

"Enjeux des corpus en didactique de la grammaire"

15 mars 2024 : envoi des projets de contribution sous la forme d'un résumé (présentation de deux pages maximum), avec le rattachement institutionnel et scientifique du (des) auteur(s), un titre explicite, cinq mots-clés et une bibliographie sélective.

1 avril 2024 : notification d'acceptation ou de refus des propositions aux auteurs.

15 juin 2024 : envoi des articles (première version) pour évaluation anonymée par deux relecteurs (membres du comité de lecture et du comité de rédaction de la revue Scolagram).

15 octobre 2024 : envoi de la version définitive des articles après éventuelles réécritures demandées aux auteurs.

La linguistique outillée a pris une place très importante dans le champ de la recherche en linguistique et l'utilisation de corpus y semble devenue incontournable au risque d'ailleurs de lui accorder un crédit illimité¹.

S'il importe de ne pas confondre l'outil destiné à investiguer le réel avec le réel lui-même, il est indéniable que l'accent mis sur les corpus de linguistique a permis de mettre au jour des caractéristiques de la langue peu identifiables sans cela. Il reste que les corpus ne sont pas directement lisibles pour les profanes, ils demandent à être soutenus d'une description de la langue quand bien même ce serait pour la confronter à l'usage et d'une réflexion sur les usages statistiques pour en mesurer la part subjective.

Or, l'appréhension des objets de recherche par les corpus de langue investit aussi aujourd'hui la recherche en didactique de la langue. Malgré les défis nombreux que peut poser la constitution de corpus non normés d'apprenants, nombreux sont les chercheurs en didactique de la langue qui consacrent une partie de leur temps de recherche à la constitution de corpus d'élèves, dans une démarche souvent artisanale constituée de repérages et codages manuels dont l'investissement en temps souligne *in fine* l'intérêt des équipes d'enseignants-chercheurs pour cette exploration.

Depuis 2005 et l'initiative de M.-L. Elalouf et Catherine Boré qui ont appelé à la diffusion de corpus scolaires, plusieurs corpus scolaires de grande envergure ont été élaborés et diffusés ou sont en cours d'élaboration et de diffusion, en général sur Ortolang. Par exemple, le *Corpus Ecriscol* (J. David et C. Doquet), le corpus ÉMA, écrits scolaires (C. Boré, M.-L. Elalouf, M.-N. Roubaud), le *Corpus Resolco* (C. Garcia-Deban et K. Bonnemaïson), le corpus *Dynascript* (S.-D. Vogüé *et al*) et le corpus *Scoledit* (C. Ponton, C. Brissaud et C. Totereau).²

¹ De nombreux chercheurs abordent cette question épistémologique (Milner, Charaudeau, Cori et David) devenue de plus en plus vive, semble-t-il et la discussion des limites de la linguistique basée sur des corpus est devenu un passage obligé dans les thèses de linguistique actuelles.

² Cette liste s'appuie sur l'état des lieux exhaustif et récent et effectué par Claire Wolfarth dans sa thèse (2019).

La plupart de ces corpus sont consacrés aux écrits d'élèves et cherchent à donner une image de l'écriture scolaire dans différentes dimensions, dont celle, notable, de la progression de la compétence scripturale.

Même si certains de ces corpus d'écriture scolaire en contexte écologique sont utilisés pour décrire la production linguistique des élèves (par ex, Capeau et Roubaud, 2018), la démarche complémentaire visant construire des corpus scolaires de grammaire pour donner une visibilité des compétences grammaticales des élèves et de leur progression est plus débutante. Et cela alors que les enjeux du développement de ces compétences et de la façon dont elles sont enseignées et mises en œuvre ont pris une importance renouvelée dans l'économie générale de la formation des élèves en français et en particulier dans leur accession à la conceptualisation de la langue. Or, rares sont encore les grands corpus en français destinés à l'analyse de la conceptualisation grammaticale des élèves et aux obstacles qu'elle rencontre.

Ce n'est que récemment, en 2019, que l'équipe de Realang pilotée par P. Gourdet, M. Beaumanoir-Secq et J.-P. Sautot (Sautot *et al.*, 2021) s'est lancée dans l'élaboration de deux corpus quantitativement imposants, l'un en 2019-2020, l'autre en 2022-2023. Ceux-ci regroupent des données autour de l'appréhension des catégories du verbe et de l'adjectif par les élèves de CE2, CM1 et CM2, qui se complètent de données sur des 6^{ème} et des 3^{ème} en 2022-2023. Le premier de ces corpus avec son outillage statistique a déjà permis de mettre au jour des découvertes sur ces compétences scolaires et de réaliser de nombreuses publications.

Cependant, une revue de littérature permet de rendre compte de l'existence de plus petits corpus, rarement diffusés, constitués au gré de thèses ou de recherches plus ciblées. Citons par exemple, Cordary (2005), Geoffre (2013) Beaumanoir-Secq (2016), Bonnal (2016), Le Levier (2019), Fureman (2022), Tallet (2012) mais aussi les travaux Cogis (2004) qui s'appuie sur un corpus constitué par des évaluations nationales en 6^e, ceux de C. Brissaud (1999, 2006, 2009, par exemple) qui interroge la grammaire à travers l'orthographe, ceux de J.-P. Sautot et S. Lepoivre-Duc (2017, par exemple) et au Québec, de M.-C. Boivin (2017, 2018, par exemple). Ces quelques exemples témoignent de l'intérêt pour la constitution de corpus constitués de productions grammaticales d'élèves souvent associées à leurs commentaires métalinguistiques.

Ce numéro de *Scolagram* se donne pour ambition d'entamer une réflexion sur les usages faits des corpus par les didacticiens de la langue dans le domaine de l'enseignement de la grammaire et plus particulièrement dans le cadre de la formation des enseignants.

Tout d'abord, en ce qui concerne les **grands corpus du français écrit ou parlé** (comme Eslo, Frantext, le CNRTL, BDTS, BELTEXT, CERF, CFPP, CIEL, CLAPI, L'Est Républicain), encore peu sollicités dans le cadre de leur didactisation il y a dix ans, selon M.-L. Elalouf (2012) :

- Dans quelle mesure, ces corpus du français sont-ils aujourd'hui utilisés dans le cadre de la formation en didactique de la langue ? Avec quels objectifs et quelles précautions ?
- S'agit-il de fonder des démarches de formation s'appuyant sur les résultats de recherches faits sur ces corpus pour s'extraire de fausses représentations des propriétés des propriétés de certaines unités linguistiques incluant éventuellement une collaboration avec le chercheur à l'origine de ces résultats ? (comme la didactisation de la thèse de J. Thuillier sur la place de l'adjectif en Français Langue Étrangère (Camussi-Ni *et al.*, 2016).

- S'agit-il de modifier la représentation de la langue des enseignants pour donner une représentation plus proche de la réalité, moins normative ? (par exemple, Gadet & Guerin, 2022)
- En dehors de cet objectif de prise en compte des variations de la langue, dans quelle mesure les corpus d'oral spontané ont-ils leur place dans une formation à l'enseignement de la grammaire ? L'appréhension de la grammaire de la langue par l'oral, par exemple des flexions audibles à l'oral, peut-il s'appuyer sur des corpus ?
- Est-il envisageable de promouvoir directement l'utilisation des grands corpus du français pour enseigner la langue comme cela se pratique en langue étrangère (*data-driven*) ? Ce modèle d'enseignement utilisé en Français Langue Étrangère (par exemple, Boulton et Tyne, 2014) peut-il fournir des pistes pour la formation en Français langue première ou pour la formation des enseignants ?³ Faut-il plutôt envisager de traiter ces corpus pour mettre à disposition d'enseignants des ressources plus facilement exploitables, présélectionnées ou annotées pour un objectif d'enseignement spécifique (suivant l'exemple proposé par Parisi et Grossmann, 2009) ?
- Quand un corpus est entré dans l'usage comme la liste de fréquence de Brunet diffusée par Eduscol, fait-il alors l'objet d'une formation ou tout au moins d'une sensibilisation auprès des enseignants qui l'utilisent ou l'utiliseront ? Inversement, à quelle dérive peut conduire dans le domaine scolaire l'emploi d'un corpus sans formation auprès des enseignants ?
- Initie-t-on les enseignants de français du secondaire au corpus *Frantext* composé d'un réservoir d'œuvres littéraires interrogeables linguistiquement ? Son interrogation en vue de construire des exercices de grammaire ou des corpus d'enseignement peut-il constituer un mode de formation des enseignants ?

Ensuite, en ce qui concerne plus particulièrement les **corpus scolaires du français** :

- quel usage est-il fait ou pourrait-il être fait des corpus scolaires dans le cadre de la formation de l'enseignement de la grammaire ?
- Les corpus d'écrits scolaires s'y prêtent-ils ? À quelles conditions ?
- Faudrait-il promouvoir la diffusion des corpus à visée de formation grammaticale ? Quels en sont ou pourraient être les axes et les méthodes de constitution et d'outillage ? Quels travaux préalables faciliteraient la constitution de ces corpus de grammaire ? Inversement, quelles méthodes d'analyse de ces corpus permettent d'en conserver la sérendipité ?
- Comment sont utilisés les corpus scolaires de grammaire dans les formations initiales et continues des enseignants ? Quels apports spécifiques peuvent-ils avoir ?
- Peuvent-ils permettre de faire évoluer les conceptions des enseignants sur la langue à disposition de l'élève et les lacunes d'emploi ?
- Ou encore asseoir une formation des maîtres en termes de progression et d'objectifs se basant sur des âges de compétences des élèves ?
- Quels usages de ces corpus sembleraient les plus pertinents ? Statistiques, comparaisons, inscriptions de ces corpus dans la construction de l'individu : relation avec les autres disciplines enseignées, avec l'écriture, exploration des champs de la lecture et de la compréhension ?
- Dans quelle mesure ces corpus scolaires pourraient-ils motiver une démarche de construction de corpus d'enseignement en mettant l'accent sur des difficultés récurrentes que les enseignants auraient à cœur d'anticiper ?

³ Le numéro 24/2023 de la revue *Corpus* est ainsi consacré à la question de la formation des enseignants à l'usage des corpus numériques pour la didactique des langues.

Enfin, en ce qui concerne ces **corpus d'enseignement**, préconisés pour extraire l'élève d'une posture de pure application et le placer dans celle d'un chercheur en herbe, on peut se demander dans quelle mesure on peut former les enseignants à construire leurs propres corpus pour répondre au plus proche aux besoins de leurs élèves.

En effet, J.-L. Chiss et C. Muller en définissant cette démarche, en souligne la complexité : « L'activité métalinguistique en classe suppose de problématiser des données de langue au lieu d'en fournir immédiatement la description réglée, de s'interroger sur l'appareil notionnel existant et donc de prévoir, à côté des leçons magistrales, des démarches effectives de recherche apprenant à poser un problème à partir d'un corpus, d'une règle, d'une définition et à réunir les conditions permettant de le résoudre ». (Chiss et Muller, 1993, p. 47).

De leur côté, Parisi et Grossmann analysent très finement les écueils de la démarche inductive attachée à ces corpus d'enseignement et en particulier son exigence :

« La démarche dite inductive, correspond bien en didactique à une stratégie visant à ce que les élèves réfléchissent et raisonnent « pour de vrai ». Entendue en ce sens, elle est particulièrement exigeante, puisqu'elle requiert du maître qu'il envisage avec précision les obstacles possibles au cheminement intellectuel de l'élève, y compris en ce qui concerne les éléments liés à la notion linguistique elle-même. » (Parisi et Grossmann, 2009, p. 168)

Fabriquer un corpus d'enseignement, c'est en somme extraire de la langue un « échantillon » qui rende compte des caractéristiques d'un fait de langue, caractéristiques qui seraient diluées dans le corpus plus large, celui de l'usage, auquel est soumis l'élève. Ce peut être aussi repérer et déconstruire une difficulté récurrente dans l'appréhension de la langue. Cette articulation entre grands corpus, corpus scolaires et corpus d'enseignement est-elle ou peut-elle être l'objet de didactisations destinées aux enseignants en formation initiale ou continue ? Quel modèle de formation des enseignants est-on amené à préconiser si l'on souhaite qu'ils aient les compétences pour évaluer la pertinence des corpus de découverte proposés dans les manuels et pour fabriquer des corpus qui répondent au plus près, voire dans l'immédiateté, aux besoins de leurs élèves ?

Beaucoup de questions se posent et ce numéro de *Scolagram* ne pourra pas répondre à toutes mais pourra amorcer des réflexions. Il acceptera aussi bien des articles qui rendront compte d'expériences de formation à l'utilisation et à la fabrication de corpus que ceux qui s'inscriraient dans une réflexion sur ces questions didactiques.

Format des contributions :

Les dossiers de la revue *Scolagram* tentent de proposer des contenus de types variés.

Le but est d'y réunir, autant que faire se peut, des documents de statuts divers qui rendent compte de la chaîne de transposition didactique au sein du thème abordé. C'est pourquoi, si les articles scientifiques y sont fréquemment majoritaires, il convient de ne pas s'interdire d'imaginer des propositions de nature différente :

- Des articles scientifiques n'excédant pas 40 000 signes, éventuellement accompagnés d'une annexe de volume raisonnable.
- Des propositions pédagogiques ou didactiques commentées
- Des exemples de corpus accompagnés de leur contexte d'utilisation
- Des comptes-rendus d'expérimentation en classe
- Des mémoires de fin d'étude dédiés à la question

Contact :

marie-armelle.camussi-ni@univ-rennes2.fr

Références :

Beaumanoir-Secq M. (2016). *Le tri de mots : pour une grammaire utile aux élèves, dans la continuité et la cohérence*. [Thèse de doctorat, Université de Cergy-Pontoise], sous la direction de M.-L. Elalouf

Beaumanoir-Secq M. (2018). *Conceptualiser les classes de mots - A la recherche d'une grammaire utile aux élèves, dans la continuité et la cohérence*. Gramm-R, Peter Lang.

Boivin, M.-C., Roussel, K. & Pinsonneault, R. (2017). Phrases complexes et maturité syntaxique : une comparaison entre des écrits d'élèves de 13 et 16 ans. *Lidil*, 55.

<http://lidil.revues.org/4206>

Boivin, M.-C. & Pinsonneault, R. (2018). Les erreurs de syntaxe, d'orthographe grammaticale et d'orthographe lexicale des élèves québécois en contexte de production écrite. *Revue canadienne de linguistique appliquée*, 21(1), 43-78.

<https://journals.lib.unb.ca/index.php/CJAL/article/view/25121>

Bonnal, K. (2016). *L'orthographe telle qu'elle s'enseigne : pratiques d'enseignement de l'accord sujet-verbe observées à la fin de l'école primaire*. [Thèse de doctorat, Université de Toulouse 2], sous la direction de C. Garcia-Debanç.

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., Roubaud, M.-N. & Elalouf, M.-L. (2022). Corpus ÉMA, écrits scolaires [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage)

www.ortolang.fr, v3, <https://hdl.handle.net/11403/ema-ecrits-scolaires-1/v3>.

Boulton, A. & Tyne, H. (2014). *Des documents authentiques aux corpus : Démarches pour l'apprentissage des langues*. Paris. Didier.

Brissaud, C. & Sandon, J.-M. (1999). L'acquisition des formes verbales en /E/ à l'école élémentaire et au collège, entre phonographie et morphographie. *Langue Française*, 124, 40-57.

Brissaud, C., Chevrot, J.-P. & Lefrançois, P. (2006). Les formes verbales homophones en /E/ entre 8 et 15 ans : contraintes et conflits dans la construction des savoirs sur une difficulté orthographique majeure du français. *Langue française*, 151, 74-93.

Brissaud, C. (2009). Quels enseignements tirer de quatre-vingt-dix-huit dictées de Troisième ? Du décalage entre prescription et acquisition des élèves. Brissaud, C., Grossmann, F. (coord.) *Repères*, 39, La construction des savoirs grammaticaux. Lyon : INRP.

Camussi-Ni, M.-A., Coatéval A. & Thuilier, J. (2016). Exploiter un travail de corpus sur la place de l'épithète pour élaborer une progression didactique en FLE. Dans A. Kamber & M.

Dubois (dirs.). Corpus, grammaire et français langue étrangère : une concordance nécessaire, *Linguistik online*, 78. 41-54.
<https://bop.unibe.ch/linguistik-online/article/view/2949>.

Charaudeau P. (2009). Dis-moi quel est ton corpus, je te dirai quelle est ta problématique. *Corpus*, 8.
<http://journals.openedition.org/corpus/1674>.

Chiss, J.-L. & Muller, M. (1993), *Recherches en didactique de la langue et des discours*, Paris, INRP.

Cogis, D. (2004). Une approche de la morphographie. L'exemple d'une séquence sur l'accord de l'adjectif. *Lidil*, 30.

Cori, M. & David, S. (2008). Les corpus fondent-ils une nouvelle linguistique ?. *Langages*, 171, 111-129.
<https://www.cairn.info/revue-langages-2008-3-page-111.htm>.

Corpus 24 | 2023, Les corpus numériques pour la didactique des langues : de la formation des enseignants à l'élaboration de dispositifs d'apprentissage.
<https://journals.openedition.org/corpus/7438>.

Doquet, C. & David, J. (2018). Collecter, interpréter, enseigner autrement l'écriture. Analyse linguistique des écrits d'élèves. *Repères*, 57.

Doquet, C., Enoiu, V., Fleury, S. & Mazziotti, S. (2017). Problèmes posés par la transcription et l'annotation d'écrits d'élèves. *Corpus*, 16.
<http://www.univ-paris3.fr/corpus-ecriscol-300513.kjsp?RH=1416243625396>

Elalouf, M.-L. (dir.) (2005). *Écrire entre 10 et 14 ans. Un corpus, des analyses, des repères pour la formation*. Paris, Scérén, CRDP Versailles, CDDP Essonne.

Elalouf, M.-L. (2012). La didactique de la grammaire dans 20 ans de la revue *Repères*, *Repères*, 46.
<https://doi.org/10.4000/reperes.86>

Furman, A. (2022). L'orthographe des verbes dans les dictées et les rédactions des collégiens. [Thèse de doctorat, Sorbonne Université], sous la direction de S. Plane.

Gadet, F. & Guérin, E. (2022). Décrire le 'français tout court' : pourquoi et à quelles conditions ?. *Travaux de linguistique*, 84-85, 159-172.
<https://doi.org/10.3917/tl.084.0159>

Gagnon, R. (2013). De l'analyse de productions écrites d'élèves et de ses usages potentiels pour la formation des enseignants du secondaire en grammaire. *Lidil*, 47.
<http://journals.openedition.org/lidil/3261>

Garcia-Debanc, C. & Bonnemaïson, K. (2014). La gestion de la cohésion textuelle par des élèves de 11-12 ans : réussites et difficultés, *Actes du Congrès Mondial de Linguistique Française* (CMLF 2014), Berlin, Germany, 961-976.

Geoffre, T. (2013). Vers le contrôle orthographique en cycle 3 de l'école primaire. Analyses psycholinguistiques et propositions didactiques. [Thèse de doctorat, Université de Grenoble], sous la direction de C. Brissaud.

Gourdet, P. & Roubaud, M.-N. (2022). Les discours grammaticaux sur le verbe par des élèves de CE2 et de CM2. *Scolia*, 36.

<https://doi.org/10.4000/scolia.2225>

Lefrançois, P. (2009). Évolution de la conception du pluriel des noms, des adjectifs et des verbes chez les élèves du primaire. *Repères*, 39.

Le Levier, H. (2019). Mise en œuvre et perception de l'orthographe française chez des élèves de collège et des étudiants de sections de technicien supérieur. [Thèse de doctorat, Université de Grenoble], sous la direction de C. Brissaud.

Nonnon, É. (1999). "Tout un nuage de philosophie condensé dans une goutte de grammaire" : interactions verbales et élaboration de règles dans la mise en œuvre d'une "démarche inductive" en grammaire. *Pratiques*, 103/104, 116-148.

Parisi, G. & Grossmann, F. (2009). Démarche didactique et corpus en classe de grammaire : le cas du discours rapporté. *Repères*, 39.

Roubaud, M.-N. & Garcia-Debanc, C. (2014). L'approche d'"anomalies" dans des textes narratifs d'élèves de fin d'école primaire (10-11 ans). Dans Avanzi M. et alii (dirs.). *Enseignement du français : les apports de la recherche en linguistique*, Peter Lang, Bruxelles-Bern, 307-326.

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

Sautot, J.-P & Lepoivre-Duc, S. (2017). Saisir les relations temporelles dans des récits d'élèves : quels enjeux pour l'analyse de corpus ? *Corpus*, 16, 265-287.

Sautot, J.-P., Gourdet, P. & Beaumanoir-Secq, M. (dirs.) (2021). Méthodologie pour l'étude des rendements des classes. *Scolagram*, 8, Realang.

Sémidor, S. (2015). La genèse orthographique chez l'élève de cours préparatoire : description de la langue et pratiques d'enseignement. *Éducation*. Université de Bordeaux.

Tallet, C. (2012). Mots-outils homophones hétérographes. Leur enseignement à l'école primaire. [Thèse de doctorat, Université Paris 3], sous la direction de Danièle Manesse.

Vaubourg, J.-P. (2017). Apprendre à réaliser les accords au cycle 3 de l'école primaire. Aspects linguistiques, psycholinguistiques et didactiques. [Thèse de doctorat, Université Paris 4], sous la direction de Sylvie Plane.

Vogüé, S. D., Espinoza, N., Garcia, B., Perini, M., Sitri, F. & Watorek, M. (2017). Constitution d'un grand corpus d'écrits émergents et novices: principes et méthodes. *Corpus*, 16.

Voiriot-Cordary, N. (2005). Acquisition et gestion de la morphologie verbale flexionnelle en français à l'entrée au lycée. [Thèse de doctorat, Université de Dijon], sous la direction de Ghislaine Haas.

Wolfarth, C. (2019). Apport du TAL à l'exploitation linguistique d'un corpus scolaire longitudinal. [Thèse de doctorat, Université Grenoble-Alpes], sous la direction de C. Brissaud et C. Ponton.