

# INTRODUCTION

Virginie PRIVAS-BRÉAUTÉ

Pour le chercheur en didactique des langues, le travail sur échantillon constitue une voie envisageable pour nourrir la réflexion autour de questions de recherche données. Pour ce faire, il est recommandé que la taille des échantillons d'appui soit « représentative<sup>1</sup> » afin de satisfaire aux exigences de robustesse, validité et normes académiques habituelles de la sphère scientifique : enregistrer quelques heures d'interaction dans une salle de cours permet en effet de recueillir quelques éléments exploitables mais ne tient pas lieu d'échantillon représentatif de la langue parlée en situation réelle, par exemple.

Construire un corpus s'inscrit dans une démarche scientifique globale (Dalbera, 2002) et implique une analyse spécifique en arrière-plan et le type de données que le chercheur sélectionne traduit des préoccupations sous-jacentes précises, d'où la multiplicité des formes que peuvent adopter les corpus : corpus de textes, corpus d'unités de langue, ceux qui relèvent de l'écrit et ceux qui relèvent de l'oral, corpus conçus comme des échantillons de faits linguistiques, ceux qui se veulent exhaustifs dans un champ donné entre autres. Dans une optique « responsable » de recherche (Narcy-Combes, 2005), il est donc essentiel pour le chercheur de penser son corpus et de le construire en fonction de ce qu'il souhaite analyser, étudier, démontrer.

Cet ouvrage est l'occasion de diffuser la variété de possibilités de recueil de données, de constitution de corpus, d'exploitation et d'analyse des données qu'ils contiennent. Les auteurs des contributions qui constituent ce recueil sont des spécialistes reconnus et incontournables de la didactique des langues en France. Ils présentent leurs corpus et les outils qu'ils utilisent pour analyser leurs données et répondre aux questions de recherche qu'ils poursuivent dans le cadre de leurs activités scientifiques. Si l'anglais est la langue cible citée en exemple dans la plupart des contributions, il n'en est pas moins pertinent de transposer les outils méthodologiques proposés ici aux autres langues vivantes étrangères.

La première partie de cet ouvrage s'adresse tout particulièrement aux néochercheurs, étudiants en master ou en thèse en sciences du langage ou langues et cultures étrangères et propose de clarifier les enjeux d'une recherche sur corpus. Séverine Behra, Maud

---

1. Par exemple, en statistiques, les échantillons strictement supérieurs à 30 individus sont considérés comme de grands échantillons.

Ciekanski, Guillaume Nassau et Dominique Macaire expliquent ainsi comment et pourquoi le jeune chercheur constitue son corpus à des fins d'analyse. Les auteurs s'appuient sur le travail de recherche d'un jeune enseignant-chercheur afin d'illustrer et d'accompagner pas à pas tout étudiant entamant des recherches en sciences du langage.

Une deuxième partie présente les travaux de recherche effectués sur des corpus textuels et oraux. Ils sont analysés au moyen d'outils visant à annoter les données recueillies. Ainsi, Justine Paris et Pauline Beaupoil-Hourdel se penchent sur la constitution de corpus oraux dans le cadre de l'acquisition d'une langue maternelle et d'une langue seconde, l'anglais est ici pris en exemple, et montrent comment exploiter des outils de transcription tels que CLAN (utilisé pour la transcription langagière), ELAN (utilisé pour des annotations gestuelles), PRATT (utilisé pour des annotations prosodiques) et EXCEL pour faire émerger les régularités et irrégularités de ces acquisitions. Elles reviennent ainsi sur les étapes décisives lors de la constitution, du traitement et de l'analyse d'un corpus oral et soulignent la valeur des outils permettant des analyses multimodales. Dans la contribution suivante, Pascale Manoïlov et Agnès Leroux-Béal présentent les modalités de valorisation et d'exploitation d'un corpus d'interactions orales filmées en classe d'anglais entre des apprenants de cycle 4 (classe de 3<sup>e</sup>), complétées par des interactions avec des enseignants dans une perspective comparative. Les deux enseignantes-chercheuses s'interrogent alors sur les enjeux du traitement des données et de leur exploitation ainsi que sur les analyses qui peuvent en découler. Elles montrent également comment utiliser CLAN pour transcrire et traiter quantitativement et qualitativement les données.

La troisième partie porte sur les analyses qualitatives des données recueillies à l'oral ou à l'écrit auprès d'élèves français et la nécessité de constituer des corpus pour comprendre les enjeux de l'apprentissage des langues dès les petites classes de maternelle. Ainsi, à la 4<sup>e</sup> contribution, Marie Potapushkina-Delfosse et Anne-Marie Voise analysent les occurrences à l'oral d'élèves de maternelle lors de la sensibilisation à la langue anglaise par le dessin, le jeu et le conte. Les corpus de leurs recherches sont essentiellement composés d'enregistrements audio et vidéo des productions des élèves, d'interactions élèves-enseignants et de transcription d'entretiens menés individuellement, ou en petits groupes. Elles exposent quelques dispositifs d'élaboration de différents types de corpus oraux et multimodaux, leurs fondements, leurs traitements, leurs usages dans le cadre de recherches menées au tout début de l'apprentissage des langues en contexte scolaire. Elles illustrent leur propos à l'aide d'exemples d'extraits de corpus analysés à la lumière des questions qui guident leurs investigations. Stéphane Soulain et Caroline Raymond rendent, quant à eux, compte des défis méthodologiques associés à la production et à l'analyse de données qualitatives auprès d'enseignantes en langues vivantes étrangères. Leur étude, de type recherche collaborative, porte sur la proposition à un groupe de six enseignantes des niveaux primaire et secondaire de participer à une formation de l'*Apprendre par corps*, les engageant à vivre un programme d'activités corporelles liées au rythme de la langue, au phénomène d'interaction et à la production gestuelle. Après avoir défini l'*Apprendre par corps* en langues vivantes, les visées et les contenus de cette démarche d'apprentissage, ils développent les méthodes retenues pour documenter le processus de transposition didactique interne (Chevallard, 1991) convoqué par les enseignantes à l'issue de leur formation. Grâce à la méthode de l'entretien d'explicitation (Vermersch, 2017), ils ont

eu à cœur d'accompagner les enseignantes dans la verbalisation de leur expérience des activités corporelles et de leur transposition en classe, au service de la pratique orale de la langue anglaise. Puis, Margot Kuligowska Esnault se penche sur l'utilisation de poèmes pour l'apprentissage de l'anglais dans le second degré et examine le corpus qu'elle a constitué à partir de son vécu en classe de langue. Cette contribution s'appuie sur une partie d'une étude doctorale axée sur la poésie dans l'enseignement-apprentissage des langues. Kuligowska Esnault expose et interroge les choix qui ont été retenus pour analyser un corpus de 32 textes poétiques en L2 anglais, produits par des élèves du secondaire en France. L'analyse de ces productions écrites était à la fois de type qualitatif (analyse de contenu, Bardin, 2013 [1977]), avec un traitement manuel, et quantitatif (utilisant le logiciel d'analyse lexicale en ligne VocabProfile-Classic et le logiciel SPSS<sup>2</sup> d'IBM).

L'ouvrage se termine par une quatrième partie portant sur une tendance grandissante au sein de la communauté d'enseignants-chercheurs en France : l'apprentissage des langues sur corpus. Si le corpus est déjà un dispositif d'apprentissage à l'étranger depuis de nombreuses années, ce n'est que depuis très récemment qu'il est utilisé en France. Ainsi, Clive E. Hamilton propose une contribution portant sur son utilisation de corpus textuels auprès d'étudiants en anglais LANSAD et en anglais de spécialité dans l'enseignement supérieur. Sa contribution souligne l'importance des corpus non seulement en tant qu'objets d'analyse mais aussi en tant qu'objets d'apprentissage.

## BIBLIOGRAPHIE

- BARDIN Laurence, 2013 [1977], *L'analyse de contenu*, 2<sup>e</sup> édition, Paris, Presses universitaires de France.
- CHEVALLARD Yves, 1985, *La transposition didactique. Du savoir savant au savoir enseigné*, Paris, La Pensée sauvage.
- DALBERA Jean-Philippe, 2002, « Le corpus entre données, analyse et théorie », *Corpus*, n° 1, [<https://journals.openedition.org/corpus/10>], consulté le 1<sup>er</sup> mars 2022.
- NARCY-COMBES Jean-Paul, 2005, *Didactique des langues et TIC. Vers une recherche-action responsable*, Paris, Éditions Ophrys.
- VERMERSCH Pierre, 2017, *L'entretien d'explicitation*, 9<sup>e</sup> édition, Paris, ESF Éditeur.

---

2. *Statistical Package for the Social Sciences*.

