

Le corpus de l'étude

Le « corpus » désigne l'ensemble des documents (œuvres, peintures, statistiques, romans, nouvelles, essais, poèmes, etc.) sur lesquels portera l'étude. C'est le matériau de base et le support principal de la recherche. C'est pourquoi il faut prêter une attention particulière à plusieurs aspects du corpus :

- **Le choix du corpus** est très important parce qu'il permet de déterminer la nature et le cadre de l'étude. Ce choix est tributaire des objectifs fixés pour la recherche : monographie, étude comparée, étude historique, etc. Car pour chaque type d'étude, il faut constituer un corpus spécifique et, si possible, original. Celui-ci peut englober l'ensemble des œuvres d'un auteur pour une thèse de Doctorat (étude exhaustive) ou une partie seulement des œuvres en Maîtrise ou en Master (étude thématique partielle).

Le choix du corpus doit respecter trois critères : *la pertinence* (du corpus par rapport au sujet choisi), *la cohérence* (entre les différents textes constitutifs du corpus) et *la consistance* (taille du corpus et faisabilité de l'étude).

- **La constitution du corpus** : Jusqu'à une époque récente, la matière de la recherche (textes, romans, nouvelles, etc.) existait seulement sous format papier (publications imprimées). Ainsi pour maîtriser son matériau, le chercheur devait lire et relire sans cesse les documents qui constituent le support de son étude, opération fastidieuse et coûteuse en temps et en effort. Mais l'on assiste depuis quelques années à une véritable révolution de la recherche, conséquence directe de la généralisation de l'Internet et de la disponibilité d'immenses ressources documentaires en ligne. Non pas que le chercheur n'a plus à lire les œuvres de son corpus, mais il a désormais la possibilité de faire des requêtes précises sur les fonds documentaires des principales bibliothèques nationales et internationales.

De la sorte, la constitution d'une bibliographie ou d'un corpus d'étude devient une opération aisée et éminemment stimulante. Ainsi par exemple, un doctorant travaillant sur William Blake peut réunir et graver sur un même cédérom la totalité des œuvres de l'auteur à partir de

l'Internet et sous un format réutilisable (texte brut ou avec mise en page, en faisant du copier / coller ou du téléchargement). Mais pourquoi faire ?

• **L'utilisation du corpus** : La disponibilité du corpus sous format électronique (sur disquette ou cédérom) ouvre des perspectives extraordinaires d'exploitation, inconnues jusqu'ici des chercheurs non habitués à l'outil informatique. Ainsi, notre jeune chercheur sur William Blake peut envisager, une fois en possession des œuvres complètes de l'auteur sur un support électronique, de faire les opérations automatiques suivantes :

- recherche d'un mot clé ou d'un nom propre dans toutes les œuvres, avec indication de sa fréquence et sans passer par une relecture exhaustive de la totalité du corpus ;
- recherche rapide d'une phrase ou d'une expression particulière dans tous les textes susceptibles d'éclairer l'étude, avec indication du contexte ;
- possibilité d'effectuer une analyse linguistique fine et pointue en s'aidant d'outils spécialement dédiés à l'étude des corpus électroniques (tel que *Tropes Zoom* pour le français ou *Wordsmith Tools* pour l'anglais) ;
- possibilité de comparer automatiquement différents textes d'un même auteur ou différentes versions d'un même texte.

Bref, l'outil informatique offre une multitude de possibilités qui vont du simple dépouillement statistique aux subtiles analyses thématiques. C'est pourquoi le chercheur moderne se doit d'acquérir une maîtrise minimale des TICE (Technologie de l'information et de la communication pour l'enseignement).