



**HAL**  
open science

# L'enseignement et l'apprentissage de l'écrit académique à l'aide de corpus numériques

Cristelle Cavalla, Laura Hartwell

## ► To cite this version:

Cristelle Cavalla, Laura Hartwell. L'enseignement et l'apprentissage de l'écrit académique à l'aide de corpus numériques. LIDIL - Revue de linguistique et de didactique des langues, 2018, L'enseignement et l'apprentissage de l'écrit académique à l'aide de corpus numériques. hal-01918278

**HAL Id: hal-01918278**

**<https://hal.science/hal-01918278>**

Submitted on 10 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Version n-1 de l'article suivant :

Cavalla Cristelle et Hartwell Laura, « Présentation », *Lidil* [En ligne], 58 | 2018, mis en ligne le 02 novembre 2018, consulté le 10 novembre 2018. URL : <http://journals.openedition.org/lidil/5367>

## Présentation du numéro *Lidil* 58

### L'ENSEIGNEMENT ET L'APPRENTISSAGE DE L'ÉCRIT ACADÉMIQUE À L'AIDE DE CORPUS NUMÉRIQUE

Coordinatrices : **Cristelle Cavalla** (Université Sorbonne Nouvelle-Paris3, *DILTEC*)  
**Laura Hartwell** (Université Toulouse 1 Capitole, *Lairdil*)

Le numéro 58 de *Lidil* s'inscrit dans la tradition des études sur les corpus numériques (cf. par exemple : Fløttum, Dahl & Kinn, 2006 ; Hunston & Francis, 2000 ; Swales, 1990/2004 ; Tutin & Grossmann, 2014 ; Quirk, Greenbaum, Leech & Svartvik, 1985) et, cette fois, l'enseignement et l'apprentissage des langues étrangères sont à l'honneur. Nous prenons le parti de définir un corpus numérique, selon le point de vue linguistique, comme un ensemble de documents écrits ou enregistrés et assemblés dans l'objectif de faire une analyse de leur contenu langagier à l'aide d'outils informatiques. Ces corpus servent directement ou indirectement pour l'enseignement et l'apprentissage via des modalités multiples. Dans les contextes d'apprentissage ou d'enseignement, les corpus peuvent prendre des formes diverses, depuis des textes ciblés jusqu'à de grandes bases de données. Nous trouvons aussi des produits dérivés de corpus tels que des listes de vocabulaires (Coxhead, 2000), des dictionnaires (Collins, 2017 ; MacMillan, 2017), des matériaux pédagogiques (Chambers, 2010). Les avantages réels de l'exploitation des corpus numériques à des fins didactiques et les difficultés de cette pratique, rencontrées ou imaginées, demeurent un vaste terrain d'étude.

Dans ce numéro, nous nous concentrons sur l'enseignement et l'apprentissage de l'écrit académique à l'aide des corpus numériques. Les auteurs expliqueront alors l'intérêt d'utiliser de telles ressources relevant des méga-données pour rédiger des écrits académiques dans une langue étrangère. Objets de nombreux travaux linguistiques et didactiques depuis les années 90 dans le monde francophone (cf. *Lidil* 18, *Pratiques de l'écrit et modes d'accès au savoir dans l'enseignement supérieur*, coordonné par Michel Dabène et Yves Reuter). Nous trouvons

actuellement des corpus disponibles en ligne, tels que les corpus anglais du BAWE<sup>1</sup> ou du Michigan Corpora<sup>2</sup>, et des corpus français/anglais tel que Scientext<sup>3</sup> interrogeable via l'interface *ScienQuest*<sup>4</sup>, permettent aux chercheur.e.s d'identifier les caractéristiques du discours académique des expert.e.s et des étudiant.e.s.

Sans revenir sur les avantages des grands corpus numériques pour la description linguistique (cf. *Lidil* 31, *Corpus oraux et diversité des approches*, coordonné par Marie Savelli), cette publication est l'occasion de découvrir de nouvelles pistes facilitant l'utilisation en classe de langue de corpus numériques pour l'aide à la rédaction universitaire. En termes didactiques, les articles rassemblés ici traitent de l'utilisation directe ou indirecte des corpus numériques, tant par les enseignant.e.s que par les apprenant.e.s dans un contexte universitaire. Depuis quelques années, l'essor des outils informatiques permettant le stockage et surtout l'interrogation en ligne par des non spécialistes a permis le développement des corpus numériques à vocation pédagogique. De fait, les articles de ce numéro explorent l'exploitation de ces ressources numériques pour l'enseignement et l'apprentissage de la langue académique et certains y ont ajouté la phraséologie scientifique.

Dans la lignée des travaux de Boulton et Tyne (2014) sur l'utilisation des corpus en classe de langue, les contributions de ce numéro apportent des réponses quant à l'utilisation effective des corpus numériques par les équipes enseignantes à l'université : comment et pourquoi les utilisent-ils ? Nous verrons que les approches linguistiques et didactiques se font écho d'une langue à l'autre et que la didactique des langues fédère ces différentes études.

Autour de l'anglais académique, Clive Hamilton et Natalie Kübler, qui enseignent à l'université Diderot à Paris, développent des cours autour des corpus d'anglais scientifique pour des étudiants francophones spécialistes de différentes disciplines universitaires. Le cas d'étudiant.e.s utilisant directement des corpus en ligne reste peu courant à l'université française et cette étude de cas nous aide à mieux saisir les enjeux de cette pratique.

Benet Vincent et Hilary Nesi, de l'université de Coventry, enseignent l'anglais académique à des étudiants internationaux à l'aide d'une interface du BAWE corpus dans laquelle des liens hypertextuels sont pensés pour guider et aider les apprenants dans leur écriture académique. La question de l'insertion de tels outils en classe de langue est posée, permettant de mettre en

---

<sup>1</sup> BAWE : <http://ota.ahds.ac.uk/headers/2539.xml>, voir aussi : [www.coventry.ac.uk/bawe](http://www.coventry.ac.uk/bawe).

<sup>2</sup> Michigan Corpora : <https://lsa.umich.edu/eli/language-resources/micase-micusp.html>.

<sup>3</sup> Scientext : <http://scientext.msh-alpes.fr/scientext-site/spip.php?article1>.

<sup>4</sup> ScienQuest : <http://corpora.aiakide.net/scientext18/>.

exergue les spécificités de l'outil et du retour de l'enseignant, ainsi que la complémentarité des approches.

Autour du français académique, Rui Yan, Agnès Tutin et Thi Thu Hoai Tran, des universités de Grenoble et d'Arras, étudient les routines verbales dans l'écrit académique et leur didactisation dans le cadre de l'enseignement du Français Langue Etrangère. L'analyse d'un corpus d'apprenants a révélé les erreurs les plus fréquentes des étudiants à propos des routines verbales. Un corpus d'experts en français académique a permis de comparer et de mettre en exergue les routines à enseigner. Leur article se termine sur des exemples d'exercices de réflexion pour les apprenants à partir d'extraits du corpus d'experts et montre clairement en quoi ces routines sont structurantes pour l'écrit académique.

Thi Thu Hoai Tran de l'Université d'Arras et Achille Falaise de l'Université Paris Diderot, s'intéressent au Français Langue Étrangère et tandis que l'un (A. Falaise) développe un outil informatique, proche du dictionnaire, pour une aide à la rédaction universitaire à partir d'un corpus numérique, l'autre (T. T. H. Tran) expérimente l'utilisation de cet outil auprès d'étudiants internationaux en FLE. Ces expérimentations ont permis une véritable réflexion didactique quant à l'amélioration de l'outil. Ceci est un exemple de collaboration fructueuse entre didacticien-linguiste et taliste<sup>5</sup>.

Autour de l'allemand académique, Marina Kogan, Anna Yaroshevich et Olga Ni, toutes trois de l'université Pierre le Grand de St. Pétersbourg, enseignent l'allemand à de futurs ingénieurs russes. Pour ce faire, les auteures élaborent des activités pédagogiques à partir d'un corpus numérique, en fonction des spécialités de leurs apprenants en s'arrêtant sur les séquences lexicales les plus fréquentes et en tentant une approche de consultation directe du corpus par les apprenants.

Le dossier thématique est clos par la contribution d'Eva Schaeffer-Lacroix, de l'ÉSPÉ<sup>6</sup> de Paris, qui enseigne l'allemand à de futurs enseignants de cette même langue. Elle présente une entrée pédagogique dans laquelle les étudiant.e.s, confrontés à un corpus numérique, doivent développer une réflexion leur permettant de comprendre l'utilisation de la virgule en allemand. Elle nous invite alors à réfléchir aux apports des approches inductives et déductives dans un tel contexte.

Nous remercions les auteur.e.s et les évaluateurs/trices d'avoir aimablement contribué à l'élaboration de ce numéro. Nous souhaitons que les recherches sur les corpus numériques pour

---

<sup>5</sup> TAL : Traitement automatique des langues

<sup>6</sup> École Supérieure du Professorat et de l'Éducation.

l'enseignement des langues académiques étrangères soient une entrée prometteuse pour l'enseignement des langues en général à l'aide des corpus numériques.

## THE TEACHING AND LEARNING OF ACADEMIC WRITING WITH THE AID OF NUMERIC CORPORA

Coordinators: **Cristelle Cavalla** (Université Sorbonne Nouvelle-Paris3, *DILTEC*)  
**Laura Hartwell** (Université Grenoble Alpes, *LIDILEM*)

This issue 58 of *Lidil*, following in the tradition of digital corpus studies (cf. for example: Fløttum, Dahl & Kinn, 2006; Hunston & Francis, 2000; Swales, 1990/2004; Tutin & Grossmann, 2014 ; Quirk, Greenbaum, Leech & Svartvik, 1985), centers upon the teaching and learning of foreign languages. We adopt the position here, from a linguistic perspective, that a digital corpus is a collection of written or recorded documents assembled for the purpose of analyzing their language content with the help of computer tools. These corpora serve directly or indirectly for teaching and learning through multiple modalities. In learning environments, corpora can take various forms, from targeted texts to large databases. We also find language resources derived from corpora such as vocabulary lists (Coxhead, 2000), dictionaries (Collins, 2017; MacMillan, 2017), teaching materials (Chambers, 2010). The actual advantages of using digital corpuses for didactic purposes and the difficulties therein encountered or imagined remain a vast field of study.

In this issue, we focus on teaching and learning academic writing using digital corpora. The authors thus explain the interest of using such resources to support writing academic texts in a foreign language. Subject of numerous linguistic and didactic studies in the French-speaking world since the 1990s (cf. *Lidil* 17, *Pratiques de l'écrit et modes d'accès au savoir dans l'enseignement supérieur*, coordinated by Michel Dabène and Yves Reuter). Today, we find corpora freely available on-line, such as the English BAWE<sup>7</sup> or Michigan Corpora<sup>8</sup> and, in both French and English, the Scientext<sup>9</sup> corpora available via the ScienQuest<sup>10</sup> interface, allowing researchers to identify the characteristics of academic discourse of both experts and students.

Without repeating the benefits of large digital corpora for linguistic description (cf. *Lidil* 31, *Corpus oraux et diversité des approches*, coordinated by Marie Savelli), this publication is an opportunity to discover new ways to support scholarly writing through the use of digital corpora. In didactic terms, the articles gathered here deal with the direct or indirect use of digital

---

<sup>7</sup> BAWE : <http://ota.ahds.ac.uk/headers/2539.xml>, see also : [www.coventry.ac.uk/bawe](http://www.coventry.ac.uk/bawe).

<sup>8</sup> Michigan Corpora : <https://lsa.umich.edu/eli/language-resources/micase-micusp.html>.

<sup>9</sup> Scientext : <http://scientext.msh-alpes.fr/scientext-site/spip.php?article1>.

<sup>10</sup> ScienQuest : <http://corpora.aiakide.net/scientext18/>.

corpora by both teachers and learners in a university context. In recent years, the development of greater storage and especially computer tools allowing online query options accessible by non-specialists has enabled the development of digital corpora for educational purposes. Indeed, the articles in this issue explore the use of these digital resources for the teaching and learning of academic language and, in some cases, scientific phraseology.

Following Boulton and Tyne's book (2014) on the use of corpora in language classes, these contributions provide answers about the actual use of digital corpora by university teaching teams: How do they use them and why? We will see that linguistic and didactic approaches echo from one language to another and that language didactics federate these different studies.

Concerning academic English, Clive Hamilton and Natalie Kübler, who teach at the Diderot University in Paris, are developing courses, based on corpora of scientific English, for French-speaking students who specialize in different disciplines. The case of students directly using online corpora remains uncommon in French universities and this case study helps to better understand the issues.

Benet Vincent and Hilary Nesi, from Coventry University, teach academic English to international students using a BAWE-based interface to introduce hypertext links designed to guide and assist learners in their academic writing. The question of the insertion of such tools in language classes is raised, making it possible to highlight the specificities of the tool and the teacher's feedback, as well as the complementarity qualities of these approaches.

Concerning academic French, Rui Yan, Agnès Tutin, and Thi Thu Hoai Tran, from Grenoble Alpes and Arras Universities, are involved in the teaching of French as a foreign language and more particularly in identifying the verbal routines of French academic writing. Their study of a learner corpus reveals the students' most common errors related to verbal routines. A corpus of academic French texts by experts made it possible to compare and highlight the relevant routines to be taught. Their article concludes with examples of reflective exercises for learners based on extracts from the expert corpus, which clearly show how these routines structure academic French.

Thi Thu Hoai Tran, from the Université d'Arras, and Achille Falaise, from the Université Paris Diderot, are interested in French as a Foreign Language. While one (A. Falaise) is developing a dictionary-like, corpus-based computer tool with the objective of helping university students with their writing, the other (T. T. H. Tran) is introducing the tool to international students learning French. These experiments have allowed an in-depth didactic reflection on improving the tool. This is an example of a successful collaboration between an educator-linguist and a natural language processor.

Marina Kogan, Anna Yaroshevich, and Olga Ni, from the Peter the Great University of St. Petersburg, teach German to future Russian engineers. By focusing on the most frequent lexical sequences and by attempting an approach by direct learner consultation of the corpus, the authors develop activities based on a digital corpus to support teaching and learning in accordance with the learners' specialties.

This thematic issue concludes with the contribution of Eva Schaeffer-Lacroix, who teaches German at the ÉSPÉ<sup>11</sup> de Paris to future German teachers. Here, she presents students' reactions to introductory consultations of a digital corpus, tracing their reflection on the practice and understanding of the use of the comma in German. She then invites readers to think about the contributions of inductive and deductive approaches in such a context.

We thank the authors and evaluators for their gracious contributions to this issue. We hope that the research on digital corpora for teaching foreign languages in university settings will be a promising entry point for the general use of digital corpora within language teaching.

#### **BIBLIOGRAPHIE/BIBLIOGRAPHY**

- BOULTON, Alex & TYNE, Henry. (2014). *Des documents authentiques aux corpus : Démarches pour l'apprentissage des langues*. Paris : Les Éditions Didier.
- CHAMBERS, Angela (2010). L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé. *Revue française de linguistique appliquée*, 15(2), 9-20.
- COLLINS FREE ON-LINE DICTIONARY. (2017). Disponible en ligne sur : <https://www.collinsdictionary.com/dictionary/english>.
- COXHEAD, Averil. (2000). A New academic word list. *TESOL Quarterly*, 34(2), 213-238.
- DABÈNE, Michel & REUTER, Yves (Dir.). (1998). *Lidil, 17, Pratiques, de l'écrit et modes d'accès au savoir dans l'enseignement supérieur*.
- FLØTTUM, Kjersti, DAHL, Trine & KINN, Torodd. (2006). *Academic voices*. Amsterdam: John Benjamins Publishing Company.
- HUNSTON, Susan & FRANCIS, Gill. (2000). *Pattern grammar: A Corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins Publishing Company.
- HYLAND Kenneth & BONDI, Marina. (Dir.) (2006). *Academic discourse across disciplines*. Bern: Peter Lang.
- MACMILLAN FREE ON-LINE DICTIONARY. (2017). Disponible en ligne sur : <http://www.macmillandictionary.com/>.
- SINCLAIR, John. (1991). *Corpus, Concordance, Collocation: Describing English language*. Oxford : Oxford University Press.
- QUIRK, Randolph, GREENBAUM, Sidney, LEECH, Geoffrey & SVARTVIK, Jan. (1985). *A comprehensive grammar of the English language*. London: Longman Publishing Company.
- SAVELLI, Marie (Dir.). (2005). *Corpus oraux et diversité des approches*. *Lidil*, 31. Disponible en ligne sur : <https://journals.openedition.org/lidil/125>.

---

<sup>11</sup> École Supérieure du Professorat et de l'Éducation.



- SWALES, John M. (1990/2004). *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- TUTIN, Agnès & GROSSMANN, Francis (Dir.). (2014). *L'écrit scientifique : du lexique au discours*. Rennes: Presses Universitaires de Rennes.