

Du corpus traditionnel au corpus numérique

From the traditional corpus to the digital corpus

Radhia CHERAK
Centre universitaire Si El-Haouès -Barika / Algérie
radhiacherak@cu-barika.dz

Reçu le : 09/02/2022, **Accepté le:** 10/04/2022, **Publié le:** 30/10/2022

Résumé:

Le corpus est considéré comme le noyau d'une recherche scientifique car pour vérifier son ou ses hypothèse (s), le chercheur doit faire appel à un corpus bien déterminé. Il existe plusieurs types de corpus comme les enregistrements, les vidéos, les copies d'étudiants, les œuvres littéraires, les affiches publicitaires, etc. Ce corpus, une fois collecté, l'étudiant peut le conserver facilement et l'utiliser selon ses besoins. Mais actuellement, la nature des corpus a bien changé et l'étudiant possède une variété de choix, par exemple il suffit d'avoir un téléphone portable ou un PC connecté pour arriver à choisir son corpus disponible sous ses différentes formes : photos, textes, vidéos, documents sonores, etc. Il s'agit bel et bien d'un corpus numérique. Donc, dans cette contribution, nous allons voir comment se fait le choix d'un corpus et quelles différences y a-t-il entre un corpus traditionnel et un corpus numérique.

Mots clés: Corpus traditionnel- corpus numérique- recherche scientifique- différences- étudiant

Abstract:

The corpus is considered as the core of a scientific research because in order to verify his hypothesis (es), the researcher must use a well-defined corpus. There are several types of corpus such as recordings, videos, students' copies, literary works, advertising posters, etc. Once collected, this corpus can be easily preserved and used by the student according to his needs. But nowadays, the nature of the corpus has changed and the student has a variety of choices, for example, it is enough to have a cell phone or a PC connected to choose his corpus available in its different forms: photos, texts, videos, sound documents, etc. It is indeed a digital corpus. Therefore, in this contribution, we will see how the choice of a corpus is made and what differences there are between a traditional corpus and a digital corpus.

Key words: Traditional corpus- digital corpus- scientific research- differences- student.

Introduction

Le corpus est défini par Charaudeau et Maingueneau (1988 :2002) comme étant «*un recueil large quelquefois exhaustif de documents ou de données* ». Il est considéré comme le noyau d'une recherche scientifique car pour vérifier son ou ses hypothèse (s), le chercheur doit faire appel à un corpus bien déterminé : ce corpus peut être oral ou écrit, donc son analyse et son interprétation diffèrent d'une étude à une autre et dépendent des hypothèses émises au début du travail, de la ou des problématique(s) et du positionnement du chercheur dans son cadre théorique. Dans toute partie pratique, il faut qu'il y ait un corpus qui désigne l'ensemble des documents (écrits ou oraux, enregistrements audio ou vidéo, statistiques, ...) sur lequel portera l'étude.

1. Problématique

Il existe plusieurs types de corpus comme les enregistrements, les vidéos, les copies d'étudiants, les œuvres littéraires, les affiches publicitaires, ...etc. surtout en sciences humaines et sociales.

Ce corpus, une fois collecté, l'étudiant peut le conserver facilement et l'utiliser selon ses besoins. Mais actuellement, avec l'avènement de nouvelles approches dans un monde numérique, la nature des corpus a bien changé et l'étudiant se trouve devant une variété de choix de corpus à savoir le corpus numérique.

Donc, dans cette contribution, nous allons nous interroger : Selon quels critères choisir son corpus ? Quelles sont les différences entre un corpus traditionnel et un corpus numérique ?

Telles sont les questions qui nous permettraient de proposer des pistes de réflexion sur le choix du corpus et sa place dans une recherche scientifique.

2. Définitions du corpus

On parle de corpus représentatifs : de quoi ? Pour quoi ? Et comment ? (il y a quoi dans votre corpus ? dans votre panier ?) Il faut toujours justifier le choix du corpus qui doit fournir une représentation fidèle, précise et suffisamment d'éléments pour pouvoir identifier des comportements significatifs : c'est le corpus qui fait la théorie. Selon Rastier (2004) :

« Un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications. »

De son côté, Cuq (57 : 2003) le définit comme étant : « *un ensemble de données collectées par enregistrement, par observation directe, par questionnaire ou entretien, et réunies pour décrire et analyser un phénomène.* »

3. Les types de corpus

Il faut admettre qu'il n'existe pas de corpus « *parfait* » en soi ; l'essentiel est de sélectionner ou de constituer un corpus adapté à la finalité de l'utilisation que l'on souhaite en faire. L'étudiant doit donc choisir son corpus en fonction de sa problématique et en fonction des hypothèses qu'il veut tester sur le terrain. Selon Charaudeau (2009) : « *On envisage trois types de corpus selon qu'ils se réfèrent à une problématique dite cognitive, communicative ou représentationnelle.* »

De cette citation, nous remarquons que l'auteur relie le choix du corpus directement à la problématique. Dans certains cas de recherches, les étudiants font appel à un corpus témoin.

On parle souvent de corpus oraux, corpus écrits, corpus ouverts, corpus clos « *des corpus ouverts (auxquels on ajoute constamment de nouveaux textes) et fermés (qui restent « stables »).* Les corpus ouverts, bien que plus à jour, impliquent un entretien constant et méticuleux. » (Marshman, 3 :2003). De corpus diachroniques, corpus synchroniques, et actuellement, on est passé au corpus numérique ou « *Web-corpus* ».

4. Corpus traditionnel VS corpus numérique

Autrefois, pour préparer son mémoire de fin d'étude ou sa thèse de doctorat, l'étudiant devait fournir de grands efforts et se déplacer d'un endroit à un autre pour arriver à collecter les données qui seront considérées comme corpus, ce dernier qui est à la base de sa recherche par exemples des enregistrements, des vidéos, des copies d'étudiants (productions écrites ou autres), des œuvres littéraires, des affiches publicitaires, etc. Ce corpus, une fois collecté, l'étudiant peut le conserver facilement et l'utiliser selon ses besoins.

Pour ce type de corpus (que nous avons désigné par corpus traditionnel), ce qui pose problème c'est, d'abord la recherche du lieu de l'enquête et des personnes qui acceptent d'être enregistrées ou filmées puis avoir l'autorisation du haut responsable et des personnes qu'il va filmer ou enregistrer. C'est le cas des enregistrements qui se font au niveau des établissements scolaires (écoles primaires, collèges et lycées).

Par la suite, la collecte et le traitement de ces données qui nécessitent un temps et un effort considérables notamment quand il s'agit d'une étude quantitative.

Mais dernièrement la nature des corpus a bien changé et l'étudiant possède une variété de choix, il peut ne pas se déplacer, il suffit d'avoir un téléphone portable ou un PC connecté pour arriver à choisir son corpus disponible sous ses différentes formes : photos, textes, vidéos, documents sonores, etc. sans toutefois oublier de citer la source Internet : c'est le corpus numérique. Mayaffre a écrit en 2010 que :

« Là où le corpus était seulement perçu, traditionnellement, comme une idée ou une potentialité, le numérique désormais l'incarne, le réifie, le matérialise en le rendant, quelle que soit sa longueur, palpable et manipulable, exploitable et réexploitable, archivable et échangeable. (...) tous les linguistes de corpus travaillent aujourd'hui sur données numérisées, et tous réfléchissent à des méthodes numériques pour traiter leur objet numérique. »

Ce qui pose problème quand il s'agit de ce genre de corpus c'est l'accès à ces données qui n'est pas toujours autorisé et gratuit. Le deuxième problème c'est la durée de validité de certaines données qui ont une période de vie bien déterminée et aussi des conditions d'utilisation très strictes.

Pour ce type de corpus, il existe différents logiciels qui facilitent sa collecte et son analyse, ce qui permet de gagner l'effort et le temps et donner des résultats fiables et exacts.

Conclusion

Enfin, nous admettons que le choix d'un corpus et son analyse posent d'insurmontables difficultés aux étudiants. Qu'il soit traditionnel ou numérique, le corpus constitue une étape importante dans la vie d'un chercheur qui doit le choisir minutieusement pour répondre à ses questions de départ et confirmer ou infirmer ses hypothèses.

Références bibliographiques

- CHARAUDEAU, Patrick, MAINGUENEAU, Dominique. (2002), *Dictionnaire d'analyse du discours*, Paris, Seuil.

Articles

- Marshman, Elizabeth. (2003), *Construction et gestion des corpus : Résumé et essai d'uniformisation du processus pour la terminologie*.
- Mayaffre, Damon, (2010), *Corpus et web-corpus. Réflexion sur la corporalité numérique*. In, Cahiers de praxématique, p. 233-248

Références électroniques

- CHARAUDEAU, Patrick. (2009), *Dis-moi quel est ton corpus, je te dirai quelle est ta problématique*, consulté le 16 janvier 2022. URL :

<http://journals.openedition.org/corpus/1674> ;

DOI :

<https://doi.org/10.4000/corpus.1674>

[Sommaire](#)[Document précédent](#)[Document suivant](#)

8 | 2009 : Corpus de textes, textes en corpus

- RASTIER, François. (2004), *Enjeux épistémologiques de la linguistique de corpus*. consulté le 16 janvier 2022. URL: <http://www.revue-texto.net/Inedit/Rastier/Rastier_Enjeux.html>.