

Exploitation d'un grand corpus d'écrits scolaires pour la recherche et la formation : introduction

Marie-Laure Elalouf

CY Cergy Paris Université ÉMA - ÉA 4507

Il y a une quinzaine d'années, nous observions l'avènement de thèses en didactique du français qui avaient pour point commun de s'appuyer sur des annexes organisées en corpus conséquents, susceptibles d'être lus et exploités de façon quasi autonome, tout en étant fortement articulées à la thèse par des choix théoriques et méthodologiques explicites (Elalouf, 2011). Les exigences ayant présidé à l'élaboration de ces corpus devaient permettre leur extension ou leur exploitation pour d'autres problématiques : mise à disposition de l'ensemble des éléments permettant de refaire les analyses, explicitation des modalités de recueil et de la place du chercheur, rigueur méthodologique dans la comparaison des sous-corpus afin de dégager des critères de comparabilité avec d'autres corpus, mise au jour d'éléments de typicité dépassant la singularité des recueils tout en la respectant.

Depuis, d'autres thèses ont apporté à la communauté des chercheurs en didactique un matériau riche et rigoureux, mais qui faute de mise en réseau ne contribue pas à sa juste mesure à la cumulativité des résultats de recherche. Par ailleurs, s'il est possible de susciter en master *Formation de formateurs* des corpus qui enrichissent des recherches existantes (op. cit., 37), ils restent souvent confinés dans des mémoires non publiés et leur exploitation en formation se trouve ainsi limitée à leurs auteurs.

En concevant le *Corpus-ÉMA-écrits scolaires*¹ comme une ressource à disposition des chercheurs et des formateurs, le choix a été fait de réunir, selon les mêmes principes, l'ensemble des productions d'écrit de différentes classes de tous niveaux scolaires avec leur contexte d'élaboration, en suivant des conventions de transcription, d'annotation et de traitement qui les rendent compatibles avec d'autres grands corpus (Doquet, David et Fleury, 2017), notamment Ecriscol.

Ce corpus s'inscrit dans une réflexion sur les genres scolaires (Boré, 2007 ; Boré *et al.*, 2007, Boré & Elalouf, 2017, Roubaud, 2017) et sur l'annotation (Boré & Bosredon, 2018 ; Besnard & Elalouf, 2018). Il a déjà donné lieu à des exploitations par leurs auteurs (Castagnet-Caignec, 2017, 2018, 2021 ; Elalouf & Koulibali, 2020 ; Elalouf & Gerlaud, 2021 ; Similowski, 2022 ; Similowski, Genre, Koulibali, Elalouf, 2022 ; Koulibali, 2023 ; Gerlaud & Elalouf, 2023). Avec ce numéro, nous avons souhaité

¹ <https://www.ortolang.fr/market/corpora/ema-ecrits-scolaires-1>, Cf. Annexe.

franchir une nouvelle étape dans son exploitation par des personnes ne l'ayant pas produit.

Pour favoriser l'appropriation du grand corpus, un article a été rédigé par l'équipe initiatrice, qui en présente les principes et décrit chacun des sous-corpus avec des renvois aux fichiers correspondants sur ORTOLANG : « *Corpus ÉMA, écrits scolaires, une ressource à disposition sur ORTOLANG* ». C'est le premier de ce numéro. Pour chacun des sous-corpus, des pistes d'exploitation pour la recherche ou la formation y sont suggérées, qui ne sont pas exclusives.

Les contributions s'organisent autour de quatre axes :

- une recherche didactique et/ou linguistique portant sur un ou plusieurs sous-corpus de Corpus ÉMA-écrits scolaires ;
- une recherche didactique et/ou linguistique mettant en relation un ou plusieurs sous-corpus de Corpus ÉMA-écrits scolaires avec d'autres corpus existants ;
- une recherche didactique et/ou linguistique portant sur une utilisation en formation d'un ou plusieurs sous-corpus de Corpus ÉMA-écrits scolaires ;
- un outil à destination de formateurs portant sur la description et l'analyse d'un dispositif d'écriture expérimenté lors de la constitution d'un sous-corpus de Corpus ÉMA-écrits scolaires.

1. Recherches portant sur un ou plusieurs sous-corpus de Corpus ÉMA-écrits scolaires

Dans cette section, deux articles s'ancrent en didactique de la littérature. Daphné Jacamon a retenu les brouillons qui accompagnent les travaux de lycéens ayant choisi la spécialité *Humanités, littérature et philosophie*, épreuve nouvelle au moment du recueil. Elle propose une analyse qualitative des traces de réflexivité observables sur les brouillons, en se demandant si les modalités d'évaluation, qui consistent en une question d'interprétation, favorisent les compétences de lecture littéraire.

Sonia Castagnet-Caignec, dont le corpus de thèse a été intégré et annoté dans le Corpus ÉMA-Écrits scolaires, expose la conception d'un recueil du CE1 à la première littéraire, permettant d'analyser à partir des mêmes supports filmiques l'évolution des compétences scripturales des élèves dans un genre peu enseigné : la novélisation. Elle mobilise la linguistique textuelle, l'énonciation et la narratologie pour dégager des profils allant de l'étroite fidélité au support filmique à sa re-création.

Deux autres articles proposent un traitement linguistique et didactique d'un fait de langue, tel qu'il s'actualise dans plusieurs sous-corpus. L'article « Vers une étude de la coordination dans un grand corpus d'écrits scolaires » part du constat que la seule mémorisation d'une liste de conjonctions de coordination est de peu de secours pour résoudre les problèmes complexes auxquels sont confrontés les élèves quand ils écrivent. Une alternative est recherchée en resserrant l'étude sur le « noyau dur » des

conjonctions qui satisfont à tous les critères définitoires et leurs usages dans différents genres textuels.

L'article proposé par Fatma Ben Barka Messaoudi, Yelle Koulibali et Marie-Laure Elalouf travaille la même tension à propos d'une autre notion, le subjonctif. Une analyse linguistique de ce mode permet de pointer les difficultés que suscite sa conceptualisation, dont certaines sont accentuées par la description scolaire. L'étude des usages du subjonctif dans des sous-corpus de cycle 4 conduit à des propositions didactiques visant une littératie grammaticale.

2. Recherches mettant en relation un ou plusieurs sous-corpus de Corpus ÉMA-écrits scolaires avec d'autres corpus existants

Agnès Furman a choisi d'augmenter le corpus qu'elle a recueilli dans le cadre de sa thèse – des textes produits dans trois classes de 6^e de milieux sociaux différenciés – en lui adjoignant deux sous-corpus de Corpus ÉMA, également contrastés sur le plan social, pour étudier les problèmes posés par l'homophonie verbale et la segmentation des unités dans la zone préverbale. La convergence des résultats permet d'identifier des zones de vulnérabilité.

Bernadette Kervyn, Véronique Magnant, Marie-Noëlle Roubaud et Claire Doquet ont mis en commun leurs différents corpus pour traiter une question peu étudiée : les relations qu'entretiennent le dessin et le texte dans la production de scripteurs novices au début de l'école élémentaire (6-8 ans). Une typologie de ces relations dans l'espace de la feuille et le temps de l'écriture montre la richesse de cette coprésence et sa contribution aux compétences littéraciques.

Si les interactions entre dessin et écriture mettent en lumière l'apport de la pluralité sémiotique à la compréhension des apprentissages, celles entre différents médiums, l'oral et l'écrit, méritent aussi attention. Marie-Laure Elalouf et Yelle Koulibali étudient l'oral accompagnant la production d'un écrit collaboratif en cherchant comment adapter et compléter les catégories d'analyse issues de l'analyse des rédactions conversationnelles d'adultes et proposent des observables pour l'analyse textométrique de tels corpus.

3. Recherche portant sur une utilisation en formation d'un ou plusieurs sous-corpus de Corpus ÉMA-écrits scolaires

L'un des sous-corpus comporte à la fois les textes d'élèves que Kathy Similowski a recueillis dans sa thèse et les traces que des enseignants ont porté sur ces textes dans le cadre d'une formation continue qu'elle a conduite et analysée. En confrontant ces dernières aux procédés syntaxiques utilisés par les élèves pour épaisser leur texte, Marie-Noëlle Roubaud pointe un hiatus et avance des propositions pour sensibiliser les enseignants aux choix linguistiques des élèves.

4. Outil à destination de formateurs portant sur la description et l'analyse d'un dispositif d'écriture expérimenté dans Corpus ÉMA-écrits scolaires

La singularité du dernier article tient au fait qu'il est issu d'un atelier d'écriture qui a conclu une recherche collaborative dans laquelle a été conçu un protocole d'accompagnement à l'autonomie scripturale. Karine Bensadia, qui l'a expérimenté dans ses classes, revient sur ses choix et ses ajustements, en dialogue avec les analyses de productions d'élèves menées collectivement dans le cadre du séminaire animé par Marie-Laure Elalouf et Eleni Valma.

Avec ce numéro de Scolagram, nous souhaitons franchir une nouvelle étape dans l'appropriation de ce grand corpus par des personnes ne l'ayant pas produit. Sa mise en relation avec d'autres corpus existants ouvre la voie à une montée en généralité, en poursuivant la réflexion méthodologique.

Sachant combien les pratiques résistent à s'appuyer sur les productions d'élèves pour observer les processus en jeu et faire observer la langue, malgré les ressources que la recherche en didactique a déjà constituées (Boré, 2016 ; Cappeau et Roubaud, 2005, 2018 ; Garcia-Debanc, Roubaud, Béchour, 2022), nous espérons que cette publication contribuera à la circulation des ressources, des outils et des analyses entre recherche et formation.

Bibliographie

- Besnard, M. & Elalouf, M.-L. (2018). (Re)apprendre à lire des textes de jeunes scripteurs. *Le français aujourd'hui*, 203, 75-86.
- Boré, C. (dir.). (2016). La phrase en production d'écrits, approches nouvelles en didactique. *Lidil*, 54, [En ligne].
- Boré, C. (2007, dir.). *Construire et exploiter des corpus de genres scolaires*. Namur (Belgique), Presses Universitaires de Namur, collection Diptyque, numéro 10.
- Boré, C. (2007). La métamorphose d'un genre : quelques descripteurs pour un genre scolaire de récit. Dans *Construire et exploiter des corpus de genres scolaires*. Presses Universitaires de Namur, collection Diptyque, numéro 10, pp. 141-165.
- Boré, C. & Elalouf, M.-L. (2017). Deux étapes dans la construction de corpus scolaires : problèmes récurrents et perspectives nouvelles. *Corpus* 16, 31-63.
- Boré C. & Bosredon, C. (2018). « Discours enseignant dans des écrits d'élèves d'école élémentaire : enquête sur le rôle des annotations ». *Le Français Aujourd'hui*, 203 : 99-112.

Cappeau, P., & Roubaud, M. N. (2005). *Enseigner les outils de la langue avec les productions d'élèves*. Paris, Bordas.

Cappeau, P., & Roubaud, M. N. (2018). *Regards linguistiques sur les textes d'élèves (de 5 à 12 ans)*. Presses Universitaires Blaise Pascal.

Castagnet-Caignec, S. (2017). La novélisation : une rencontre avec du déjà-là qui révèle les positionnements auctoriaux des jeunes scripteurs ». *Pratiques*, 173-174, [En ligne].

Castagnet-Caignec, S. (2018). Traitement du temps dans des récits à visée littéraire chez les élèves de primaire et secondaire. *Repères*, 57, 35-56.

Castagnet-Caignec, S. (2021). Pour des corpus transfictionnels à l'école : en quoi un récit multimédiatique peut-il servir l'enseignement ? ». *Strenæ*, 19, [En ligne].

Doquet, C., David, J. & Fleury, S. (Eds.). (2017). Spécificités et contraintes des grands corpus scolaires : problèmes de transcription, d'annotation et de traitement. *Corpus*, 16, [En ligne].

Elalouf, M.-L. (2011). Constitution de corpus scolaires et universitaires : vers un changement d'échelle ? ». *Pratiques*, 149-150, 56-70.

Elalouf, M.-L. & Koulibali, Y. (2020). Les emplois du verbe *dire* dans un corpus d'écrits scolaires : Étude exploratoire pour un lexique-grammaire des usages. *7^e Congrès Mondial de Linguistique Française*, <https://doi.org/10.1051/shsconf/20207807006>

Elalouf, M.-L. & Gerlaud, B. (2021). La mise en grille et l'activité résumante, deux dispositifs pour appréhender la sémantisation de la langue au lycée. In Buleaubronckart, E. & Garcia-Debanc, C., *Recherches en didactique du français, L'étude du fonctionnement de la langue dans la discipline français : quelles articulations ?* (pp. 113-132) Presses universitaires de Namur.

Garcia-Debanc, C., Roubaud, M.-N. & Béchour, M. (2022). *Guide pour enseigner la grammaire pour écrire, CE2 et cycle 3*. Paris, Retz.

Gerlaud, B. & Elalouf, M.-L. (2023). Annoter au lycée, un geste professionnel partageable ? *Recherches*, 79, 65-88.

Koulibali, Y. (2023). Les constructions des verbes hyperfréquents dans des corpus d'écrits scolaires : contribution d'une étude textométrique à la connaissance des usages des verbes de parole et d'états mentaux par des collégiens de 6^e et 5^e en éducation prioritaire. Thèse en sciences du langage de CY Cergy paris Université, dirigée par Marie-Laure Elalouf.



Roubaud, M-N. (2017). Le français écrit : transcription et édition. Le cas des textes scolaires. *Corpus*, 16, 113-131.

Similowski K. (2022). L'extrait et l'écriture scolaire : comment des élèves de fin d'école primaire approchent-ils un genre à partir de la lecture d'extraits littéraires ? In A. Belhadjin & M.-F. Bishop (Eds.). *Usage des extraits en classe de français*. Namur : CEDOCEF, PUN.

Similowski, K., Genre, S. Koulibali, Y. & Elalouf, M.-L. (2022). Comment les élèves d'école primaire reconfigurent les collocations en écrivant. *8^e Congrès Mondial de Linguistique Française*, CMLF SHS Web of Conferences, 138.

Annexe

Copie d'écran de la première page : cliquer sur parcourir pour accéder aux sous-corpus.

The screenshot shows a digital interface for a corpus. At the top, there are five tabs labeled "Version 1", "Version 2", "Version 3", "Version 4", and "Version 5". Below these tabs, there is a section titled "Référence à citer" (Reference to cite) which includes the citation: "BORÉ, C. & ELALOUF M.-L. (2017) : Deux étapes dans la construction de corpus scolaires : problèmes récurrents et perspectives nouvelles. In Corpus 16, 31-63." There are also "Texte" and "BibTeX" buttons. To the right of this section is a "Contactez le producteur" (Contact the producer) button with a "Envoyer un mail" (Send email) link. Further down, there is a "Téléchargement" (Download) section with a "Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International" (License: Creative Commons Attribution - Non Commercial - Share Alike 4.0 International) and a detailed description of the license terms. At the bottom of this section are "Télécharger" (Download) and "Parcourir" (Browse) buttons.