

CORPUS D'APPRENANT-ES : UN « BASQUETTE » DE RESSOURCES POUR L'ENSEIGNEMENT DES LANGUES ÉTRANGÈRES

LEARNERS' CORPORA: A "BASQUETTE" OF RESSOURCES FOR FOREIGN LANGUAGE TEACHING

● Karine Lichtenauer
| Université de Genève
Anita Thomas
| Université de Fribourg

Quel meilleur moyen que de regarder les réalisations des élèves pour comprendre la manière dont leurs apprentissages progressent ?

La linguistique de corpus appliquée aux corpus d'apprenant-es de langue peut être vue comme une méthode de recherche empirique proche de l'expérience des enseignant-es de langue: écouter, lire, observer ce que font les élèves pour mieux comprendre leurs cheminements et les écarts de leurs productions d'avec les usages les plus courants. D'autant plus que le travail sur les corpus permet de renouveler la description de la langue et de sortir des carcans normatifs: L'étude de corpus présentant la langue telle qu'elle est utilisée par ses locuteurs permet de relativiser les normes présentées dans les grammaires et donne une place légitime à la langue parlée. Ainsi, les corpus permettent de documenter de manière concrète l'usage de la langue que ce soit comme L1 ou L2 et donner un apport spécifique à la didactique des langues.

What better way to understand how students learn a foreign language than to look at what they actually do with it?

When applied to corpora of language learners, corpus linguistics can be seen as an empirical research method that shows numerous similarities to the experience of language teachers: listening to students, reading and observing what they do in order to analyse their development and the gaps between their language and more common formulations. Indeed, working with corpora lets us revisit the description of language and break free from normative constraints: investigating corpora presenting the language as used by its speakers questions the norms presented in grammars and gives a legitimate place to spoken language. Corpora provides thus concrete documentation of language use, be it L1 or L2, and makes a specific contribution to language teaching.

Landure & Boulton définissent un *corpus linguistique* comme un « grand ensemble de textes authentiques représentatif d'une variété donnée quelle qu'elle soit et exploitable sous forme électronique » (Landure & Boulton 2010 : 57, selon la définition de McEnery et al. 2006 :5). Dans ce numéro de *Babylonia* toutefois, les textes ne sont pas représentatifs d'une « variété donnée » qui serait partagée par une communauté linguistique, mais plutôt d'interlangues, individuelles, non systématiques et en évolution constante. Les corpus d'apprenant-es dont il sera question dans ce numéro reflètent ainsi la progression des apprentissages, ils fournissent des indices sur les compétences langagières et communicatives des élèves, mais aussi sur leurs attitudes et comportements, sur l'efficacité de dispositifs didactiques et bien plus encore. Les articles publiés ici montrent bien le foisonnement des intérêts de recherche, des méthodes de recueil des données, des fonctions et des applications didactiques de chaque corpus.

Nous aurions pu procéder à une catégorisation des articles par types de corpus selon le médium (écrit, oral – avec ou sans vidéos), le genre de texte (interactions enseignant-e/élèves ou entre élèves, production écrite individuelle, production orale de discours scientifique, etc.), la langue (corpus monolingues, bilingues, multilingues) ou encore la taille (grand corpus, ensemble de plusieurs corpus, corpus restreints pour des questions de recherche particulières). A une telle catégorisation de corpus, une perspective plus orientée vers les types d'applications des résultats des recherches – vers l'enseignement des langues – nous a semblé favorable.

Le numéro s'ouvre sur deux articles qui montrent les enjeux de la recherche sur des corpus d'apprenant-s et les domaines d'applications des résultats. Tout d'abord, **Gaëtanelle Guilquin** présente des **projets de recherche actuels basés sur des corpus d'apprenant-es** et l'évolution du champ au cours des dernières années, notamment en termes d'applications didactiques concrètes. Cet état des lieux éclaire ainsi les bénéfices que les enseignant-es de langue, les concepteurs-trices de matériel pédagogique et autres acteurs-actrices peuvent tirer des recherches sur le sujet mais aussi des corpus existants pour mieux répondre aux besoins indi-

Dans ce numéro de *Babylonia* toutefois, les textes ne sont pas représentatifs d'une « variété donnée » qui serait partagée par une communauté linguistique, mais plutôt d'interlangues, individuelles, non systématiques et en évolution constante. Les corpus d'apprenant-es dont il sera question dans ce numéro reflètent ainsi la progression des apprentissages, ils fournissent des indices sur les compétences langagières et communicatives des élèves, mais aussi sur leurs attitudes et comportements, sur l'efficacité de dispositifs didactiques et bien plus encore.

Landure & Boulton define a linguistic corpus as 'a large set of authentic texts which are representative of any given language variety and exploitable in electronic form' (Landure & Boulton 2010 : 57¹, following the definition of McEnery et al. 2006 : 5). In this issue of *Babylonia*, however, the data is not representative of a 'given variety' shared by a linguistic community, but rather of individual, non-systematic and constantly evolving interlanguages. Thus, learner corpora discussed in this issue reflect the progression of learning. But not only does it provide clues about learner interlanguage and communication skills, it also informs us about attitudes and learning practices, about the effectiveness of teaching methods, and much more! The various articles illustrate the wide range of research interests, data collection methods, as well as the purpose of each corpus and the possible uses in language teaching.

We could have categorised the contributions by type of corpus according to the medium (written, oral – with or without videos), type of text (interaction between learners or with the teacher, individual written production, oral production of scientific discourse, etc.), language (monolingual, bilingual, multilingual corpora) or size (large corpus, set of several corpora, restricted corpora for particular



Anita Thomas est professeure en Français Langue Étrangère à l'Université de Fribourg et actuellement directrice de l'Institut de plurilinguisme. Ses recherches et son enseignement portent sur le rôle de l'input et de son traitement en classe de langue ainsi que chez les apprenant-e-s L2. Elle s'intéresse depuis près de 20 ans à l'utilisation des corpus comme ressource pour l'enseignement et la recherche.



Karine Lichtenauer est chargée d'enseignement en didactique de l'allemand à l'IUFE de l'Université de Genève.

¹ Translated by the authors.

Explorer des corpus dont la simple présentation aiguise les envies de recherche.

viduels des apprenant-es. Les bénéfices ne vont toutefois pas sans défis à relever, comme le confirment **Katrin Wisniewski & Matthias Schwendemann** dans l'interview qu'ils nous ont accordée. Ils partagent leur expérience du projet **DAKODA (Datenkompetenzen in DaF/DaZ: Exploration sprachtechnologischer Ansätze zur Analyse von L2-Erwerbsstufen in Lernerkorpora des Deutschen)** qui vise à intégrer et rendre accessible un grand nombre de corpus d'apprenant-es d'allemand langue étrangère ou seconde dans un référentiel. Spécificité des corpus en fonction des questions de recherche, réflexions sur les métadonnées et le travail d'annotations, exigences éthiques du respect de l'anonymat, possibilités d'utilisation ou perspectives pour la recherche, l'interview nous fait ainsi entrer dans le détail du travail de corpus avant de donner un ensemble de conseils à celles et ceux qui souhaitent se lancer dans l'aventure des corpus d'apprenant-es.

La deuxième partie de ce numéro est consacrée aux facteurs qui influencent le développement des compétences linguistiques des apprenant-es de langue. **Thomas Studer & Nina Hicks** nous présentent le corpus suisse des apprenant-es **SWIKO**, un corpus multilingue de jeunes en langue de scolarisation et en langue étrangère, ainsi que deux utilisations concrètes : l'une pour mieux comprendre les corrélations entre types de tâche et réalisations des élèves ; l'autre pour une exploitation didactique de **SWIKO** en classe de langue pour soutenir l'apprentissage de la négation en allemand langue étrangère. En milieu universitaire, cette fois, **Matthias Schwendemann & Franziska Wallner** étudient la réalisation des contractions phonologiques en allemand L2 dans les discours académiques oraux produits par des étudiant-es à l'aide du corpus **GeWiss (Gesprochene Wissenschaftssprache kontrastiv)**. Outre les comparaisons avec un corpus de productions académiques produites par des locuteurs-trices dont l'allemand est la langue première, l'analyse des productions d'apprenant-es montre des différences entre locuteur-trices qui étudient en Allemagne et à l'étranger.

Gwendoline Lovey montre comment l'étude d'un **corpus audio d'interactions en classe** entre quatre enseignantes à l'école primaire et leurs élèves respectifs peuvent donner des indications sur

research questions). We preferred, however, to concentrate on the types of applications of the scientific results – to concentrate on language teaching.

The issue opens with two articles providing deep insights into research with learner corpora and showing how the results can be useful to language teaching in a large sense. **Gaëtanelle Guilquin** presents **current research projects on learner corpora** as well as the evolution of the field in recent years with an emphasis on practical applications. This overview sheds light on the benefits to language teachers, designers of teaching materials and other stakeholders from this research and from existing corpora so to better meet the individual needs of learners. However, the benefits are not without their challenges, as **Katrin Wisniewski & Matthias Schwendemann** confirm in their interview. They share their experience with the **DAKODA project (Datenkompetenzen in DaF/DaZ: Exploration sprachtechnologischer Ansätze zur Analyse von L2-Erwerbsstufen in Lernerkorpora des Deutschen)**: the creation of an open-access repository collecting a large amount of data from learners of German.

They explore characteristics of corpora according to the type of research question, metadata and annotation, issues on anonymisation requirements and the possible ways of exploiting the repository in research or in teaching. They include a series of tips for those wishing to embark on the adventure of learner corpora.

In the second part of this issue, we turn to the factors having an impact on the improvement of learner language skills. **Thomas Studer & Nina Hicks** present the **Swiss learner corpus SWIKO**, a multilingual young learner corpus in both the language of schooling and a foreign language. Using two examples, they illustrate how the corpus can be used: first to investigate the correlations between types of task and learners' achievements; and then to use **SWIKO** directly in the language classroom when teaching, here for negation in German as a foreign language lessons.

In German-speaking academic settings, **Matthias Schwendemann and Franziska Wallner** focus on phonological contractions in oral academic discourse

l'efficacité des pratiques enseignantes. Elle en dégage des pratiques d'introduction et d'accompagnement des activités propres à encourager les élèves à formuler des productions plus riches. **Rosamond Mitchell & Florence Myles** présentent des résultats de recherche sur la base du corpus d'interaction en classe de langue **Learning French** composé d'enregistrements audio et vidéo de 33 leçons de français langue étrangère à l'école primaire en Grande-Bretagne. Leurs analyses visent, d'une part, le rôle de la fréquence des mots dans l'input des enseignant-es et dans la langue française et, d'autre part, l'influence du niveau d'engagement perçu des élèves sur l'apprentissage de nouveaux mots.

L'interlangue des élèves et le développement de leurs compétences langagières seront au centre de la troisième partie, qui illustre bien la variété des phénomènes linguistiques étudiés que les corpus d'apprenant-es révèlent. En effet, les quatre contributions qui la composent visent à mieux comprendre le développement des compétences lexicales, pragmatiques ou phonologiques des élèves – en explorant des corpus dont la simple présentation aiguise les envies de recherche.

Isabelle Racine se penche tant sur la réception que sur la production de caractéristiques phonologiques du français oral sur la base des corpus **PFC (Phonologie du Français Contemporain)** et **IPFC (Inter-Phonologie du Français Contemporain)**. La pragmatique est ensuite à l'honneur avec le corpus **DiCoi (digitalisation – corpus – interaction)** à partir duquel **Anita Thomas & France Rousset** étudient l'apprentissage, réceptif et productif, de marqueurs discursifs en français. Leur étude longitudinale du développement des compétences des élèves comprend une intervention didactique qui engage les apprenant-es à observer des usages spontanés en L1.

Les articles suivants portent notre attention sur les choix lexicaux des élèves, mais par des méthodes d'investigation très différentes. Avec des méthodes d'élicitations issues de la psycholinguistique et des analyses des champs conceptuels fondées sur la sémantique cognitive, **Mireille Copin & Inès Saddour** proposent de travailler sur des **corpus de type expérimental** pour des phénomènes sémantiques complexes. De tels corpus

produced by students with German as a Foreign or Second language using the **GeWiss corpus (Gesprochene Wissenschaftssprache kontrastiv)**. In addition to comparisons with a corpus of academic productions produced by speakers of German as a first language, the analysis of learner productions shows differences between speakers studying in Germany or abroad.

Gwendoline Lovey analyses an **audio corpus of classroom interactions** between four primary school teachers and their respective learners to find out how efficient different teaching practices are. She deduces from this how to introduce or scaffold foreign language activities in order to help learners produce more complex language.

Rosamond Mitchell & Florence Myles share research results also based on a classroom interaction corpus: **Learning French** collects 33 lessons of French filmed in English primary schools. They focus on the role of word frequency in teacher input and in the French language and investigate how learners' perceived level of commitment affects their learning of new words.

Exploring corpora whose mere presentation whets the appetite for research

The third part of this issue deals with learners' interlanguage and the development of their language skills, and thus illustrates the variety of linguistic phenomena revealed by learner corpora. The four contributions seek a better understanding of the development of learner lexical, pragmatic or phonological skills by exploring corpora whose mere presentation whets the appetite for research.

Qu'il s'agisse du développement des compétences grammaticales, lexicales, pragmatiques ou phonologiques des élèves ou encore des facteurs-clés ayant un impact sur l'ampleur et la qualité de leurs réalisations linguistiques, les auteur-es de ce numéro montrent ce que les corpus d'élèves leur ont dévoilé et les implications directes de ces connaissances pour une amélioration des pratiques enseignantes ou pour la rédaction de supports didactiques.

permettent d'observer les différences et similarités dans les choix lexicaux des apprenant-es de français langue étrangère et de francophones, ce qui aide les enseignant-e-s à mieux comprendre comment enseigner (et corriger) le lexique. **Christina Lindqvist** interroge un **grand corpus de productions écrites en FLE par des élèves plurilingues du secondaire en Suède**. L'analyse détaillée des transferts lexicaux, et, plus généralement, des choix lexicaux des élèves, font émerger une connaissance en profondeur de leurs compétences lexicales plurilingues. L'article se termine sur des pistes didactiques pour travailler sur un tel corpus dans une approche plurilingue.

Qu'il s'agisse du développement des compétences grammaticales, lexicales, pragmatiques ou phonologiques des élèves ou encore des facteurs-clés ayant un impact sur l'ampleur et la qualité de leurs réalisations linguistiques, les auteur-es de ce numéro montrent ce que les corpus d'élèves leur ont dévoilé et les implications directes de ces connaissances pour une amélioration des pratiques enseignantes ou pour la rédaction de supports didactiques. Les exemples d'utilisation des corpus en classe, avec le matériel didactique, illustrent par ailleurs la variété des recherches actuelles sur le *Data-Driven-Learning*.

Nous vous souhaitons une lecture inspirante!

Isabelle Racine looks at both the reception and production of phonological features in spoken French, using the **PFC (Phonologie du Français Contemporain)** and **IPFC (InterPhonologie du Français Contemporain)** corpora. Pragmatics then takes centre stage in the contribution by **Anita Thomas & France Rousset**. They investigate the receptive and productive learning of discourse markers in French with the corpus **DiCoi (digitalisation - corpus - interaction)**. Their longitudinal study of the development of students' skills includes an intervention that engages learners in observing spontaneous uses in L1.

The next articles both focus on learners' lexical choices, but with very different methods of investigation. With a corpus elicited with psycholinguistic methods and analysed along cognitive semantic lines, **Mireille Copin & Inès Saddour** demonstrate how to work on experimental types of corpora to address complex semantic phenomena. They observe differences and similarities in the lexical choices of students of French as a Foreign Language on one hand, and Francophones on the other in order to understand how to teach (and correct) these phenomena.

Christina Lindqvist examines a **large corpus of written production by plurilingual secondary school learners of French as a Foreign Language in Sweden**. The detailed analysis of lexical transfers and, more generally, of the pupils' lexical choices, reveals in-depth knowledge of their plurilingual lexical skills. The article concludes with teaching suggestions for working with such a corpus in a plurilingual approach.

Whether they study the development of learner grammatical, lexical, pragmatic, or phonological skills or the key-factors impacting on quantity and quality of learner linguistic achievements, the authors of this issue show us what learner corpora have revealed to them as well as practical uses of this knowledge for improving teaching practices or creating teaching materials.

The examples of the use of corpora in the classroom, along with teaching materials, also illustrate the variety of current research on *Data-Driven-Learning*.

We wish you inspiring reading!

Whether they study the development of learner grammatical, lexical, pragmatic, or phonological skills or the key-factors impacting on quantity and quality of learner linguistic achievements, the authors of this issue show us what learner corpora have revealed to them as well as practical uses of this knowledge for improving teaching practices or creating teaching materials.



Landure, Corinne & Alex Boulton (2010). « Corpus et autocorrection pour l'apprentissage des langues », *ASP [En ligne]*, 57 | 2010, consulté le 5 octobre 2022. URL : <http://journals.openedition.org/asp/931>

McEnery, T., R. Xiao et Y. Tono (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. Londres : Routledge.