

## **E-CALM. Analyser de grands corpus scolaires et universitaires : des questions pour la recherche et pour la formation**

### **E-CALM. Analysing large school and university corpora: questions for research and for training**

Sara Mazziotti - Université de Lorraine - INSPE de Montigny-lès-Metz, CREM

Luca Pallanti - Université de Lyon, ECP

Arnaud Moysan - Université Sorbonne Nouvelle, CLESTHIA

Clément Reversé- CERTOP - Centre Emile Durkheim, UT2J

## **L'écriture scolaire et universitaire : questionnements multiples**

Les corpus scolaires suscitent de plus en plus l'intérêt des chercheurs, linguistes et didacticiens, qui à partir de différents questionnements étudient les productions d'élèves, des écritures approchées aux rédactions de textes de genres divers. En 2018, dans le cadre du projet ANR E-CALM (Écriture scolaire et universitaire : Corpus, Analyses Linguistiques, Modélisations didactiques) dirigé par Claire Doquet (Université de Bordeaux), trois équipes de Sciences du langage (CLESTHIA, CLLE-ERSS, LIDILEM) et une équipe de Sciences de l'éducation (CIRCEFT-ESCOL) ont initié une collaboration afin de rendre visibles et lisibles en grand nombre des écrits scolaires et permettre leur exploration outillée pour agir sur l'enseignement de l'écriture.

Cette équipe de 23 chercheurs a constitué et analysé le corpus E-CALM (environ 4500 écrits disponibles), aujourd'hui repéré par l'Inspection Générale des Lettres comme une ressource importante pour les enseignants et utilisé en formation (<https://eduscol.education.fr/document/40190/download>). Les travaux menés à partir de ce corpus avaient pour objectif de décrire l'acquisition des compétences à l'écrit au cours de la scolarité et d'observer la manière dont les enseignants, par leurs interventions sur les copies, orientent l'écriture.

En juin 2022, le colloque de clôture “*Analyser de grands corpus scolaires et universitaires : des questions pour la recherche et pour la formation*” a rassemblé des linguistes et didacticiens de l’écriture autour de la question de l’analyse des écrits scolaires et universitaires et de son exploitation pour l’enseignement. Trois axes étaient travaillés : le développement de la compétence orthographique ; les formes de la cohérence et de la cohésion textuelles au cours de la scolarité ; les interventions des enseignants sur les copies et leur rôle dans l’apprentissage de l’écrit.

Les différentes communications ont montré que le recueil et l’exploitation de corpus scolaires et universitaires peuvent se décliner à partir de questionnements, de méthodes et de choix terminologiques très hétérogènes. Le volume que nous présentons reflète cette variété théorique et méthodologique. Il réunit, après un premier article d’ouverture sur la constitution et le traitement du corpus, quinze contributions organisées en deux parties. La première met en avant l’analyse de corpus d’apprenants de différents niveaux et avec différentes portées : orthographique, syntaxique, textuelle. La deuxième partie se focalise sur les pratiques des enseignants en termes de choix d’outils et de gestes professionnels pour enseigner.

La première contribution de Claire Doquet et de Claude Ponton, qui ouvre notre volume, présente le corpus E-CALM, les choix de transcription et les perspectives envisagées. Les auteurs explicitent **les choix théoriques et techniques opérés pour la transcription et l’annotation des manuscrits**, en vue de la numérisation du corpus. Les auteurs mettent l’accent sur des perspectives à venir du projet : augmentation du nombre de copies disponibles sur la plateforme et amélioration des recherches critériées qui permettront de sélectionner par exemple les écrits répondant au mieux aux questionnements des utilisateurs.

## **1. Analyse de corpus d’apprenants**

La première partie de ce volume réunit huit articles qui analysent des corpus d’apprenants à différents niveaux linguistiques, orthographique, morphologique, syntaxique ou textuel, pour proposer des réflexions didactiques et des leviers pédagogiques.

Trois contributions explorent la **dimension orthographique**. La première, de Anne Dister et Marie-Louise-Moreau, considère les spécificités du système alphabétique

français en étudiant les correspondances phonèmes-graphèmes dans un corpus de près de 12 000 dictées produites lors d'un championnat d'orthographe. Les élèves utilisent un graphème plausible pour transcrire le phonème-source, avec un écart inférieur à 1%, alors que les difficultés majeures sont concentrées autour d'unités moins familières pour l'enfant.

Le deuxième article, de Christian Surcouf, propose une réflexion sur l'orthographe du pluriel articulant les domaines de la morphographie, de la linguistique diachronique et de l'apprentissage de l'orthographe. L'étude part du concept d'intuition linguistique pour montrer comment les élèves s'inscrivent dans l'évolution de leur langue et ne font, somme toute, que poursuivre un processus historique qui stabilise le rôle décisif des déterminants et réduit l'importance accordée aux marques du pluriel dans les autres formes. L'étude permet de relativiser un discours souvent fort stigmatisant de l'erreur et de repenser le statut de la norme orthographique.

Le troisième article, d'Agnès Furman, explore la très grande variabilité du verbe français en situation de production de texte, à partir d'un corpus de rédactions de collégiens. En se focalisant sur les formes de la personne 2 ainsi qu'aux formes de passé simple à la personne 3, l'étude adopte une perspective didactique et vise à identifier les difficultés de collégiens en matière d'orthographe verbale.

Les trois contributions suivantes s'intéressent aux **compétences syntaxiques** des scripteurs.

La première, de Dominique Legallois et Christophe Benzitoun, présente une analyse exploratoire d'un corpus longitudinal du CE1 au CM2 ayant l'objectif d'identifier les évolutions des performances rédactionnelles grâce à une méthode informatisée qui met en lumière l'élaboration syntaxique des textes. Leurs résultats « donnent un aperçu de ce que pourrait être le répertoire par niveau des formes linguistiques relevant de la compétence rédactionnelle », ce qui permettrait « aux enseignants d'identifier chez les élèves la présence ou l'absence de formes caractéristiques de leur niveau ».

La deuxième contribution d'Eduardo Calil, Véronique Paolacci, Kall Anne Amorim et Cristina Felipeto analyse en revanche l'emploi des marques de ponctuation. Cette étude comparative révèle des emplois contrastés de marques de ponctuation dans des histoires inventées par des élèves de 6 ans de deux classes : l'une brésilienne, l'autre française. À

travers une capture multimodale des processus d'écriture en temps et en espace réels, des différences de nombre et d'usage des marques de ponctuation ont pu être constatées.

Marie-Noëlle Roubaud, quant à elle, s'intéresse aux procédés syntaxiques employés par les élèves pour renforcer et épaissir leurs textes. L'article s'appuie de manière originale, entre autres, sur les méthodes de description microsyntaxique pour analyser les textes d'élèves et décrire des comportements scripturaux chez les apprenants. Les résultats de l'étude fournissent au lecteur une typologie des procédés syntaxiques à l'œuvre dans les textes du corpus étudiés. Cette typologie pourrait permettre aux enseignants en formation d'outiller les élèves sur les procédés pour améliorer syntaxiquement leurs écrits.

Deux dernières recherches clôturent cette partie en mettant en avant des **analyses textuelles**.

Anne Sardier revisite la notion d'isotopie lexicale, entité « à la croisée de la compétence scripturale et de la compétence lexicale », à partir d'un extrait de corpus EcriScol. Elle étudie la manière de gérer le déploiement des isotopies « pour mieux appréhender la cohérence et la cohésion » dans des récits du CE1 au CE2. La présence d'une seule isotopie lexicale liée au texte source fournit des pistes didactiques pour les étapes de réécriture.

Myriam Bras et Laure Vieu présentent ensuite dans leur article une démarche d'annotation du corpus ResolCo selon un modèle d'annotation capable de définir des indicateurs de la progression de la cohérence dans des textes d'élèves de cycle 3. Les auteures se focalisent sur les relations de discours dans des écrits narratifs éloignés de la norme et fournissent des mesures de la cohérence prometteuses. Les résultats montrent la manière dont évolue la compétence textuelle des élèves et ouvrent la voie à une réflexion sur des leviers pédagogiques en matière d'enseignement-apprentissage de la cohérence textuelle.

## **2. Analyse des pratiques enseignantes**

La deuxième partie regroupe les articles qui s'attachent aux pratiques enseignantes, dans leurs interactions avec les élèves et dans les situations d'apprentissage impliquant l'écrit, sous diverses formes. Les articles prennent appui, d'une part, sur les

gestes professionnels des enseignants – et notamment celui de la correction des écrits d'élèves – mais également sur d'autres aspects de l'écrit dans le cadre de la classe : affichages, discours de l'enseignant dans l'interaction avec les élèves.

Deux premières contributions mettent en avant ces « **outils pour enseigner** ». Le premier, de Belinda Lavieu-Gowdz, analyse les affichages didactiques présents dans les classes de fin d'école élémentaire, réalisés par les enseignants dans le cadre de l'étude de la langue au cycle 3. Pour l'auteure, leur analyse permet de mettre au jour les indices révélateurs du rapport à la langue (et à son enseignement) des enseignants, l'amenant à considérer ces affichages davantage comme des aides-mémoires qui ne favorisent pas l'autonomie des élèves dans le raisonnement métalinguistique attendu d'eux au regard de l'étude de la langue au cycle 3.

Le deuxième article, de Gilles Tabourdeau et de Marianne Vergez-Couret, s'intéresse à la structuration progressive du discours narratif des enfants de maternelle. Pour ce faire, les auteurs reposent leur analyse sur une recherche exploratoire menée à partir de situations de dictées à l'adulte. Pour rendre compte de cette structuration progressive, ils proposent une typologie des stades de développement du discours narratif, tout en soulignant l'importance du discours de l'enseignant durant ces dictées, ce qui leur permet de présenter différentes postures étayantes et d'analyser les micro-gestes de négociation destinés à guider les élèves.

Cinq contributions s'intéressent de près à un outil indispensable de l'enseignant : le geste professionnel de **correction des écrits d'élèves**. Parmi ces cinq articles, trois s'attachent à l'enseignement scolaire primaire et secondaire, ainsi qu'aux cycles 2, 3 et 4 des apprentissages. Les deux autres articles s'intéressent au geste de la correction tel qu'utilisé en contexte adulte, auprès de publics d'étudiants.

Le premier article, de Claudine Garcia-Debanc, Myriam Bras et Laure Vieu, s'intéresse à la cohérence dans les écrits d'élèves d'école primaire et du début de collège. S'appuyant de productions écrites provenant du corpus ResolCo, les auteures se proposent notamment de mettre en perspective l'annotation des aspects linguistiques liés à la cohérence dans des textes d'élèves et les marques des enseignants relatives à la cohérence dans ces mêmes textes. L'article permettra au lecteur d'analyser en profondeur

les enjeux liés aux indications des enseignants en matière de cohérence, ce qui se révélera utile, entre autres, dans le cadre de la formation à l'enseignement.

Le deuxième article, de Veronica Ujcich, cherche à interroger les pratiques des enseignants italiens de primaire (équivalents au niveau CM2), à travers un questionnaire destiné à recueillir à la fois leur(s) conception(s) de la correction, mais également leurs pratiques déclarées. Les réponses obtenues ont incité l'auteure à mener une recherche-action sur la formation à la correction auprès des enseignants. Elle détaille cette recherche-action, réalisée dans trois contextes différents, et les retours recueillis des suites de cette formation.

Le troisième article, d'Arnaud Moysan, présente une synthèse d'un travail doctoral mené autour du geste professionnel de correction. Prenant appui sur un recueil multimodal constitué d'écrits de collégiens (6<sup>e</sup> et 3<sup>e</sup>) et d'entretiens menés auprès d'enseignants, l'auteur interroge la fonction de la correction individuelle des copies. Pour ce faire, il part du discours des enseignants sur leur pratique, qu'il confronte ensuite avec les observations menées sur les textes d'élèves corrigés, ce qui l'amène à reconsidérer les fonctions de la correction comme étant non pas destinées aux élèves mais plutôt à l'enseignant correcteur qui s'en sert de support pour ajuster les apprentissages.

Du côté de l'enseignement supérieur, deux articles se sont penchés sur la question du geste professionnel de correction. Le premier, de Caroline Scheepers, s'intéresse aux annotations réalisées par des assistants sur des écrits de primoétudiants en droit et notamment à travers le prisme de leur objectif méthodologique puisque ces annotations visent à acculturer les étudiants à la méthodologie du travail universitaire. Le lecteur pourra ainsi découvrir des résultats à la fois inédits et dont les conclusions diffèrent des travaux menés sur le sujet.

Enfin, le dernier article de cette partie, écrit par Beatrice Dal Bo et Corinne Gomila, se propose d'interroger le geste professionnel de correction en situation professionnelle. Pour cela, les auteures ont analysé les corrections réalisées par deux enseignantes sur des copies d'examen blanc d'étudiants en master 2 MEEF (Métiers de l'enseignement, de l'éducation et de la formation). Si leurs résultats confirment ceux de recherches précédentes, notamment du point de vue de la typologie des marques issues de la

correction, ils font également apparaitre des variations amenées par le contexte de l'écrit corrigé.