plusieurs reprises, est particulièrement peu traité, non seulement dans la description de l'usage des corpus, mais aussi dans les références bibliographiques.
- 13 Malgré ces quelques limites, cet ouvrage sera utile aux enseignants de langue étrangère désireux de s'initier à de nouvelles méthodes appuyées sur les technologies de l'information. Il ouvre également la voie à de futures publications proposant des séquences pédagogiques qui répondent, dans la perspective de l'ASC, à tout l'éventail des besoins de l'apprentissage linguistique, de la découverte de la langue écrite à celle de la langue orale, en passant par les langues spécialisées et la traduction.

BIBLIOGRAPHIE

Biber, Douglas et alii. 1999. *Longman Grammar of Spoken and Written English*. Londres - New York : Longman.

COBUILD English Language Dictionary. 1987. Londres : Collins.

Fillmore, Charles. 1992. « Corpus linguistics or computer-aided armchair linguistics ». In Svartvik, J. (dir.), *Directions in Corpus Linguistics*. Berlin : Mouton de Gruyter, 13-38.

Johns, Tim. 1997. « Contexts: The background, development and trialling of a concordance-based CALL program ». In Fligelstone, S., T. Mac Enery & G. Knowles (dir.), *Teaching and Language Corpora*. Londres : Longman, 100-115.

Porcher, Louis. 1992. « Omniprésence et diversité des auto-apprentissages ». *Le Français dans le Monde*, « Les auto-apprentissages », numéro spécial, fév-mars, 6-14.

NOTES

1. <http://www.webcorp.org.uk/live/>
 2. <http://bootcat.sslmit.unibo.it/>
 3. http://www.laurenceanthony.net/antconc_index.html
 4. <http://www.lextutor.ca/>
-

AUTEURS

GENEVIÈVE BORDET

Université Paris Diderot, CLILLAC-ARP. gbordet@eila.univ-paris-diderot.fr