



HAL
open science

Des Documents Authentiques aux Corpus : introduction

Alex Boulton, Henry Tyne

► **To cite this version:**

Alex Boulton, Henry Tyne. Des Documents Authentiques aux Corpus : introduction. Alex Boulton ; Henry Tyne. Des Documents Authentiques aux Corpus : Démarches pour l'Apprentissage des Langues, Didier, pp.309, 2014. hal-00937418

HAL Id: hal-00937418

<https://hal.science/hal-00937418>

Submitted on 12 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.

Alex Boulton & Henry Tyne. Introduction à : *Des Documents Authentiques aux Corpus: Démarches pour l'Apprentissage des Langues*. Paris: Didier.

Introduction et organisation de l'ouvrage

Apprendre une autre langue que sa langue maternelle a été un enjeu pour l'homme à travers les âges et cela n'a jamais été une mince affaire. Depuis longtemps on essaie d'inventer des moyens pour faire en sorte que l'apprentissage soit le plus facile, le plus efficace et le plus rapide possible, car le nombre d'heures de contact avec la langue cible à l'école est d'ordinaire si réduite par rapport à un enfant qui grandit avec sa langue. On perfectionne aussi les outils qui pourraient aider à faciliter l'apprentissage.

Au fur et à mesure que nos connaissances sur l'appropriation des langues progressent, on peaufine des approches en didactique. Et on en vient également à repenser les objectifs d'apprentissage : concrètement, que veut-on apprendre, comment et pourquoi ? A l'heure actuelle, on est à peu près certain des choses suivantes :

- l'apprenant n'est pas une simple copie du locuteur natif ;
- l'apprenant n'est pas non plus une éponge ; il doit construire sa langue ;
- les apprenants ne sont pas tous identiques ; il existe des objectifs, des styles, des motivations, des attitudes, des capacités qui diffèrent ;
- le contact avec la langue cible, sous une forme ou sous une autre, est primordial.

Par ailleurs, la conjecture actuelle fait qu'il est possible qu'un apprenant entre en contact relativement facilement, où qu'il soit, avec des données en langue cible, notamment grâce aux technologies de l'information et de la communication. La technologie a rendu possible un ensemble de pratiques de gestion du contact avec les données en langue cible, notamment sur Internet. Il existe aussi des ensembles plus ou moins ordonnés de données, qu'on appelle des « corpus », qui permettent de spécifier les recherches que l'on souhaite faire et, surtout, d'axer celles-ci sur des questions proprement linguistiques.

Ce livre se donne comme objectif de présenter une approche en didactique des langues que l'on peut nommer « apprentissage sur corpus » (en anglais : « *data-driven learning* »¹), approche dans laquelle l'apprenant est amené à mettre à profit ses différentes observations de la langue à partir de données qui se présentent sous forme de corpus ou collections de textes/transcriptions. L'apprentissage sur corpus (désormais ASC) s'est développé au cours des vingt dernières années, grâce notamment aux possibilités offertes par les technologies de l'information et de la communication (TIC). Mais grâce aussi aux avancées en matière de description des langues (l'arrivée relativement récente de méthodologies sur corpus) et en didactique des langues (approche communicative centrée sur l'apprenant, utilisation de documents authentiques, etc.). Le « *data-driven learning* » en est venu à s'imposer en tant que véritable approche nouvelle en didactique des langues (même si elle arrive à peine en France) et en tant que paradigme de recherche.

Si la méthodologie d'observation sur corpus développée en linguistique a surtout eu tendance à privilégier les données linguistiques (les formes, les occurrences, la distribution, etc.), il paraît clair que les corpus peuvent être pris en compte en tant que documents

¹ Littéralement l'apprentissage « conduit ou porté par les données » (voir Tyne 2013).

authentiques, au sens le plus complet du terme, permettant de nombreuses utilisations et exploitations pédagogiques, et devenant une riche source d'activités. La didactique des langues a ainsi pu se doter de méthodologies nouvelles : l'enseignant a recours à des données authentiques, à des exemples d'emplois récurrents, à des méthodologies d'exploitations avec ses apprenants ; l'apprenant, quant à lui, devient chercheur (il observe, exploite ou fabrique des corpus), fait des découvertes, s'informe sur la langue à partir de données authentiques, etc.

On peut résumer les différentes approches de l'apprentissage des langues à partir de corpus de la manière suivante :

- on considère les apprenants comme la *source* des données : on travaille sur des données produites par des apprenants afin de mieux comprendre les productions d'apprenants tout au long de leur parcours d'acquisition, offrant des possibilités de description en fonction des différents stades observés à partir de données réunies sous forme de corpus ;
- on considère les apprenants comme étant les *bénéficiaires* des données : les enseignants ou concepteurs de manuels, syllabus, dictionnaires, etc. travaillent sur des corpus de données produites généralement par des locuteurs natifs de la langue cible afin de créer des supports pédagogiques, l'accent est mis sur le recours aux données authentiques pour éviter les pièges de l'intuition – c'est une utilisation dite « indirecte » ;
- on considère les apprenants comme étant les *utilisateurs* des données : les apprenants travaillent sur les corpus mêmes, observent les données, induisent les règles – c'est une utilisation dite « directe ».

Ce sont principalement ces deux dernières approches que nous retiendrons ici. Dans ce livre, nous présenterons le paradigme d'apprentissage sur corpus en nous appuyant sur des exemples mais aussi sur des recherches récentes et en élargissant la réflexion vers le cadre plus général de la découverte en didactique des langues et l'apport des technologies à celle-ci.

Ce livre s'appuie sur des données et des expériences de terrain et tient compte des recherches les plus récentes dans le domaine. Il a en outre une visée pratique, étant destiné, entre autres, aux acteurs de terrain qui ne sont pas dans le monde de la recherche universitaire, qui n'ont pas nécessairement des compétences informatiques très développées et qui ne disposent pas d'installations technologiques très sophistiquées sur leur lieu d'exercice. Les exemples que nous donnons dans ce livre concernent différentes langues et ne seront donc pas focalisés sur une seule. Nous espérons ainsi que le chercheur, l'enseignant, l'apprenant, ou tout autre utilisateur de langues étrangères, quelles qu'elles soient, trouvera parmi les méthodologies et les réflexions, parmi les outils et les techniques, parmi les comptes-rendus d'expériences que nous présentons ici, de quoi éveiller sa curiosité, de quoi avancer dans ses propres pratiques.

Les deux premiers chapitres de ce livre servent à introduire le contexte dans lequel s'est développé l'ASC avant d'introduire le travail sur corpus dans le chapitre 3. Ensuite viennent deux chapitres (4 et 5) dédiés aux différentes questions lors des applications de l'ASC,

d'abord sous un angle général et méthodologique, ensuite en résumant les résultats des études existantes. Les deux derniers chapitres (6 et 7) sont avant tout destinés à présenter des outils et des possibilités de collecte ou d'analyse, et sont conçus pour une lecture plus sélective, non linéaire en fonction des intérêts des lecteurs. Une liste des sites et des outils est donnée en annexe.