



Lidil

Revue de linguistique et de didactique des langues

56 | 2017

Grammaires et littéracies

Céline Poudat et Frédéric Landragin, *Explorer un corpus textuel : Méthodes – pratiques – outils*

Louvain-la-Neuve, De Boeck Supérieur, 2017, 240 p.

Eva Schaeffer-Lacroix



Édition électronique

URL : <http://journals.openedition.org/lidil/4720>

DOI : [10.4000/lidil.4720](https://doi.org/10.4000/lidil.4720)

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

ISBN : 978-2-37747-024-2

ISSN : 1146-6480

Référence électronique

Eva Schaeffer-Lacroix, « Céline Poudat et Frédéric Landragin, *Explorer un corpus textuel : Méthodes – pratiques – outils* », *Lidil* [En ligne], 56 | 2017, mis en ligne le 01 novembre 2017, consulté le 25 septembre 2020. URL : <http://journals.openedition.org/lidil/4720> ; DOI : <https://doi.org/10.4000/lidil.4720>

Ce document a été généré automatiquement le 25 septembre 2020.

© Lidil

Céline Poudat et Frédéric Landragin, *Explorer un corpus textuel : Méthodes – pratiques – outils*

Louvain-la-Neuve, De Boeck Supérieur, 2017, 240 p.

Eva Schaeffer-Lacroix

RÉFÉRENCE

Céline Poudat et Frédéric Landragin, *Explorer un corpus textuel : Méthodes – pratiques – outils*, Louvain-la-Neuve, De Boeck Supérieur, 2017, 240 p.

- 1 L'ouvrage *Explorer un corpus textuel : Méthodes – pratiques – outils* de Céline Poudat et de Frédéric Landragin a paru en mars 2017. Les auteurs visent un public d'étudiants en sciences du langage, mais aussi tout étudiant ou chercheur d'une autre discipline intéressé par l'analyse des données textuelles à l'aide des outils et méthodes de la linguistique de corpus.
- 2 Les auteurs ont pour objectif « d'expliquer de manière progressive et pédagogique ces méthodes de la linguistique outillée qui sont orientées vers l'exploration de corpus, et donc vers l'accès aux données, leur visualisation, le calcul de statistiques » (p. 10). Il ne s'agit pas d'une annonce vaine : le style est limpide et abordable tout en restant scientifiquement valide. Certains éléments ou caractéristiques sont évoqués de façon brève parce que la connaissance de leur existence est considérée comme une condition pour la bonne compréhension de la suite ; citons comme exemple le langage SQL (Structured Query Language) dont on se contente de mentionner l'existence et l'importance (p. 23), ce qui est suffisant dans le cadre posé par les auteurs.
- 3 On peut également lire que le propos des auteurs « ne relève pas de la technique : il est méthodologique avant tout et vise à fournir des repères sur le fonctionnement des outils en général, et non sur l'utilisation de tel ou tel outil particulier » (p. 10). Il ne s'agit visiblement pas de présenter toutes les fonctionnalités d'un outil d'exploration de

corpus en particulier. On peut tout de même constater qu'au fil du texte, des noms d'outils sont cités, certains principalement pour illustrer un point précis (Glozz ou Annis). D'autres comme TXM ou IRaMuTeQ apparaissent de façon récurrente ; de plus, en annexe, un paragraphe descriptif leur est dédié. Le lecteur désireux d'approfondir ses connaissances sur l'un ou l'autre outil est invité à lire certains des ouvrages ou articles cités dans Poudat & Landragin (2017). Il est également recommandé de consulter les synthèses descriptives attestant des usages d'outils d'exploitation de corpus par la communauté francophone, publiées en ligne par Céline Poudat pour le compte du groupe de travail Exploration de corpus du consortium CORLI (CORpus, Langues, Interactions), affilié à la très grande infrastructure de recherche des humanités numériques Huma-Num.

- 4 Afin de compléter les suggestions des auteurs, on peut proposer la lecture du livre *Discovering English with Sketch Engine* de James Thomas (2017) pouvant convenir aux personnes désirant connaître le système de gestion de corpus Sketch Engine qui a des vertus pédagogiques avérées. N'oublions pas de mentionner les autres moyens de s'informer sur l'un ou l'autre outil de façon exhaustive et conviviale : des manuels en ligne (CQPweb et TXM), des tutoriels filmés (Corpus-Workbench, 2012-2017) et des listes de diffusion (Corpus, TXM).
- 5 En conclusion, les auteurs rappellent les principes qui sous-tendent la rédaction de l'ouvrage et qui caractérisent leur positionnement général : des méthodes d'exploration de corpus répandues et ayant fait leurs preuves dans la communauté francophone et au-delà, le choix de logiciels libres, l'insertion dans des communautés de chercheurs comme le consortium CORLI, et une clarté dans le propos pour donner aux lecteurs novices l'occasion de progresser dans un domaine en pleine expansion. On y apprécie le rappel de certains débats encore actuels, en particulier celui autour du sens de l'annotation : « [...] l'idée de décompter ce que l'on a annoté et déjà interprété, et de s'éloigner de la matérialité textuelle peut encore poser problème à certains. » (p. 215) Les auteurs plaident pour une meilleure coopération entre communautés travaillant sur l'annotation et communautés spécialistes en exploration quantitative des données (p. 215). Le développement de formats standards, par exemple pour les annotations de type XML-TEI, est également considéré comme une urgence.
- 6 Pour finir, on peut saluer la publication de cet ouvrage qui soutient parfaitement une évolution nécessaire, à savoir la banalisation des méthodes et outils de corpus dans le monde de la recherche.

AUTEURS

EVA SCHAEFFER-LACROIX

ESPE de l'académie de Paris, Université Paris-Sorbonne, CeLiSo