

Avant-propos

Depuis les années 1990, la linguistique a progressivement connu un tournant méthodologique fondamental. D'une discipline essentiellement rationaliste depuis le milieu du XX^e siècle suite aux travaux du linguiste américain Noam Chomsky, elle s'est progressivement (ré)ouverte aux approches empiriques représentées par la linguistique de corpus et la linguistique expérimentale. Cette transition s'est encore accélérée durant la dernière décennie, à tel point que la majorité des travaux en linguistique publiés dans des revues internationales font actuellement usage de données empiriques. Ainsi, les corpus linguistiques se sont progressivement imposés comme des outils fondamentaux pour les linguistes et leur utilisation s'est généralisée à l'ensemble des domaines de la linguistique, y compris ceux traditionnellement rationalistes comme la syntaxe. Le développement de la linguistique de corpus a entraîné la création de nouvelles méthodes pour collecter et analyser des données linguistiques, rendues possibles grâce au développement de l'informatique et à l'arrivée d'Internet. Cette nouvelle orientation de la linguistique a permis des avancées spectaculaires pour l'étude scientifique des nombreuses facettes du langage humain dans toute leur complexité. Cet ouvrage vise à faire découvrir cette richesse à des lectrices et lecteurs peu habitués à lire la littérature spécialisée en linguistique.

À l'heure actuelle, la capacité à analyser de manière quantitative des données de corpus fait ainsi partie intégrante de la boîte à outils des linguistes. Toutefois, l'utilisation de telles données repose sur des principes théoriques et méthodologiques précis, qui doivent être bien maîtrisés. Ce tournant en linguistique implique donc la nécessité d'initier les nouvelles générations d'étudiant·e·s à l'usage de ces méthodes afin de leur permettre de comprendre les enjeux qui sous-tendent leur utilisation dans la littérature scientifique, d'évaluer de manière critique les résultats obtenus, ainsi que d'utiliser ces méthodes dans le cadre de leurs travaux universitaires. Ce livre est conçu comme un support pédagogique pour les étudiant·e·s et,

de manière générale, toutes les personnes qui souhaitent s'initier à l'utilisation de corpus en linguistique.

La matière présentée dans cet ouvrage ne présuppose pas de compétences préalables autres que des connaissances de base en linguistique, ainsi qu'une maîtrise minimale des outils informatiques les plus courants, comme les logiciels de tableurs. L'ouvrage est pensé à la fois comme un support pour accompagner un enseignement universitaire d'initiation à la linguistique de corpus et comme outil permettant aux étudiant·e·s qui souhaitent se former à l'utilisation de corpus de travailler de manière autonome grâce aux questions de révision présentées à la fin de chaque chapitre et à leur corrigé détaillé en fin d'ouvrage.

De par son caractère introductif, cet ouvrage est nécessairement partiel et ne traite pas toutes les questions soulevées par l'usage de corpus dans différentes disciplines de la linguistique. Il ne couvre pas non plus certaines méthodes d'analyses avancées nécessitant un haut niveau de compétences en informatique et en statistiques afin d'analyser les données. Toutefois, des suggestions de lectures présentées à la fin de chaque chapitre permettront à celles et ceux qui le souhaitent d'approfondir l'un ou l'autre des aspects présentés.

Enfin, cet ouvrage met une emphase particulière sur le français comme objet d'étude. En effet, si la linguistique de corpus s'est imposée de manière incontestable dans le monde anglo-saxon et qu'une proportion importante des chercheurs francophones utilise actuellement ces méthodes, son enseignement reste encore marginalisé en France. Ce livre a donc aussi pour objectif de montrer la vitalité et la richesse des études de corpus portant sur le français, ainsi que de recenser les ressources les plus importantes qui ont été développées pour cette langue, en espérant apporter ainsi une contribution à l'essor de cette discipline dans le monde francophone.