

A la croisée des corpus et de la phraséologie : une proposition d'outil informatique

Jean-Pierre Colson

Résumé : Les corpus connaissent aujourd'hui un succès grandissant et ne sont plus l'apanage des linguistes de corpus. Ils nous renvoient l'image de la langue telle que la décrivait John Sinclair (1991): une bonne moitié de tous les textes est sans doute constituée d'unités phraséologiques au sens le plus large. Vouloir enseigner les langues étrangères sans tenir compte de cette composante majeure du langage est donc une gageure.

Le recours aux corpus pour la didactique de la phraséologie est toutefois semé d'embûches : la fréquence des unités phraséologiques est faible et il faut manipuler de très vastes corpus dans l'espoir de découvrir des exemples pertinents. Les simples concordances d'unités phraséologiques sont parfois contradictoires ou difficiles à interpréter ; enfin, les outils nécessaires à la manipulation d'aussi vastes collections de textes pèchent souvent par un manque de convivialité.

Nous proposons de revenir à une notion déjà proposée par Michael Lewis en 1993 : la prise de conscience phraséologique, par le biais de confrontations nombreuses à des extraits de corpus, mais en mettant à profit de nouveaux outils qui doivent encore être perfectionnés. Nous proposons un tel outil expérimental, l'application *IdiomSearch* (J.-P. Colson 2016), qui offre aux apprenants un retour immédiat sur la présence d'unités phraséologiques dans un texte source, ainsi qu'une classification par fréquence et par figement.

Mots-clés : phraséologie, didactique des langues, corpus, sémantique statistique, collocations.