

Écriture et forme scolaire : spécificités de transcription et de traitement

Claire Doquet, Luisa Revelli, Arnaud Moysan

DANS **LANGUE FRANÇAISE** 2021/3 N° 211, PAGES 21 À 36
ÉDITIONS **ARMAND COLIN**

ISSN 0023-8368

ISBN 9782200933616

DOI 10.3917/lf.211.0021

Article disponible en ligne à l'adresse

<https://shs.cairn.info/revue-langue-francaise-2021-3-page-21?lang=fr>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour Armand Colin.

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur cairn.info/copyright.

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.

Claire Doquet

Université de Bordeaux (France)

Luisa Revelli

Université de la Vallée d'Aoste (Italie)

Arnaud Moysan

Université Sorbonne Nouvelle (France)

Écriture et forme scolaire : spécificités de transcription et de traitement

On n'envisage guère les écrits des élèves comme des espaces où se déploient les potentialités de l'écriture. Tendus vers le texte final dont ils ne paraissent souvent que des prémices imparfaites, les brouillons ne semblent mériter d'attention qu'en vertu du texte auquel ils sont censés conduire. L'esprit de la génétique textuelle – considérer les « avant-textes » comme des écrits qui ont constitué un aboutissement, même provisoire, et n'ayant *a priori* « pas d'après » (Bellemin-Noël, 1982 : 163) – et la notion d'*écrits intermédiaires* (Chabannes & Bucheton 2000), ni intrinsèquement brouillons, ni textes terminaux, n'ont pas vraiment infléchi cette tendance. La vertu heuristique et créative de l'écriture est encore peu sensible dans le système scolaire, dans lequel les productions écrites s'inscrivent dans des attendus fortement codifiés, aux plans graphique et linguistique à la fois, comme si ce qui se joue dans l'écriture au moment de l'apprentissage était extérieur à ce qu'y cherchent les scripteurs experts, écrivains ou non.

Réunir un nombre important d'écrits scolaires permet d'en donner à voir la variété, les standards et les écarts, les règles et leurs détournements. Nonobstant l'hétérogénéité de l'ensemble constitué, le caractère scolaire de l'écriture met en jeu des caractéristiques propres à la sphère socio-discursive de l'école et rend visible ce que les élèves s'approprient de l'écrit, dans ses spécificités : organisé sur un espace plan, l'écrit joue de la spatialité pour exprimer des variations d'ordre énonciatif aussi bien que générique (Lebrave 2009 ; Lefebvre 2007) ; échappant à l'immédiateté de l'oral, la parole écrite se double d'une couche métalinguistique née de la (re)lecture et matérialisée par les ratures et rectifications (Fabre 1990 ; Doquet 2011). Nous réfléchirons à ces questions à partir de deux corpus d'écrits scolaires : ÉCRISCOL, qui rassemble deux mille écrits

d'élèves et leurs avant-textes ; CODISSC, qui réunit les reproductions numériques de cahiers d'élèves rédigés depuis l'Unité de l'Italie (documents datés de 1863 à nos jours) en Vallée d'Aoste, région italienne transfrontalière à vocation multilingue. Il s'agit, au-delà de l'exposition d'écrits divers, de faire émerger des caractéristiques propres au genre, au niveau scolaire, au type de consigne, au contexte socio-économique et didactique, entre autres, pour parvenir à décrire la manière dont les élèves s'inscrivent progressivement dans la « raison graphique » (Goody 1979) et la langue écrite, « italiano scolastico » (Revelli 2013) ou français « du dimanche » (Blanche-Benveniste 2013).

ÉCRISCOL comme CODISSC cherchent à donner à voir les écrits scolaires dans leurs spécificités, en tant qu'objets scolaires mais surtout en tant qu'écrits, et de permettre leur analyse automatisée. La question se pose alors de la manière de traiter ces écrits pour à la fois préserver la « puissance expressive de la substance graphique des manuscrits » (Lebrave, 1999 : 371) et rendre repérables lors d'investigations informatiques les critères qui les distinguent. Démarche différentielle, qui cherche à extraire d'un continuum sémiotique – le manuscrit – des éléments discriminants. Démarche catégorisante, qui construit à partir du groupement de certains éléments des paradigmes opératoires pour caractériser les données. Démarche indispensable enfin, puisqu'au-delà d'une dizaine de copies il est très difficile de se passer de l'analyse automatique pour faire des investigations précises. Cette démarche s'appuie principalement sur un modèle de traitement, celui de la génétique textuelle, privilégiée car elle permet d'analyser non seulement les écrits mais l'écriture elle-même. Si maints corpus littéraires (fac-similés et transcriptions) sont déjà en ligne ¹, les écrits scolaires n'entrent pas toujours aisément dans les conventions existantes. Nous nous pencherons ici sur certaines de leurs spécificités pour nous demander ce que leur mise en corpus permet de conserver *vs* oblige à modifier des principes et techniques de la génétique textuelle, d'un double point de vue :

- la manière de rendre compte, dans la transcription, des modalités d'effectuation des opérations d'écriture ;
- la place de la sémiotique visuelle dans les copies scolaires.

1. DONNER À COMPRENDRE L'ÉCRITURE : APPROCHE GÉNÉTIQUE

Au départ consacrée à l'étude des manuscrits d'écrivains, la génétique textuelle s'appuie sur le recensement et l'analyse de l'ensemble des avant-textes d'une œuvre (plans, brouillons, ébauches, notes, etc.) dans le but de comprendre le cheminement de l'écriture. Le brouillon, joliment situé « entre bouillon et brouillard » par P. Lejeune (1998 : 7), constitue en littérature un espace de découverte et de création. À l'école, nonobstant les programmes scolaires valorisant

1. Cf. p. ex. les manuscrits de Stendhal (<http://stendhal.demarre-shs.fr>), ceux de Flaubert (<https://flaubert.univ-rouen.fr/manuscrits/>) ou les archives de Zola (<http://www.archives-zoliennes.fr>).

le brouillon et la réécriture², les connotations péjoratives du terme demeurent. Difficile en effet de concilier une image valorisée du brouillon avec les collocations du terme et son usage adjectival qui renvoient toujours à une déficience : *travail brouillon / élève brouillon*, entre autres, et, en italien, *brutta copia*, contrastivement à la *bella copia* qu'est la copie finale. C'est pourtant à une réhabilitation du brouillon, comme porteur de tâtonnements scripturaux interprétables comme des interrogations métalinguistiques, que se sont livrés les généticiens étudiant les écrits scolaires (entre autres Fabre 1990 ; Lamothe-Boré 1999 ; Doquet 2011).

Pour rendre accessible et analysable la matière verbale propre à chaque copie, il a fallu adapter la génétique textuelle, forgées sur des manuscrits littéraires, aux brouillons et plus largement aux écrits des élèves (Doquet 2011 ; Revelli 2013). La raison d'être de l'analyse génétique, mettre au jour la singularité scripturale, entre en tension avec l'exploration d'un corpus où chaque manuscrit intéresse davantage pour sa représentativité que pour son caractère unique : l'étude des écrits d'un seul élève intéresse surtout en tant que représentant les performances d'un élève à cet âge, à ce niveau de scolarité, etc. En plus de travailler sur les nécessaires ajustements des techniques déployées par la génétique textuelle sur les manuscrits littéraires, il faut procéder à des approches (aussi) quantitatives et non (uniquement) qualitatives, qui permettent de comparer et de généraliser les phénomènes observés.

1.1. Quelques spécificités de l'activité d'écriture à l'école

Les écrits scolaires s'inscrivent dans une conventionnalité artificielle : conçus comme s'adressant à des destinataires fictifs, ils sont en fait destinés d'abord à l'enseignant ; dans la tâche scolaire, l'envie d'écrire, quand elle existe, se double pour l'élève de la nécessité de répondre à une commande. De la même manière, la lecture que fera un enseignant de la copie d'un élève, si elle peut mettre en jeu une lecture coopérative (Eco 1985), est toujours également une lecture contrôle (Pilorgé 2010). Pour permettre des investigations informatiques, il faut tenir compte de ces caractéristiques spécifiques des productions écrites scolaires.

D'abord, les écrits scolaires comportent souvent plusieurs versions : on trouve le plus fréquemment un brouillon et une version 2, parfois une version 3. Ces différentes versions sont, dans les deux corpus ÉCRISCOL et CODISSC, présentées ensemble³ ; cette mise en regard permet d'observer la genèse des écrits, les textes présents sur le brouillon étant considérés *a priori* comme les plus proches de la pratique spontanée des scripteurs.

Ensuite, les écrits scolaires reflètent une acquisition non stabilisée des compétences linguistiques et scripturales : contrairement aux écrits populaires qu'ils

2. <https://www.education.gouv.fr/programmes-et-horaires-l-ecole-elementaire-9011>

3. http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/ECRISCOL/CM2/2016_17/PNT/CLASSE_1/DEVOIR_2/EC-CM2-2016-PNT-D2.html

rappellent parfois, ceux des « peu lettrés » (Branca-Rosoff & Schneider 1994) que l'on a pu nommer « semi-cultivés » (Hans-Bianchi 2005) et qui manifestent des compétences cristallisées, les écrits scolaires peuvent être observés dans leur dimension évolutive. ÉCRISCOL et CODISSC rendent possible l'étude des écrits par auteur, ce qui permet de décrire le devenir des acquisitions, la progression des compétences, l'apparition de phénomènes d'hyper-correction et aussi l'éventuelle fossilisation de solutions résistant aux préceptes du canon scolaire. Ce troisième niveau d'étude permet d'entrer dans la genèse des scripteurs.

L'ensemble de ces éléments ne peut être compris sans tenir compte du contexte de production de l'écrit. L'analyse des seules copies manuscrites suscite en effet un certain nombre d'interrogations auquel on ne peut répondre sans informations externes, parmi lesquelles la consigne d'écriture donnée par l'enseignant, assortie des attendus qu'il cherche à évaluer. Ces attendus, explicités par la consigne d'écriture ou totalement implicites, motivent à la fois la production de l'élève et les réactions de l'enseignant. Voici un exemple de consigne, donnée en classe de troisième :

Imaginez qu'après la mort d'Antigone, le tyran Créon comparaisse devant une cour internationale de justice. Vous commencerez par rédiger un résumé objectif des faits qui sera énoncé par le juge. Ensuite vous rédigerez la plaidoirie (défense) ou le réquisitoire contre Créon.

Si la thématique et l'organisation apparaissent dans cette consigne, le barème de notation, donné en même temps, livre d'autres attentes : bonne présentation de la copie, correction de la langue, pertinence du résumé, qualité de l'argumentation dans le discours, présence de figures de style. Le commentaire de l'enseignant sur une des copies met en valeur ces éléments tout en faisant apparaître d'autres exigences, les connaissances de l'élève et la précision du rendu :

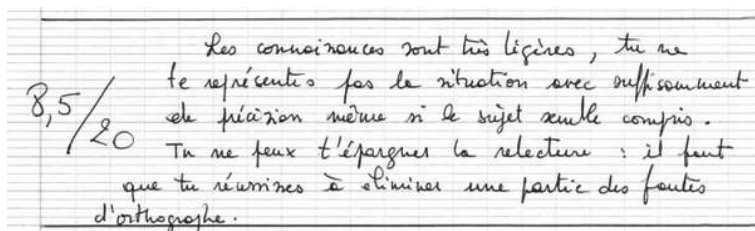


Figure 1 : Copie de 3^e – 2018 [corpus de thèse de Moysan (en prép.)]

La prise en compte de ces différents éléments permet de contextualiser les annotations des enseignants sur les copies, qu'il s'agisse de commentaires ou de corrections directes dans le texte, mais aussi les différents énoncés des élèves, au brouillon comme dans la copie finale. Une des caractéristiques des écrits scolaires est la co-écriture, au sens où l'écrit est là parce que l'enseignant l'a demandé, et que ce dernier y fait des commentaires sur lesquels l'élève s'appuiera pour réécrire (Doquet & Pilorgé 2020). Il faut donc faire entrer systématiquement

dans la transcription la notation de l'identité du scripteur, élève ou enseignant au moins, pour en déduire la stratigraphie et lire les différents états du texte comme des espaces où se déploient les potentialités de l'écriture.

1.2. Une opération scripturale problématique : le remplacement

Que veut montrer une transcription d'écrits scolaires ? Qu'est-ce qui la différencie d'une transcription de manuscrits d'autres natures ? Quelles spécificités des fac-similés doit-elle faire apparaître ? Nous ne pouvons ici traiter exhaustivement ces questions et nous prendrons l'exemple d'une opération scripturale, le remplacement, qui rassemble différents problèmes et présente, sur les écrits d'élèves, des formes spécifiques.

La première difficulté concernant le remplacement dans les écrits scolaires est son repérage. A. Grésillon (1994) le caractérise comme la combinaison d'une suppression et d'un ajout, quel que soit l'empan textuel concerné (graphème, mot, syntagme, etc.). J.-L. Lebrave (2009 : 14) précise que les opérations d'écriture sont « encadrées par un contexte stable à l'intérieur duquel elles prennent sens ». Les exemples donnés dans la littérature génétique font tous intervenir un élément A biffé et un élément B juxtaposé ou superposé à A. C'est le cas dans l'exemple suivant dans lequel un nom propre est remplacé par un pronom anaphorique :

de tous les sens. ~~Il~~ <je> regardais de tous les côtés. La sage femme ~~prend Lucien~~ <me prend> dans ses bras et ~~posa Lucien~~ <me posa> sur le ventre de ~~sa~~ <ma> maman. ~~Sa~~ <ma> maman était tres contente. ~~Lucien~~ <je> pleurait

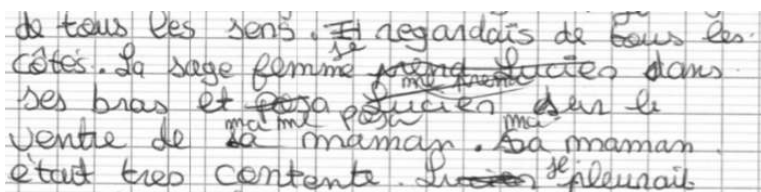


Figure 2 : Extrait d'un brouillon de 3^e [ÉCRISCOL a]

a. http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/ECRISCOL/3EME/2014_15/ARG/CLASSE_1/DEVOIR_1/TEI/CO-3-2014-ARG-D1-E5-V1-tei-1.html

Ces nombreux remplacements sont dus à la consigne d'écriture, qui demande une narration en *je* à partir d'un texte littéraire hétérodiégétique. Les références au héros, Lucien, sont systématiquement remplacées par le pronom personnel de première personne, de même que les déterminants possessifs de P3 sont remplacés par la P1. Une annotation des opérations, non encore réalisée à ce jour, pourrait signaler le changement de personne et la valeur énonciative des remplacements qui marquent une difficulté de l'élève à s'inscrire comme

Sur le plan spatial, la configuration est la même que précédemment (Fig. 3) : il n'y a pas de marque d'ajout, donc si l'on en reste à la définition du remplacement comme la combinaison d'une suppression et d'un ajout, cette séquence devrait être classée comme une succession de suppressions. Pourtant, cette succession de biffures se fait dans une logique de remplacement : l'élève formule des essais successifs qui se remplacent l'un l'autre, jusqu'à obtenir une graphie qui le satisfasse. Il s'agit d'un remplacement de forme et non de mot. La fréquence de ce type d'opération dans les écrits scolaires a conduit C. Fabre (1990) à créer la catégorie d'opérations de *mise en graphie*, portant sur le versant signifiant des signes, par rapport aux opérations de *mise en texte* qui portent sur leurs deux versants signifiant et signifié. Cette distinction a des conséquences sur la catégorisation de la chaîne ~~renvers~~ ~~renre~~ ~~renversais~~ renverser : la succession de la suppression suivie de l'inscription du même mot sous une graphie différente (opération de *mise en graphie*) n'a pas les mêmes conséquences sur l'écriture qu'une suppression suivie de l'inscription d'un autre mot (opération de *mise en texte*). Le choix d'un mot ou d'un autre ouvre bien des possibles, et dans l'exemple *supra* (Fig. 3) il n'est pas possible de savoir si le contexte droit observé après l'inscription de B aurait été le même si l'on avait maintenu A. En revanche, on peut penser que la rectification graphique observée en Figure 4 n'a pas d'effet sur le contexte et considérer une stabilité contextuelle, donc classer cela comme un remplacement.

Ce faisant, nous infléchissons les critères originellement distinctifs de l'opération de *remplacement*, mais nous pensons rester dans la logique de l'opération : les différentes formes sont ici dans un rapport de substitutionnalité indépendant du contexte, parce que travaillant uniquement au niveau du signifiant.

Nous venons de mettre au jour un type de remplacement qui ne comporte pas d'ajout. Les écrits scolaires présentent aussi des remplacements qui ne comportent pas de suppression, comme ici :

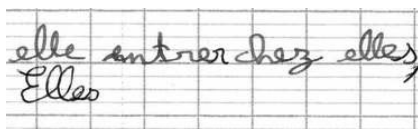


Figure 5 : Extrait d'un brouillon de CM2 [ÉCRISCOL^a]

a. http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/ECRISCOL/CM2/2015_16/EGL/CLASSE_1/DEVOIR_2/TEI/EC-CM2-2015-EGL-D2-E2-V1-tei-1.html

Si l'on a bien un élément B superposé à A, A n'est pas biffé, les deux éléments entrant en relation paradigmatique. À la relecture, l'élève a repéré une double erreur sur *elle* et a proposé, au-dessous, la forme alternative *Elles*. De toute évidence, le projet est le remplacement, comme le montre d'ailleurs le texte final où seule la forme *Elles* subsiste. Ici, comme dans les très nombreux cas où les formes alternatives cohabitent sans biffure, nous traitons cette structure comme un remplacement et, comme dans l'exemple précédent, il s'agit d'un

remplacement de forme et non de mots. Dans l'ensemble des copies examinées, cette configuration particulière de remplacement sans suppression est réservée à la mise en graphie. Elle diffère de manuscrits où cohabitent plusieurs éléments substituables les uns aux autres mais non biffés, considérés par J.-L. Lebrave (2000) comme des « lieux variants »⁴ où se déploient des paradigmes entre lesquels l'auteur n'a pas (encore) choisi. C'est ainsi que l'on pourrait qualifier cet exemple :

Dehor ~~il ferai~~
il ferai tré froi et Dehor il pleuver beaucoup



Figure 6 : Extrait d'un brouillon de CM2 [ÉCRISCOL^a]

a. http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/ECRISCOL/CM2/2016_17/EGL/CLASSE_1/DEVOIR_2/TEI/EC-CM2-2016-EGL-D2-E18-V1-tei-1.html

Si la transcription elle-même ne pose pas de problème, l'interprétation de ce qui se passe est incertaine : ces deux lignes forment-elles un tout ou bien sont-elles disjointes ? Si elles forment un tout, il faut y repérer un remplacement (~~il ferai~~ il ferai) et admettre que ce qui est donné à lire est : *Dehor il ferai tré froi et Dehors il pleuver beaucoup*. Si ce n'est pas le cas, pourquoi avoir biffé *il ferai* sur la ligne 1 ? On accède ici à l'écriture singulière d'un élève, au stade du brouillon, suite de tâtonnements non réductibles à une rationalité pré-établie.

Ces exemples montrent que les écrits scolaires présentent des spécificités d'ordre linguistique, dont l'analyse génétique doit tenir compte pour rendre compréhensible l'écriture. Ils permettent, en l'occurrence, de reconsidérer l'opération de *remplacement* telle qu'elle a été décrite par la génétique textuelle pour amorcer une réflexion sur l'activité métalinguistique constitutive de l'écriture et sur le versant cognitif de l'activité : si l'écriture est sans doute, quel que soit son contexte, une activité où l'énonciation participe de l'élaboration du référent de l'énoncé – autrement dit, si c'est en écrivant que l'on découvre ce que l'on veut écrire –, il y a vraisemblablement deux niveaux de vigilance, ayant trait à la mise en texte *vs* à la mise en graphie, le second pouvant se manifester sans affecter le projet de texte.

4. Voir p. ex. le manuscrit du poème *Été* de Valéry (BNF, Nafr. 19.002, f°57 r°), dans lequel on observe l'alternative non résolue à ce stade de *mange* et *marche* (*une mer marche/mange, bruit qui t'âtre*).

2. SÉMIOTIQUE VISUELLE : LE VERSANT SPATIO-GRAPHIQUE DES ÉCRITS SCOLAIRES

La place du visuel dans l'appréhension des écrits et l'importance de la spatialité dans l'écriture ne sont plus à démontrer : depuis les travaux fondateurs de J. Anis sur le *signifiant graphique* dans un numéro de *Langue française* (Anis 1983), différents spécialistes ont cherché à caractériser l'espace scriptural. J.-L. Lebrave (1999, 2009) donne à la mise en espace un rôle de premier plan dans l'écriture. A. Crasson et L. Hay (2013 : 7) posent comme principe fondateur de la génétique « la conception du manuscrit comme un réseau de signes qui instaure une page en image ». M. Favriaud (2014) parle de l'écrit comme d'un plurisystème où les ponctuations *blanche* et *grise* segmentent et distinguent diverses unités linguistiques et discursives. Appelant à fonder une nouvelle science, la *scripturologie*, J.-M. Klinkenberg et S. Polis (2018) posent l'écriture comme « articulant les faits langagiers et les faits spatiaux ». Nous donnons dans les lignes qui suivent un aperçu de cette articulation et de son traitement dans les copies pensées comme fruits d'un travail à la fois verbal et iconique.

2.1. La forme scolaire : pérennité spatio-graphique des copies

Une copie d'élève se reconnaît et ce qui joue d'abord dans cette reconnaissance, ce sont sa mise en espace et ses attributs (typo)graphiques. Sur les deux feuillets (Fig. 7), une copie italienne de 1915 et une française de 2018, l'espace graphique est organisé en différents blocs qui comportent :

- les éléments paratextuels inscrits par l'élève : identification, intitulé du devoir, date ;
- les espaces scripturaux dévolus à l'enseignant : commentaires portant sur l'ensemble du devoir situés « en chapeau » de la copie (Fournier 1999), marges haute et gauche ;
- le texte de l'élève.

Si des variations sont observables d'un pays à l'autre et d'une époque à une autre (Dancel 2001), on identifie au premier coup d'œil un écrit scolaire (Chartier 2003).

Cette forme graphique spatialisée, à la fois transgénérationnelle et transfrontalière, concerne les copies remises au professeur pour évaluation. Elles obéissent à un ordonnancement topographique déterminé par leur destination, être évaluées, qui nécessite à la fois des indications sur la production (genre d'écrit, discipline scolaire, consigne) et un ou des espace(s) pour les commentaires du professeur. C'est donc l'ensemble des conditions de production de l'écrit qui lui donne forme, l'univers scolaire érigeant, au-delà des formes discursives dont il s'empare et qu'il modèle (Schneuwly 2007), des formes spatio-graphiques qui lui sont propres et dont il faut rendre compte dans la transcription. L'apprentissage de l'écriture est aussi celui de ces formes. J. David (2008) estime que les premières tentatives d'écriture des jeunes enfants visent en premier lieu à remplir la page

et analyse le développement de l'écriture à l'aune de la fonction ponctuationnelle et énonciative (Anis 1983) de la mise en espace ; à l'école, les supports d'écriture constituent des instruments d'apprentissage de la spatialité tout autant que de la textualisation (Doquet 2019).

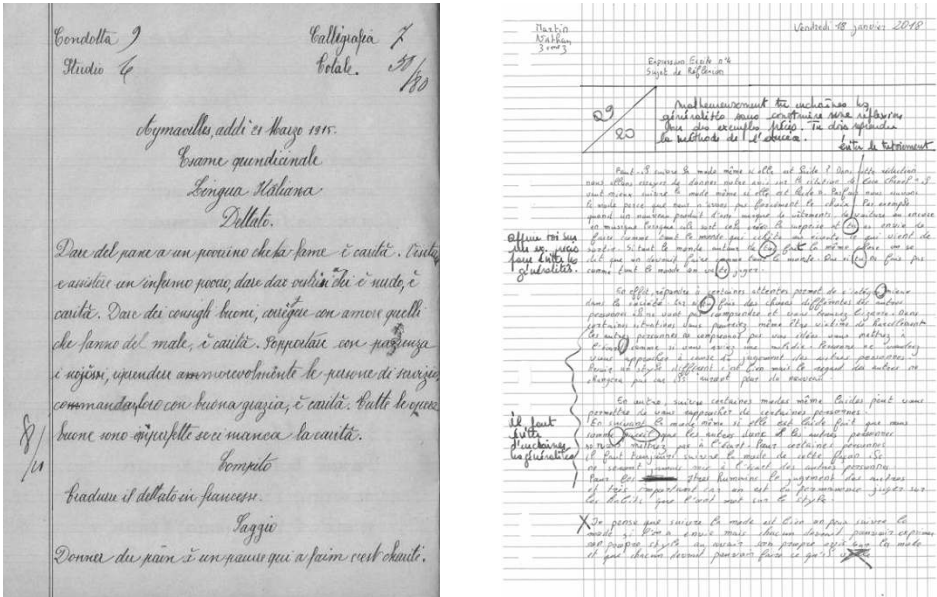


Figure 7 : [g.] Copie italienne de 1915 ^a – [d.] Copie française de 2018 ^b

- a. <http://www.codissc.it/archivio-digitale/quaderno-165>
- b. Copie de 3^e, extraite du corpus de thèse de Moysan (en prép.).

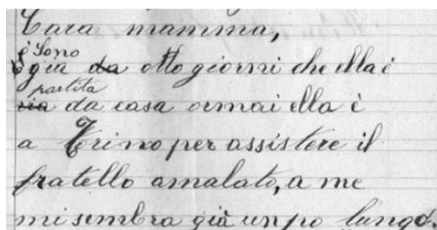
Nous pensons que la créativité scripturale se déploie, chez les scripteurs débutants comme chez les écrivains, également dans cette dimension, et que les transcriptions doivent en rendre compte. Pour comprendre l'ontogénèse de l'écriture, on ne peut limiter l'analyse à la matière verbale car cette dernière s'inscrit dans une familiarisation beaucoup plus vaste qui est celle du jeune scripteur avec l'écriture dans toutes ses dimensions. Effort de conformité et d'inscription dans un modèle pour ce qui concerne les textes finalisés, l'écriture scolaire est aussi, dans les avant-textes, une expérimentation personnelle de la textualisation dans la spatialité écrite. En voici deux exemples :

éléments, pour lesquels il n'existe pas de catégorisation stricte et qui sont donc sans cesse à réinterpréter, illustrent la part non verbale de la création scripturale.

2.2. Choix de transcription dans les corpus ÉCRISCOL et CoDISSC

Quels vont être, dès lors, les choix de transcription, dans le cadre de la constitution de corpus interrogeables par des outils informatiques ? Ce qui est interrogé ici est la manière de donner à voir la disposition spatiale et éventuellement les attributs graphiques des écrits. Comme c'est le cas pour le volet génétique, où il paraît important de matérialiser, dans la transcription, les opérations d'écriture repérées sur les manuscrits, leur versant spatio-graphique demande également à être représenté.

Le corpus valdotain CODISSC a fait le choix de la transcription diplomatique, qui tente de coller au plus près de la physionomie globale de la page manuscrite. Dans le cadre des objectifs scientifiques du projet – qui vise une analyse des écrits scolaires dans une dimension diachronique et diatopique –, il faut reproduire fidèlement les détails des manuscrits, pour identifier des conventions représentatives de l'esthétique scolaire dans différentes périodes historiques et contextes culturels : par exemple, la date est reproduite exactement comme dans l'original, on différencie les guillemets anglais et français selon les choix des apprenants, et la présence sur la feuille d'écriture de dessins ou d'images décoratives est indiquée par une abréviation (DSG = *disegno* «dessin») placée dans la zone correspondante. Sont également reproduits les retours à la ligne, indicatifs des représentations syntaxiques préférentielles des apprenants mais aussi des stratégies d'évitement de partitionnement des mots de césure incertaine :



Cara mamma,

È sono già da otto giorni che ella è via partita da casa ormai ella è a Torino per assistere il fratello amalato, a me mi sembra già un po' lungo.

Figure 9 : Transcriptions proposées dans le CODISSC

La transcription diplomatique est extrêmement utilisée aujourd'hui en génétique textuelle et, grâce aux méthodes de l'ecdotique numérique⁷ (Nougaret & Parinet 2015), il est possible d'encoder chaque détail considéré comme signifiant⁸. La TEI (*Text Encoding Initiative*) a mis au point des conventions permettant de rendre compte de la disposition de l'écrit sur la page mais aussi des diverses variations d'ordre typographique (Burnard, O'Brien O'Keeffe & Unsworth 2007). M. Arabyan (2019) propose la transcription « sémio-diplomatique » qui prend en compte les marges, les indentations, etc. ; bref, l'ensemble de ce que l'auteur nomme la « ponctuation textuelle » en les notant et en les rendant accessibles lors de recherches automatiques mais sans en donner de représentation visuelle. Ce niveau de granularité, adapté à l'étude détaillée des manuscrits, ne convient que partiellement dans un travail qui cherche à traiter automatiquement un grand nombre de données dans un but de généralisation des observations et de caractérisation d'une forme conditionnée par la sphère socio-historique de sa production. Pour tenter de concilier la représentation visuelle avec l'impératif d'explorabilité du corpus par des outils informatiques, le corpus ÉCRISCOL donne des pages écrites une image qui reflète certains traits – mais pas tous – de la spatialité du manuscrit. En voici un exemple :

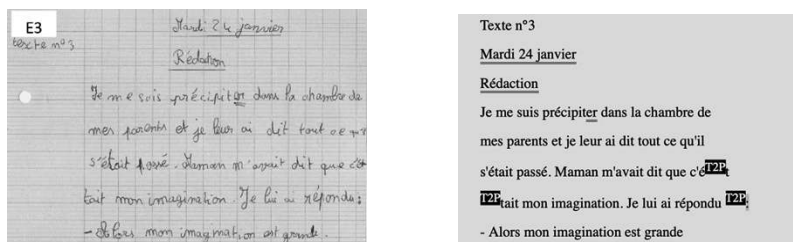


Figure 10 : Copie de CM2 [ÉCRISCOL^a], copie manuscrite [à gauche] – Transcription de la copie [à droite]

a. http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/ECRISCOL/CM2/2016_17/PNT/CLASSE_1/DEVOIR_2/TEI/EC-CM2-2016-PNT-D2-E3-V3-tei-1.html

Cette image des écrits s'appuie sur la TEI mais n'en exploite pas toutes les potentialités. Le choix est fait de respecter la disposition du texte en lignes. En revanche, la topographie du paratexte n'est pas exactement reproduite. Ce choix est lié à la nécessité de ne pas surcharger les transcriptions qui conduisent non seulement aux fichiers TEI comme celui de la Figure 10 (*supra*) mais aussi à des fichiers XML destinés à être décodés par un logiciel d'analyse de données textuelles. La transcription proposée est donc le résultat d'une prise en compte de

7. L'ecdotique est l'art d'éditer des textes d'après une méthode critique qui vise à rendre compte de l'ensemble des aspects visuels des manuscrits (mise en page, trait, etc.). La méthode est mise à l'épreuve sur un manuscrit scolaire par Capron, Giovacchini & Grignon (2020).

8. Cf. p. ex. la transcription en ligne des *Testaments de guerre de Poilus parisiens (1914-1918)* par l'École des Chartes (<http://elec.enc.sorbonne.fr/testaments-de-poilus/index.html>).

la matérialité spatio-graphique du manuscrit doublée de la nécessité de simplifier les données pour les rendre lisibles automatiquement. En faisant ce choix, ÉCRISCOL renonce pourtant à représenter, dans la visualisation des transcriptions, certains traits de la forme scripturale que constitue l'écrit scolaire. Ces derniers sont présents, comme l'ensemble des éléments visuels non directement transcriposables (type de biffure, épaisseur ou nervosité du trait, taille d'un blanc entre deux mots, tracé des lettres, etc.), dans les fac-similés des manuscrits présentés en regard des transcriptions, mais ils ne sont pas codés informatiquement.

3. CONCLUSION

La constitution de corpus d'écrits scolaires est soumise à la double nécessité de rendre compte de chaque écrit, dans sa singularité, et en même temps de caractériser l'ensemble des données réunies et leur variabilité. Une des difficultés principales, lorsque l'on veut parvenir à produire des analyses généralisables, est d'échapper à une vision stéréotypique de l'écriture scolaire. Nous avons présenté, dans cette contribution, la manière dont deux corpus d'écrits scolaires, un français et un italien, traitent les écrits qui les composent pour les rendre « vilisibles » (Anis 1983b) en s'appuyant sur les méthodes de transcription développées par l'ecdotique et la génétique textuelle. L'article expose des traits spécifiques aux écrits scolaires et les ajustements qu'ils imposent aux catégories construites sur des corpus différents, ainsi que les aménagements liés à l'exploration outillée du corpus. La nécessité de catégoriser les observables, condition de leur exploitation, est un des rôles de la transcription, qui s'accompagne toujours d'une perte : la transcription diplomatique en est une, qui sacrifie à la lisibilité des éléments visuels sémiotiquement signifiants comme la légèreté d'un trait ou la force d'une biffure ; la transcription choisie pour le corpus ÉCRISCOL, elle, renonce même à représenter fidèlement certaines caractéristiques de la spatialité, pour privilégier une exploration linéaire de l'écrit et l'exploitation des outils d'analyse de données textuelles. Chaque écrit composant les deux bases peut être analysé dans sa singularité, par l'observation humaine et éventuellement après un repérage semi-automatisé, et l'on peut explorer des ensembles et sous-ensembles de textes avec le traitement automatique pour identifier les caractéristiques évolutives de l'écriture des groupes de scripteurs à différents âges, à différents niveaux scolaires, dans différents contextes.

Il reste assurément beaucoup à faire pour éviter le risque d'aplatissement des données, notamment dans les deux domaines que nous avons explorés ici : la chronologie et la spatialité de l'écriture. La richesse sémiotique du matériau doit conduire, pour ses traits non verbaux, à la construction d'un système d'annotation spécifique, à la mesure de ce qui s'est déjà produit pour le versant verbal des productions. La solution technique réside dans une annotation précise et adaptée au matériau, annotation non réalisée à ce jour mais dont nous espérons avoir esquissé des pistes pour caractériser la forme scolaire de l'écriture.

Références bibliographiques

- [CoDISSc] *COrpus Dlgitale delle Scrittura Scolastiche*, Université de la Vallée d'Aoste (Italie). [<http://www.codissc.it/>]
- [ÉCRISCOL] *Écriture Scolaire*, Clesthia (EA 7345) & Université Sorbonne Nouvelle. [<http://www.univ-paris3.fr/ecriscol-300509.kjsp>]
- ANIS J. (1983a), « Pour une graphématique autonome », *Langue française* 59, 31-44.
- ANIS J. (1983b), « Visibilité du texte poétique », *Langue française* 59, 88-102.
- ARABYAN M. (2019), « La transcription sémio-diplomatique des manuscrits de linguistes », *Linguistique de l'écrit* 1. [en ligne]
- BELLEMIN-NOËL J. (1982), « Avant-texte et lecture psychanalytique », dans L. Hay & P. Nagy (éds), *Le texte et l'avant-texte*, Paris & Budapest, CNRS & Akadémiai Kiado, 161-166.
- BLANCHE-BENVENISTE C. (2013), « La langue du dimanche et la langue de tous les jours », *Tranel* 58, 301-305.
- BRANCA-ROSOFF S. & SCHNEIDER N. (1994), *L'écriture des citoyens : une analyse linguistique de l'écriture des peu lettrés pendant la période révolutionnaire*, Paris, Klincksieck.
- BURNARD L., O'BRIEN O'KEEFE K. & UNSWORTH J. (eds.) (2007), *Electronic Textual Editing*, New York (NY), Modern Language Association of America.
- CAPRON L., GIOVACCHINI J. & GRIGNON S. (2020), « L'ecdotique numérique à l'épreuve d'un manuscrit scolaire byzantin : décrire, déchiffrer et transcrire le *Marcianus gr. XI, 1* », *Humanités numériques* 1. [en ligne]
- CHABANNE J.-C. & BUCHETON D. (2000), « Les écrits « intermédiaires » », *La Lettre de la DFLM* 26, 23-27. [en ligne]
- CHARTIER A.-M. (2003), « Exercices écrits et cahiers d'élèves : réflexions sur des pratiques de longue durée », *Le Télémaque* 24, 81-110. [en ligne]
- CRASSON A. & HAY L. (éds) (2013), *Genesis n° 37 : Verbal/Non verbal*, Paris, Sorbonne Université Presses.
- DANCEL B. (2001), *Un siècle de rédactions : écrits d'écoliers et de collégiens*, Grenoble, CRDP de Grenoble.
- DAVID J. (2008), « Les explications métagraphiques appliquées aux premières écritures enfantines », *Pratiques* 139-140, 163-187. [en ligne]
- DOQUET C. (2011), *L'écriture débutante : pratiques scripturales à l'école élémentaire*, Rennes, Presses Universitaires de Rennes.
- DOQUET C. (2019), « Le blanc comme outil de démarcation discursive et énonciative : quelques observations dans des écrits d'élèves », *Linguistique de l'écrit* 1. [en ligne]
- DOQUET C. & PILORGE J.-L. (2020), « La correction de copies au collège entre langue et discours : une catégorisation syntactico-énonciative », *Repères* 62, 191-213.
- ECO U. ([1979] 1985), *Lector in fabula : le rôle du lecteur ou la coopération interprétative dans les textes narratifs*, trad. fr. par Myriam Bouhazer, Paris, Grasset & Fasquelle.
- FABRE C. (1990), *Les brouillons d'écoliers ou l'entrée dans l'écriture*, Grenoble, Ceditel.
- FAVRIAUD M. (2014), *Le plurisystème ponctuationnel français à l'épreuve de la poésie contemporaine*, Limoges, Lambert-Lucas.
- FOURNIER J.-M. (1999), *La rédaction au collège : pratiques, normes, représentations*, Paris, INRP.
- GOODY J. (1979), *La raison graphique : la domestication de la pensée sauvage*, Paris, Les Éditions de Minuit.

- GRÉSILLON A. (1994), *Éléments de critique génétique*, Paris, Presses Universitaires de France.
- GRÉSILLON A. (1996), « Raturer, rater, rayer, éradiquer, radier, irradié », dans B. Rougé (éd.), *Ratures et repentirs*, Pau, Presses Universitaires de Pau et des Pays de l'Adour, 49-60.
- GRÉSILLON A. & LEBRAVE J.-L. (1982), « Les manuscrits comme lieu de conflits discursifs », dans C. Fuchs et alii (éds), *La genèse du texte : les modèles linguistiques*, Paris, Éditions du CNRS, 129-175.
- HANS-BIANCHI B. (2005), *La competenza scrittoria mediale: studi sulla scrittura popolare*, Tübingen, Max Niemeyer.
- HAY L. (éd.) (1996), *Genesis n° 10 : Sémiotique*, Paris, Jean-Michel Place.
- KLINKENBERG J.-M. & POLIS S. (2018), « De la scripturologie », *Signata* 9, 9-56. [en ligne]
- LAMOTHE-BORÉ C. (1999), *Choix énonciatifs dans la mise en mots de la fiction : le cas des brouillons scolaires*, Lille, Atelier National de Reproduction des Thèses.
- LEBRAVE J.-L. (1999), « Les manuscrits entre la substance de l'expression et la substance du contenu », dans C. Cortès & A. Rousseau (éds), *Catégories et connexions. En hommage à Jean Fourquet pour son centième anniversaire*, Villeneuve d'Ascq, Presses Universitaires du Septentrion, 371-378.
- LEBRAVE J.-L. (2000), « La nouvelle philologie et l'édition électronique des dossiers génétiques », *Arzanà* 5, 121-151.
- LEBRAVE J.-L. (2009), « Manuscrits de travail et linguistique de la production écrite », *Modèles linguistiques* 59, 13-21. [en ligne]
- LEFEBVRE J. (2007), *La note comme greffe typographique : étude linguistique et discursive*, Thèse de l'Université Paris 3.
- LEJEUNE P. (1998), *Les brouillons de soi*, Paris, Seuil.
- MOYSAN A. (en prép.), *La correction comme ressource pour les productions d'écrits au collège*, Thèse de l'Université Sorbonne Nouvelle.
- NOUGARET C. & PARINET É. (2015), *L'édition critique des textes contemporains (XIX^e-XX^e siècle)*, Paris, École nationale des chartes.
- PILORGÉ J.-L. (2010), « Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves », *Pratiques* 145-146, 85-103.
- REVELLI L. (2013), *Diacronia dell'italiano scolastico*, Roma, Aracne.
- SCHNEUWLY B. (2007), « Genres écrits et oraux et forme scolaire : enseignement et apprentissage de la langue première à l'école », dans C. Boré (éd.), *Construire et exploiter des corpus de genres scolaires*, Namur, Presses Universitaires de Namur, 13-26.