Master « Linguistique »

Kheira Mérine

 **Le corpus : objet, concept et théorie**

**Section 2**

1. **Compétence :** distinguer les différentes catégories de corpus pour une meilleure utilisation.
2. Nature du corpus :

D’après D. Mayaffre (2005), le corpus peut être soit un outil qui permet *de rendre compte d’une réalité transcendante (la langue ?), d’accéder à un monde déjà-là, d’illustrer une connaissance a priori, de ‘’découvrir’’ un savoir déjà su* », soit *« un objet vivant de recherche et de connaissance, en lui-même, dont la description débouchera sur des modèles sémantiques à inventer »* (Mayaffre, 2005)[[1]](#footnote-2)

Nous avons-là une définition générale illustrant les cas de figure où on sent la nécessité de recourir à un corpus. Elle nous expose, cependant deux types de corpus :

:

(i) celui qui nous mène vers le déjà connu, le travail serait donc de confirmer ce qui a déjà été attesté dans d’autres circonstances, ce serait donc un travail mené sur la base d’une théorie, soit pour l’illustrer ou pour garantir la fiabilité de ce corpus qui répondrait à cette théorie.[[2]](#footnote-3)

Dans cette démarche la formation du corpus doit répondre à l’exigence d’une théorie, ce que Mayaffre appelle « l’observatoire d’une théorie » ou alors le »corpus support »

(ii) Celui dont la description va aboutir vers quelque chose de tout à fait nouveau et qui va conditionner le théorique. Ce type de corpus questionne « l’épistémè » de la discipline, il est le « corpus apport ».[[3]](#footnote-4)

3. **Evolution de la notion de corpus dans la sphère linguistique.**

Au départ la notion de « corpus » a divisé le monde de la linguistique en deux parties. La première regroupe les « théoriciens » qui refusaient d’emblée l’existence de corpus dans une démarche linguistique ; parmi eux, il y a surtout les générativistes, l’exemple en est cette déclaration de Chomsky « Corpus linguistics does not exist » (entretien avec BaasAarts, 1999) La deuxième partie est celle des « descriptivistes » qui ne considéraient aucune théorie sans passer par la description d’un corpus.

Pour les premiers, attachés à la théorie Saussurienne telle que l’a présentée C. Bally en précisant l’objet de la linguistique qui est « la langue envisagée en elle-même et pour elle-même », il n’y a pas lieu d’utiliser un corpus, car tout ce qui provient de corpus, surtout textuels, où il s’agit de performances dans des situations concrètes, relèverait de la sociolinguistique, de la psycholinguistique, de l’analyse du discours ou même de la littérature (Mayaffre, 2005). Ce rejet pourrait s’expliquer par le fait que le fondamentalisme grammatical n’acceptait pas d’être remis en cause par les emplois de l’usage qu’il considérait comme étant des atteintes au système.

Les descriptivistes, appartenant au deuxième groupe, rejettent toute théorie si ce n’est celle qu’ils déterminent à partir d’un corpus ; pour eux, le sens est fixé par l’usage, donc la signification d’une unité linguistique, lexie ou syntagme ne peut être définie qu’à partir d’usages recueillis sous forme de corpus. Leur courant a influencé plusieurs entreprises dont celle du TLF (Trésor de la langue française) qui n’explique le sens d’une entrée lexicale qu’à travers ses différents emplois recueillis dans des romans ou autres écrits littéraires.

Les linguistes appartenant à la position médiane considèrent que la théorie est nécessaire mais pas suffisante pour décrire des comportements langagiers qui souvent sont entachés de singularité.

4. **Forme(s) du corpus.**

Là aussi les avis se divisent en deux grands groupes : ceux qui reconnaissent une pluralité formelle au corpus, et ceux qui n’y voient qu’une seule forme liée au texte ou au macrotexte.

4.1. Pluralité formelle du corpus

 Pour les linguistes qui partagent ce point de vue ( Dalbéra,2002, Mayaffre, 2005…). C’est dans un souci méthodologique que se fait le choix du corpus.

Concernant ce point, Mayaffre (op.cit.) établit une sorte de hiérarchie en présentant trois grandes formes de corpus :

-Le corpus lexicographique dont la spécificité est de prétendre à l’exhaustivité Et de ce fait, ils présentent un caractère clos..

-Le corpus phrastique utilisé pour des recherches en syntaxe et morphosyntaxe. Ce corpus peut être recueilli comme il peut être construit (fabriqué), comme on l’a vu plus haut.

-Le corpus textuel qui pose le problème du sens. Pour Mayaffre (ibid), ce type de corpus exclut toute idée d’exhaustivité et, offrant plusieurs possibilités d’interprétations , il ne peut permettre une véritable détermination du sens. On peut fabriquer des corpus réduits (mot/phrase) mais pas un texte.

1. [↑](#footnote-ref-2)
2. K. Mérine (2015) [↑](#footnote-ref-3)
3. ibid [↑](#footnote-ref-4)