

Retours sur les aspects quantitatifs pour la description en syntaxe.

Mireille Bilger

(Université de Perpignan-via-Domitia, CRESEM)

Dans le domaine de la lexicographie, le fait d'avoir recours aux corpus est une démarche acceptée depuis longtemps (Teubert 2002), concernant le domaine grammatical, cela est beaucoup plus récent et a même suscité, au début, quelques résistances, qui semblent aujourd'hui avoir disparu. De fait, actuellement, il existe un réel consensus pour considérer les corpus comme d'excellents instruments d'observation, de description et d'enseignement-apprentissage de la langue (McEnery & Xiao, 2010, Boulton & Tyne, 2014).

L'exploitation, à la fois quantitative et qualitative, de corpus de médium et de genres différents, (écrit littéraire, journalistique, scientifique / oral de parole privée, publique, professionnelle, etc.), permet de révéler des faits de langue susceptibles de catégoriser à la fois les productions langagières et les usages (Biber, 2006, Cheng, 2012). Les travaux, plutôt anglo-saxons, issus de cette exploitation ont ainsi permis de réinterroger, entre autres, la notion de « norme unique d'acceptabilité » pour laisser place à des normes linguistiques variées dépendantes du genre, des registres et du contexte énonciatif (Biber & *alii* 1999). Pour le français, même si comme le dit Blanche-Benveniste (2010 : 84) : « la description syntaxique doit tenir compte des répartitions en *genres* et ne peut pas faire totalement abstraction de la fréquence d'attestation dans les données », les études restent encore parcellaires (Bilger & Cappeau 2015). S'intéresser aussi bien au nombre d'occurrences recueillies qu'à la variété « des genres », d'où sont issues les données (ce qui est aujourd'hui rendu possible par une disponibilité plus grande de corpus de types différents), semble donc nécessaire pour enrichir la description.

La comparaison entre les données, quand celle-ci est menée, s'effectue en grande partie, via l'utilisation de concordanciers, par le comptage des formes (et des collocations entre les formes), ces comptages ne sont cependant pas anodins (Poibeau 2014) et méritent que l'on y revienne.

Nous nous proposons ici, à travers diverses illustrations issues d'études sur corpus, de revenir brièvement sur les interrogations déjà formulées par Habert (2000):

- comment déterminer un nombre d'occurrences significatif ? Et significatif de quoi ? Comment interpréter des faits de langue très fréquents ou au contraire peu attestés dans les corpus étudiés ? Quels sont les écueils ? (requêtes difficiles à formaliser, taille du contexte, etc.)

Selon que le travail porte sur la description de tournures syntaxiques ou sur le « genre » des productions, le nombre d'occurrences nécessaire peut être différent et la dimension quantitative ne pas remplir les mêmes services. Cette notion de fréquence, régulièrement exploitée dans une perspective variationniste, nous permettra, par ailleurs, d'interroger certaines oppositions, encore trop souvent présentées comme fondamentales, en particulier celle que l'on pose entre oral et écrit.

Biber, D., 2006. *University Language. A corpus-based study of spoken and written registers*. Amsterdam. John Benjamins Publishing Company.

Biber, D., & *alii*, 1999, *Longman Grammar of spoken and Written English*. Harlow : Pearson education Limited.

Bilger, M. & Cappeau, P. , 2015, « Comment les données de corpus pourraient renouveler les manuels de grammaire ? », H. Tyne (éd), revue *LINX*, 68-69.177-199

- Boulton, A., & Tyne, H., 2014, *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*, Paris, Didier.
- Blanche-Benveniste, C., 2010. *Le français : Usages de la langue parlée*. Leuven. Peeters
- Cheng, W., 2012. *Exploring Corpus Linguistics. Language in Action*. Abingdon. Routledge.
- Habert, B., 2000. « Des corpus représentatifs : de quoi, pour quoi, comment ? » dans Bilger, Mireille (éd). *Linguistique sur corpus – Etudes et réflexions*. Perpignan. P.U. de Perpignan. 11-58.
- McEnery, T. & Xiao, R., 2010, « What corpora can offer in language teaching and learning », In E. Hinkel (éd.), *Handbook of Research in Second Language Teaching and Learning*, Londres, Routledge, p. 364-380.
- Poibeau, T., 2014, « La linguistique est-elle soluble dans la statistique ? », *Revue Sciences/Lettres*, 2 | 2014.
- Teubert, W., 2002, « Corpus Linguistics and Lexicography », *International Journal of Corpus Linguistics*, n° 6, 2002, p. 125-154.