

Atelier Corpus et méthodes pour l'analyse outillée des variétés de français à travers le monde (CoMeVar)

Organisé par Sascha Diwersy (Praxiling- UMR 5267, U. Montpellier) & Mathieu Avanzi (Université de Neuchâtel)

Plus de quinze ans après la naissance du *Corpus International Écologique de la Langue Française* (CIEL-F)¹ et au vu des avancées constantes que connaissent les méthodes et outils de traitement et d'exploration de données textuelles, l'atelier que nous nous proposons d'organiser a pour objectif de dresser l'état de l'art et d'esquisser des perspectives de recherche dans le domaine de la linguistique de corpus appliquée aux variétés de français à l'échelle mondiale, non sans tenir compte de l'essor considérable que connaît cette discipline dans les recherches sur les « World Englishes » depuis une trentaine d'années². Notre atelier sera l'occasion de réunir des chercheur.e.s qui portent un certain intérêt à l'analyse outillée de la variation diatopique du français et qui sont spécialistes de différentes variétés, que celles-ci relèvent des francophonies américaines, africaines ou européennes. Le programme de l'atelier s'articulera autour des questions de recherche suivantes :

- Quel est le panorama des corpus dédiés aux différentes zones géographiques de l'espace francophone ? Quelle est la place qu'ils prennent au sein des consortiums au niveau de la TGIR HumaNum ou des GDR qui bénéficient d'un soutien du CNRS ? Quelles sont les perspectives de curation de ces ressources et de leur pérennisation à travers des infrastructures de dépôt centralisées comme Ortolang ou encore CLARIN au niveau européen ?
- Quels sont les défis qui sont posés d'une manière générale par la constitution d'un corpus de référence des variétés de français à l'échelle mondiale au regard des expériences faites dans le cadre d'un projet comme le CIEL-F ? En quoi le modèle de corpus du même type existant pour d'autres langues, et notamment celui du *International Corpus of English*, est-il transposable à la francophonie ?
- Dans quelle mesure les ressources et outils de TAL dont on dispose actuellement pour le français sont-ils adaptés à la prise en charge de la variation diatopique ?
- Quels sont les enjeux spécifiques aux procédés lexico-statistiques déployés dans l'analyse des variétés de français ? Que peuvent apporter à ces études des méthodes multi-variées avancées³ ?
- Quels sont les logiciels et plateformes d'interrogation spécifiquement dévolus à l'exploration des (données issues de) corpus de variétés de français ? Quel est leur impact d'un point de vue applicatif dans les domaines de la lexicographie, grammaticographie ou de la didactique de la langue française ?

¹ Gadet et al. 2012 ; Mondada & Pfänder 2016.

² En témoignent entre autres le manuel *Corpus Linguistics for World Englishes* (Lange & Leuckert 2019) ainsi que l'existence, depuis le début des années 1990, du *International Corpus of English ICE* (Greenbaum 1990), qui fait preuve d'une grande vitalité (cf. Kirk & Nelson 2018).

³ On peut penser ici au paradigme méthodologique MuPDAR(F) déployé dans l'étude des variétés d'anglais par Heller, Bernaisch & Gries (2017) ou Gries (2022).

Le programme prévoit les contributions suivantes :

1. Mathieu Avanzi et Laure Anne Johnsen, Université de Neuchâtel
2. Christophe Benzitoun, ATILF, Université de Lorraine Nancy
3. Anne Dister, Université Saint-Louis
4. Sascha Diwersy, PRAXILING, Université Paul-Valéry Montpellier 3
5. Stefan Gries, University of California, Santa Barbara
6. Stefan Pfänder, Université de Freiburg
7. Wim Remysen, Université de Sherbrooke / CRIFUQ

La matinée sera consacrée à des présentations de corpus francophones, dans une perspective décentralisée. L'après-midi comportera des présentations plus longues mettant en lumière les problèmes liés aux méthodes (outils disponibles et à développer, interopérabilité, etc.)

Références

- Gadet, F., Ludwig, R., Mondada, L., Pfänder, S. & Simon, A. (2012). Un grand corpus de français parlé : le CIEL-F: Choix épistémologiques et réalisations empiriques. *Revue française de linguistique appliquée*, XVII, 39-54. <https://doi.org/10.3917/rfla.171.0039>
- Greenbaum, S. (1990). The International Corpus of English. *ICAME Journal*, 14, 106–108.
- Gries, S. T. (2022). MuPDAR for corpus-based learner and variety studies: two (more) suggestions for improvement. In Flach, S. & Hilpert, M. (eds.), *Broadening the spectrum of corpus linguistics: New approaches to variability and change*, 257-283. Amsterdam & Philadelphia: John Benjamins.
- Heller, B., Bernaisch, T. J., & Gries, S. T. (2017). Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal*, 41. 111-144.
- Kirk, J. & Nelson, G. (2018). The International Corpus of English project: A progress report. *World Englishes*, <https://doi.org/10.1111/weng.12350>.
- Lange, C. & Leuckert, S. (2019). *Corpus Linguistics for World Englishes: A Guide for Research*, London: Routledge.
- Mondada, L. & Pfänder, S. (2016). Corpus international écologique de la langue française (CIEL-F) : un corpus pour la recherche comparée sur le français parlé, *Corpus*, 15.
URL : <http://journals.openedition.org/corpus/2951> ; DOI : <https://doi.org/10.4000/corpus.295>.