



MémoDico

Recherche

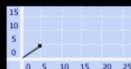
Liens utiles

Trucs & astuces

UPV

Plan du site

Selon *Wikipédia*, la **statistique** est « l'étude d'un phénomène par la collecte de données, leur traitement, leur analyse, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques ».



MémoInfos

Néologismes

Notes en vrac

Veille linguistique

Moteur de recherche

Contact

Nous allons tenter de voir ci-dessous plus simplement de quelle façon cette science peut nous rendre service et dans quelle mesure.

→ Dans la présentation de *Wikipédia* déjà citée en haut de cette page, on nous rappelle aussi que la statistique est pour les uns un domaine des mathématiques, et l'on ajoute (ce qui est plus intéressant pour nous ici) que pour les autres (en particulier les anglo-saxons), il s'agit d'une discipline à part entière hors des mathématiques, enfin de plus en plus.

Elle fait partie de ce que l'on appelle aujourd'hui la **science des données** (en anglais : **Data Science**). Elle possède une composante théorique ainsi qu'une composante appliquée, celle que nous utiliserons ici. La statistique appliquée est utilisée dans presque tous les domaines de l'activité humaine : ingénierie, management, économie, biologie, informatique, etc. La statistique utilise des règles et des méthodes sur la collecte des données, pour que celles-ci puissent être correctement interprétées, souvent comme composante d'une **aide à la décision**, ce qui peut être appréciable par exemple en matière de lexicographie.

→ Les statistiques sont une aide que j'ai adoptée et adaptée petit à petit au cours de ma carrière et qui peut constituer parfois une sorte de marqueur d'usage intéressant en lexicographie bilingue notamment : les **statistiques lexicales**, qui sont de nos jours à la portée de tout un chacun grâce aux différents moteurs de recherche présents sur le web. En écrivant un article intitulé « Deux langues romanes à la sauce anglaise », qui avait d'ailleurs pour sous-titre « De l'anglo-américain en français et en espagnol » (à voir [ici](#) si vous le souhaitez), je m'étais à plusieurs reprises « amusé » à voir par exemple à travers les chiffres rendus par *Google* (entre autres outils de même type) si telle ou telle lexie dans une langue donnée était plus ou moins fréquente que son ou ses équivalents de traduction.

→ Dans les dictionnaires que j'ai confectionnés dernièrement, notamment le dernier, *l'Adelantel - Diccionario español-francés / francés-español contextual y comentado de la lengua en marcha* (2017) (dont vous pourrez trouver une brève description [ici](#)), j'ai souvent utilisé les statistiques lexicales pour décider du choix de telle ou telle traduction et pour élaborer des commentaires destinés à éclairer le lecteur sur les fréquences d'usage et autres détails plus ou moins importants.

Pour traduire par exemple la locution verbale **faire fureur**, la plupart des dictionnaires bilingues ou multilingues proposent en espagnol **causar furor** ou **hacer furor**. Or il faut savoir que ces deux équivalents de traduction ne sont pas les meilleurs (surtout le premier), car le plus fréquent en espagnol actuel est de dire **ser furor**, et ceci nous a été "soufflé" par une rapide étude statistique réalisée grâce à un ou plusieurs moteur(s) de recherche. Les statistiques peuvent donc être déterminantes pour certains choix de pur lexique.

Autre exemple, pour traduire fr. **buffet libre**, nous hésitions entre esp. **buffet libre** et **bufé libre**. Or, même si le mot **buffet** a été espagnolisé en **bufé**, après une rapide double recherche, il est apparu que cette seconde graphie apparaît seulement dans 6 % des occurrences trouvées sur Internet (le samedi 15 février 2015). Cela participe d'ailleurs d'une tendance générale de l'espagnol actuel à conserver le mot emprunté tel quel. Ici donc, le sondage de fréquence nous renseigne clairement sur l'orthographe à privilégier.

Parfois aussi, les statistiques lexicales peuvent nous renseigner sur le genre de la lexie en question. Nous avons pu constater à travers une rapide étude statistique sur Internet en date du 28 février 2014 que la locution nominale **3D** est employée en français au féminin dans l'immense majorité des cas (99,8 %), alors qu'en espagnol, la tendance est à l'inverse et on dit **el 3D** dans 81 % des cas et **la 3D** dans seulement 19 %.

En ce qui concerne enfin le mot fr. **anticopie**, alors qu'en français le pluriel est marqué sur l'adjectif (**des systèmes anticopies**) dans 31 % des cas (contre 69 % de cas où l'adjectif est invariable en nombre), en espagnol on peut considérer que l'adjectif correspondant est invariable puisqu'on ne trouve pratiquement que **sistemas anticopia** (dans 99 % des cas, contre 1 % seulement avec **anticopias** et un pluriel marqué). La statistique en comparaison entre les deux langues apporte dans ce cas encore une précision importante (morphologique en l'occurrence), car le français se montre hésitant parfois sur le pluriel du nom en apposition (on dit **des enfants prodiges**) alors que l'espagnol laisse le pluriel de côté pour le second substantif (**niños prodigio**).

→ Même s'il faut absolument ne pas faire une confiance aveugle en les résultats que peuvent nous retourner les **moteurs de recherche** (quels qu'ils soient d'ailleurs), même s'il faut interpréter leurs réponses chiffrées avec circonspection, leur aide est donc tout de même indéniable et leur utilité en linguistique devrait se confirmer encore plus largement dans les années à venir.



Jean-Louis BARREAU, le 26 novembre 2018

Afficher la page en 'live'

Fin de la page. Cliquez ← pour retourner à la précédente.



[Haut de page]

Retour à l'accueil



[Haut de page]