



HAL
open science

Le traitement de transcriptions d'écrits scolaires

Sara Mazziotti

► To cite this version:

Sara Mazziotti. Le traitement de transcriptions d'écrits scolaires. Colloque international des Etudiant×e×s chercheur×se×s en Didactique des langues et Linguistique, CEDIL'18, May 2018, Grenoble, France. ⟨hal-02648588⟩

HAL Id: hal-02648588

<https://hal.univ-grenoble-alpes.fr/hal-02648588v1>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-SA 4.0 - Attribution - Non-commercial use - ShareAlike - International License

LE TRAITEMENT DE TRANSCRIPTIONS D'ÉCRITS SCOLAIRES : UN PROTOCOLE D'ANNOTATIONS POUR ANALYSER LES INTERVENTIONS DES ENSEIGNANTS

Sara MAZZIOTTI^a

sara.mazziotti@studio.unibo.it

^aLaboratoire CLESTHIA, Paris, France

À l'heure de la digitalisation, la linguistique bénéficie autant que d'autres disciplines d'innovations de pointe qui facilitent et améliorent les études de corpus. Cependant, l'apport humain demeure essentiel notamment dans le cas de corpus d'écrits scolaires et se concrétise lors de l'étape de transcription, nécessaire pour en étudier ses spécificités. Dans notre corpus d'écrits d'élèves de CE2 et de CM2 italiens et français, l'application de balises qui identifient toutes les procédures d'écriture et les interventions et commentaires de l'enseignant, nous permet de passer du manuscrit au fichier exploitable par des logiciels de textométrie. Ainsi, il est possible de dessiner des profils d'« enseignants-correcteurs » grâce à l'extraction et le calcul automatique de toutes les interventions de l'enseignant. Le linguiste Jean-Luc Pilorgé (2010) avait distingué cinq « postures de correction » en analysant justement le type d'intervention de l'enseignant et l'aspect du texte de l'élève sur lequel elle porte. Nous nous baserons sur cette distinction pour explorer les pratiques de correction des enseignants italiens et français et pour en évaluer successivement la prise en compte de la part de l'élève lors d'une consigne de réécriture.

1. Constitution d'un corpus d'écrits scolaires français et italiens

À partir d'un corpus de productions écrites recueillies dans deux écoles italiennes et deux écoles françaises, nous nous sommes intéressés aux pratiques de correction des enseignants français et italiens en CE2 et en CM2 et à l'incidence des deux systèmes linguistiques sur la fréquence de certaines postures de correction. Pour cette étude expérimentale, nous avons soumis aux élèves participant au projet de recherche une consigne de rédaction commune, puis nous avons demandé aux enseignants de corriger les productions écrites selon leurs propres habitudes correctives. Ensuite, nous avons donné aux élèves la possibilité de relire leurs textes corrigés, quelques jours après l'avoir écrit ; et enfin d'en rédiger une version définitive. La consigne était « Que feras-tu quand tu seras grand ? Raconte une de tes journées » ; le temps à disposition pour la rédaction du brouillon et ensuite de la version définitive était d'une heure pour chacune des deux séances. Des métadonnées concernant la profession des parents et la langue maternelle des élèves, ainsi qu'un questionnaire écrit destiné aux enseignants ont été également recueillis. L'enseignant a toujours été présent dans la classe lors du recueil et il était autorisé à répondre aux doutes des élèves et/ou à les aider en cas de blocage. Les élèves avaient également accès aux dictionnaires et aux affiches dans la classe qui rappelaient des règles de conjugaison ou d'accord verbal. En résumé, nous disposons actuellement du brouillon et de la version définitive d'environ 250 élèves, en sachant que le recueil s'achèvera en décembre 2019.

Successivement, chaque version du texte a été transcrite et annotée selon le protocole de transcription utilisé dans le cadre du projet ECRISCOL (Doquet) qui se donne pour objectif de constituer une base de données ouverte en ligne d'un corpus scolaire et d'associer à chaque manuscrit sa transcription génétique présentée en version TEI et les métadonnées concernant l'école, le devoir et l'élève-scripteur. En tenant compte de l'aspect génétique du texte, de l'aspect temporel de chaque procédure d'écriture et d'une duplicité de voix, notamment celle

de l'élève et celle de l'enseignant-correcteur, les corpus scolaires présentent des spécificités que la transcription et l'annotation peuvent mettre en évidence. Bien que chacune de ces spécificités peut être observable par quiconque sur les copies constitutives de notre corpus, l'intérêt de l'outil de transcription réside dans la possibilité de faire ressortir et de mettre en exergue des informations telle que la construction du processus d'écriture de l'élève, l'organisation de son flux d'idées, l'impact des corrections apportées par l'enseignants chez l'élève et la diversité de ces corrections portant sur différents aspects de la langue et du texte. De plus, une transcription difficile, riche en enchâssements de procédures d'écriture et dans laquelle le taux d'intervention de l'enseignant est très élevé, peut suggérer un problème de lisibilité et donc un obstacle pour l'élève dans la phase de réécriture. Souvent, l'annotation même de ce type de manuscrit devient presque illisible en confirmant la présence excessive du stylo de l'enseignant : seule la conversion en version TEI, destinée à la lecture et à l'exploitation par les usagers, rend lisible ce type d'annotations, comme dans le cadre du projet ECRISCOL.

Dans notre étude, la transcription et l'analyse des productions écrites seront destinées en particulier à l'identification des profils de chaque enseignant-correcteur, afin de les comparer avec leurs témoignages et considérations personnelles et d'évaluer l'impact de chacune de ces postures de correction chez les élèves. N'étant qu'au début de l'analyse de notre corpus, nous considérerons ici, d'une part, l'importance accordée à la transcription et à l'élaboration d'un protocole très détaillé pour traiter un corpus scolaire, et, d'autre part, quelques exemples d'exploitation automatique des transcriptions.

2. Transcription et annotation des manuscrits : deux vrais outils d'analyse

Afin d'accéder au contenu des productions écrites des élèves et l'analyser automatiquement à l'aide des outils numériques de textométrie, il est nécessaire de transcrire le texte selon un protocole de transcription très précis. Plusieurs groupes de recherche en France qui s'intéressent aux écrits scolaires ont des exigences de transcription similaires et se confrontent avec les mêmes problèmes liés à la transcription des copies. Dans le cadre du projet E-CALM, financé par l'ANR et coordonné par Claire Doquet, quatre laboratoires (CIRCEFT-ESCOL, LIDILEM, CLESTHIA et CLLE-ERSS) sont en train de collaborer pour constituer, mettre à disposition de la communauté scientifique et étudier un vaste corpus d'écrits d'élèves et d'étudiants. L'idée fondamentale de cette coopération entre les différents laboratoires participant au projet est de rendre les transcriptions et les annotations les plus homogènes possible, afin que chaque laboratoire ait le même protocole, et ce, dans l'optique d'explorer ce corpus « dans des analyses sociologiquement contextualisées » et « d'étudier les modalités de l'écriture dans les avant-textes ». Notre corpus qui intégrera les données déjà collectées et disponibles en ligne dans le cadre du projet ECRISCOL respecte les objectifs suivants :

1. les quatre écoles participant au recueil proviennent de contextes sociologiques comparables ;
2. les productions écrites finales des élèves sont étudiées en comparaison avec le premier texte produit et corrigé par l'enseignant.

Mais quelles spécificités présente un corpus scolaire ?

Tout d'abord, une copie d'élève conserve les traces de la construction du processus d'écriture : tout ajout, rature, remplacement, déplacement et dessin est signalé dans nos transcriptions ; mais aussi elle conserve les traces de la co-construction de l'écriture, en particulier dans le cas des versions finales qui ont été précédées d'un brouillon sur lequel l'enseignant a apporté des corrections ou des commentaires. Il est donc nécessaire de signaler, d'une part, la multiplicité

des voix qui interviennent au sein de la copie et, d'autre part, la temporalité derrière l'action de ces acteurs.

Il ne serait pas inutile de préciser qu'il existe une distinction au sein de notre projet de recherche entre les fichiers texte appelés « transcriptions » et les fichiers texte appelés « annotations ». Le traitement automatique est mis en place à partir des annotations qui sont des transcriptions des manuscrits incluant une correction de toute forme non normée : si dans un premier temps, nous transcrivons le texte de l'élève tel qu'il apparaît dans le manuscrit ; dans un deuxième temps, nous y ajoutons la correction des erreurs d'orthographe et nous effectuons des modifications de certaines structures du protocole de référence, en préparation de l'exploitation à l'aide des outils de textométrie. Regardons l'extrait d'un manuscrit de CM2 (figure 1) et l'alignement de sa transcription et de son annotation :

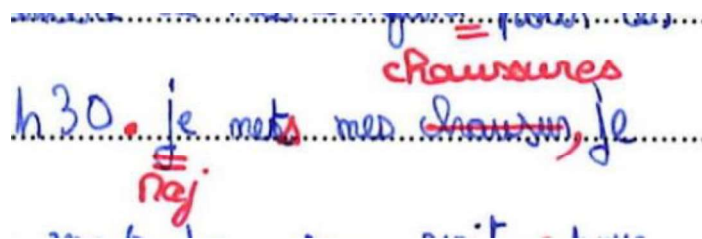


Figure 1 - Extrait d'un manuscrit français de CM2 corrigé.

Transcription	Annotation
{T2Pσ#j}{T2P#maj}e me[s]<t><T2P#s> mes [T2P#chausur]<T2p#chaussures> <T2P#,> je£	<{T2Pσ#j}{T2P#maj}e> <Je> me®[s]//t//®//T2P#s// mes ®[T2P#chausur]//T2P#chaussures//® //T2P#,,// je £

Tableau 1 - Comparaison entre le segment transcrit et le segment annoté de l'extrait de copie en figure 1

Les chevrons (< >), qui signalent dans la transcription les ajouts et donc aussi l'ajout à l'intérieur d'un remplacement, effectué ici d'abord par l'élève (remplacement du « s » par le « t ») et ensuite par l'enseignant (remplacement de « chausur » par « chaussures »), deviennent des slashes dans l'annotation. Dans cette dernière, nous délimitons la totalité de cette procédure avec ®, symbole qui indique le début et la fin du remplacement. De plus, toutes les erreurs d'orthographe et de majuscules (ici, nous avons « Je ») sont corrigées à l'aide des chevrons, toujours dans la phase d'annotation (<je>_<Je>). Le protocole de transcription des soulignements et des commentaires de l'enseignant, que nous expliquerons plus loin, ne varie pas lors du passage aux annotations : dans ce cas, nous situons le soulignement du « j » minuscule et du commentaire « maj » (« majuscule ») à l'intérieur du premier segment entre chevrons que nous avons corrigé avec « Je ». Le premier segment entre chevrons indique en effet le mot écrit par l'élève et le deuxième sa forme normée ajoutée par le transcripteur. La phase de passage des transcriptions aux annotations nous permettra de signaler et de corriger ainsi toute forme non normée dans les transcriptions et d'aboutir à un fichier qui sera converti en format TEI.

Dans la figure 2, un extrait d'une copie de niveau CE2 faisant partie de notre corpus italien. Ici, nous retrouvons en noir le premier jet de l'élève et en rouge les corrections de l'enseignant situées dans le texte, en marge et en bas de page.

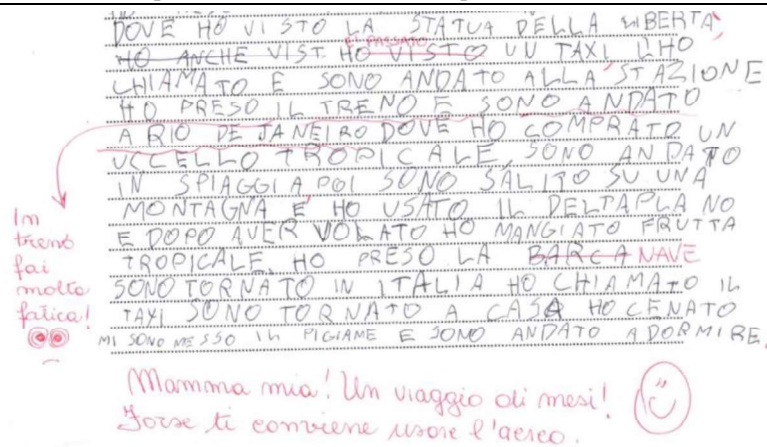


Figure 2 - Extrait d'un manuscrit de CE2 italien corrigé.

Dans ce cas, les interventions de l'enseignant sont très hétérogènes : on retrouve des corrections orthographiques (ajout de l'accent sur le « a » final de « libertà »), des soulèvements, des remplacements lexicaux (ex : [barca]<nave>), des commentaires sur le contenu (en marge, pour souligner qu'il est difficile d'aller à Rio De Janeiro en train), des appréciations globales (en bas de page, et qui insistent sur l'idée qu'il s'agit d'un long voyage et qu'il faudrait peut-être prendre l'avion) et des dessins (des smileys). En ce sens, la transcription n'est pas seulement un outil de constitution des données en vue du traitement automatique des textes, elle est aussi un moyen de préparation à la réflexion, un instrument pour rentrer dans le texte et adopter à la fois le point de vue de l'élève et celui de l'enseignant. En effet, nous avançons déjà des hypothèses concernant la lisibilité du texte après la correction de l'enseignant et nous remarquons les spécificités des commentaires de cet enseignant qui suggère implicitement à l'élève de faire attention à sa représentation du monde plus qu'à la forme. Il pourrait s'agir de la posture du « lecteur-naïf » définie par J-L. Pilorgé qui « commente l'action inventée par l'élève en référence au monde d'aujourd'hui » (Pilorgé, 2010). Cependant, nous avons besoin, pour cela, d'abord d'une analyse de toutes les interventions de l'enseignant sur la copie : c'est pour cela qu'il est important de distinguer dans notre protocole tout soulèvement, commentaire et remaniement direct par l'enseignant sur le discours de l'élève. Ainsi, nous aboutirons à des annotations qui sépareront ces trois types d'interventions et qui pourront être repérées automatiquement.

3. Des annotations au logiciel de textométrie

3.1. Les différents types d'interventions des enseignants à l'écrit

À partir du manuscrit (figure 2), nous allons proposer l'annotation et ensuite les trois types d'interventions de l'enseignant que nous pourrions rechercher, compter et classer automatiquement :

dove ho visto la statua della libert@[T2P#a]//T2P#à.//@£

[ho anche vist] @[T2P#ho visto]//T2P#è passato//@ un taxi //T2P#,,/l'ho£

chiamato e sono andato alla stazione£

ho preso {T2P#il treno e sono andato£

a Rio de Janeiro} {T2P#In treno fai molta fatica!}{T2P#DESSIN} dove ho comprato un£

uccello tropicale, sono andato£

in spiaggia //T2P#,,/poi sono salito su una£

montagna e ho usato il deltaplano£

e dopo aver volato ho mangiato frutta

tropicale. Ho preso la [T2P#barca]//T2P#nave//[R]£

sono tornato in Italia ho chiamato il

taxi sono tornato a casa[o]//a//[R] ho cenato£

mi sono messo il <pigiame>_<pigiama> *e sono andato a dormire* //T2P#.§

{T2P#Mamma mia! Un viaggio di mesi! Forse ti conviene usare l'aereo.}{T2P#DESSIN}§

- Ajout de l'accent sur le « a » final de « libertà » (ligne 1). Il s'agit d'un cas de remplacement du « a » sans accent avec le « a » accentué qui est annoté comme une suppression suivie d'un ajout et limité par ce symbole marquant le remplacement [R]. Dans notre protocole de transcription, les interventions de l'enseignant sont signalées avec un « P » précédé de l'indication temporelle « T2 » (temps 2, par rapport au premier jet de l'élève) :

libert[R][T2P#a]//T2P#à//[R]

- Soulignement d'une phrase (ligne 4 et 5). La phrase est limitée par des accolades et précédée toujours par la marque temporelle T2, l'indication de l'acteur P et une lettre grecque, le sigma :

{T2Pσ#ho preso il treno e sono andato a Rio de Janeiro}

- Commentaire en marge ou en bas de page. Tout ajout ou commentaire en marge se doit d'être restitué dans le texte, donc nous allons l'insérer juste après la phrase soulignée entre accolades, mais sans le petit sigma :

{T2P#In treno fai molta fatica!}

Pour conclure, avec cette dernière structure nous transcrivons aussi la présence d'un smiley, toujours dans l'extrait du manuscrit proposé dans la figure 1, (ou si présent aussi un tableau, un schéma ou tout autre dessin) qui dans ce cas apparaît juste après le commentaire, toujours en marge : {T2P#DESSIN}

Comme nous l'avons vu plus haut, l'extraction automatique de toutes les procédures d'écritures dans le brouillon et dans la version finale et les interventions des enseignants précédemment annotées dans nos transcriptions nous est utile lors du classement des différents profils de correction qui toutefois n'ont pas l'ambition d'associer un seul type d'intervention à chaque enseignant. Puisqu'au sein d'une même copie nous retrouvons plusieurs types de correction, il est nécessaire de les extraire automatiquement pour les calculer et les analyser. Ceci est possible grâce à un script que nous avons créé, mais que nous ne présenterons pas ici ou à l'aide du logiciel d'analyse textométrique *LeTrameur* conçu par Serge Fleury qui nous montre également le contexte gauche et droit. Cette dernière démarche se révèle très utile également pour investiguer le discours de l'élève.

En effet, le logiciel nous permet de sélectionner un mot-clé (prenons l'exemple d'un mot utilisé dans la consigne, comme « grande ») et de comparer la construction du texte chez les filles et chez les garçons, ou bien de la comparer chez les élèves français et chez les élèves italiens. Voici un exemple d'interface du concordancier de la version en ligne *iTrameur*, à partir du mot « grande » dans l'une des quatre classes italiennes de niveau CE2 (figure 3).

LeTrameur Analyse textométrique de données

Afficher items Recherche :

N°	Partie	Contexte Gauche	Pôle	Contexte Droit
1	EC-CE2-2017-FOR1-D1-E10-V1	a casa non (d)n(/d) stanca per la cena .] § § (d)Io da grande, vorrei esse(/d) § § Io da	grande	(i)T2#(/i) vorrei essere una pasticceria . E nel mio negozio , vorrei come
2	EC-CE2-2017-FOR1-D1-E11-V1	di più , ma di cosa farai in una giornata quando sarai	grande	} § / T3#Io da grande sarò l ' archeologa . § Oggi dovrò andare a scavare
3	EC-CE2-2017-FOR1-D1-E12-V1		sarò grande	farò la (an)(b)pittrice(/b)(a)pittrice(/a)(/an) perché mi piace (an)(b)diseg(/)T2=n(/)are(/b)(a)disegnare(/a)(/an) , (d)=XX#(/d) andrò { T2P#ha } (i)T2#a(/i)
4	EC-CE2-2017-FOR1-D1-E14-V1	racconto una giornata di quando sarò	grande	. § Il 12 luglio , il mio compleanno . § Io forse farò la scrittrice

Figure 3 - Capture d'écran de la concordance du mot « grande » dans la version en ligne du logiciel *LeTrameur*

Nous repérons facilement, dans le contexte droit, les premières idées que la consigne a suscitées chez l'élève, comment il les a organisées pour débiter son texte et quels temps verbaux il leur a associés.

3.2. *Le passage du brouillon à la V2 : la prise en compte des corrections*

L'étude des différents types d'intervention des enseignants ne se limite pas aux textes sur lesquels ces corrections ont été apportées, mais elle s'étend également à l'analyse de la version finale produite par l'élève. L'hypothèse de base de cette recherche considère les enseignants français comme étant davantage focalisés sur le langage, sur l'aspect orthographique, sur la forme qu'ils protègent et corrigent comme des « gardiens du code » (Pilorgé, 2010). En revanche, les enseignants italiens aborderaient plus rapidement dans la scolarité des élèves des aspects tels que la cohérence et corrigeraient plus souvent la construction du texte ou la vraisemblance du contenu. Nous avançons l'hypothèse que cela est dû initialement aux caractéristiques intrinsèques aux deux systèmes linguistiques : une plus forte correspondance graphie-phonie propre à la langue italienne accélère l'apprentissage de l'orthographe et permet aux enseignants de fournir des conseils finalisés à l'amélioration globale du texte initial et non seulement de la forme.

Pour vérifier cette hypothèse et surtout pour évaluer quelle est la posture de correction la plus propice dans le cadre d'un exercice de réécriture de ce type, nous allons comparer la transcription du brouillon corrigé et la transcription de la version définitive, à l'aide du logiciel *mkAlign* conçu par Serge Fleury. Il s'agit d'une première étape d'analyse des deux versions pour évaluer le taux de remaniement du brouillon : le logiciel crée un tableau qui signale en rouge tous les éléments supprimés lors du passage à la version finale et en vert tous les éléments que l'élève a ajoutés. Avec cette mise en comparaison des deux versions, il est possible de déterminer la prise en compte des corrections de la part de l'élève, en sachant que la consigne de réécriture leur donnait aussi la possibilité de ne pas réécrire le premier texte corrigé, mais d'en rédiger un nouveau s'ils le préféraient. Les élèves choisissent-ils de ne pas retravailler le brouillon et de proposer un nouveau texte quand le premier est surchargé de corrections et/ou assez illisible ? Essayent-ils, en revanche, de prendre en compte les corrections et les commentaires de l'enseignant quand ils portent plutôt sur la globalité du texte ?

4. Quelques résultats

Comme précédemment souligné, la transcription et ensuite l'annotation du corpus scolaire constituent une première phase importante de l'analyse de nos données. Nous en déduisons certains éléments qui pourraient être utiles au moment de l'exploitation automatique des annotations qui intervient dans un deuxième temps. Nous ne sommes pas encore en mesure de proposer des chiffres exacts concernant les pratiques de correction des enseignants dans notre corpus, mais nous remarquons cependant une insistance des enseignants français sur l'aspect orthographique et grammatical de la langue ainsi qu'une surcharge de signes dans la façon de corriger les productions écrites de leurs élèves. En outre, les corrections orthographiques sont moins fréquentes en contexte italien car, à parité de niveau, le système linguistique fait en sorte qu'il y ait moins d'erreurs d'orthographe. Nous avons choisi de ne proposer ici que deux extraits de copies corrigées, l'une italienne (figure 1) et l'autre française (figure 4), dans le but de montrer l'hétérogénéité des pratiques de correction. La figure 4 présente un extrait de texte, écrit par un élève de CE2 d'une des deux écoles françaises participant au projet, qui met en évidence la multiplicité de traits, de signes et de couleurs au sein d'une même copie et qui devraient donner des pistes à l'élève en vue de la réécriture du brouillon.

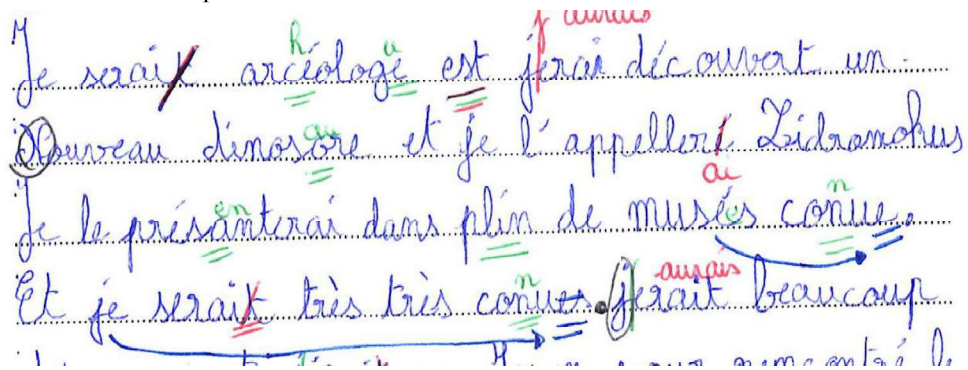


Figure 4 - extrait d'un manuscrit d'un élève de CE2 corrigé.

Si nous comparons ce manuscrit au manuscrit italien de la figure 1, nous remarquons une intervention massive de la part de l'enseignant qui se sert de différentes couleurs pour suggérer à l'élève sur quel aspect porte l'erreur (lexical, grammatical, manque de majuscule). Nous signalons que dans une étude précédente à cette recherche menée dans deux classes françaises de CM2, nous avons constaté qu'une surcharge d'interventions de l'enseignant inhibait le travail de réécriture de l'élève, lequel procédait à une mise au propre du brouillon corrigé en essayant de recopier toutes les corrections du maître. Même si une telle fréquence d'interventions n'a pas été retrouvée dans notre corpus italien, nous avons repéré toutefois la posture de « gardien de code » chez les enseignants italiens également et dans tous les deux niveaux scolaires analysés, avec des nuances différentes que nous allons rapidement présenter maintenant.

Premièrement, les quelques cas de remplacements lexicaux que l'on a retrouvés dans le manuscrit italien (figure 1) sont en réalité très fréquents et imputables à une volonté des enseignants de faire varier le plus possible le vocabulaire dans le texte de l'élève. Il ne s'agit pas uniquement de fournir des synonymes pour éviter les répétitions : pourrions-nous avancer l'hypothèse d'une véritable tendance que les enseignants italiens font émerger dans leur pratique de correction ? Dans la plupart des cas, ils remplacent le mot choisi par l'élève avec un mot appartenant à un registre plus soutenu qui se distingue nettement du discours habituel de l'élève. Cette tendance avait été déjà signalée dans des études linguistiques et didactiques des années 70 (Berretta, 1977) et reconfirmée trente ans après par le linguiste A. De Benedetti

(2009). Ce dernier insiste sur l'idée que l'instruction s'est vue confier le rôle de contraster les changements dans la langue et de mettre en place un mouvement de retour en arrière, vers le prestige et la pureté originelle de la langue italienne. Le traitement automatique de l'ensemble des annotations à l'aide du logiciel *LeTrameur* nous aidera donc à confirmer cette tendance visible chez les enseignants italiens et à considérer sa présence dans le corpus français également.

L'hétérogénéité de nos écrits scolaires et la variété des styles de correction des enseignants nous empêchent d'investiguer de manière plus détaillée ici notre corpus, notre analyse n'étant d'ailleurs qu'au début. Nous concluons en mettant l'accent sur quelques résultats qui seront successivement accompagnés de pourcentages exacts : tout d'abord, nous constatons une tendance générale de l'enseignant à apporter des corrections locales, qui portent souvent sur un mot précis et sur un aspect linguistique et normatif, indépendamment de la « posture de correction » que l'on lui attribuera. De plus, à travers la phase de transcription et d'annotation, nous remarquons que les élèves italiens ont plus de facilité à modifier la structure et le contenu dans la version finale, malgré la présence de corrections de l'enseignant dans le premier jet. Cela pourrait être lié au fait que les élèves italiens soient moins inhibés par l'idée de commettre de nouvelles erreurs orthographiques. Le système linguistique joue un rôle important dans l'approche de l'erreur : l'erreur d'orthographe est considérée dans les deux contextes linguistiques comme étant plus grave par rapport aux répétitions dans la langue. Cependant, les difficultés orthographiques imposées par la langue écrite en français provoqueraient chez les élèves français un attachement au premier texte qui les rassure au moment du passage à la version finale et qui présente déjà un grand nombre de corrections, de suggestions et de commentaires de l'enseignant. Ce que l'on constate est une tendance à la mise au propre ou à l'abandon du premier jet, en faveur de l'écriture d'un nouveau texte entièrement différent, (peut-être dû aussi au fait que notre protocole de recueil ne prévoyait pas une deuxième correction du maître). En revanche, nous faisons l'hypothèse que les élèves italiens remanieraient davantage leur premier texte et qu'ils rajouteraient plus facilement du contenu : cela ne doit pas nous faire croire pourtant qu'ils aboutissent toujours à un texte de meilleure qualité dans la version définitive. Nous n'avons pas suivi les enseignants dans leur travail de réécriture avec les élèves pendant l'année scolaire, mais pouvoir investiguer la représentation de l'écriture des enseignants et ce qu'est une rédaction aboutie pour ces enseignants en particulier nous aiderait à éclaircir ce dernier point.

Références bibliographiques

- BERRETTA, Monica (1977). *Linguistica ed educazione linguistica. Guida all'insegnamento dell'italiano*, Torino : Piccola Biblioteca Einaudi.
- BORE, Catherine (2007). Construire et exploiter des corpus de genres scolaires, Namur (Belgique) : Presses Universitaires de Namur, collection Diptyque.
- DAVID, Jacques & DOQUET, Claire (2016). Les écrits d'élèves : un corpus de référence pour le français contemporain. Congrès Mondial de Linguistique Française, Tours, France. hal-01659765f.
- DAVID, Jacques & DOQUET, Claire (2018). Collecter, interpréter, enseigner l'écriture », Repères [En ligne], 57 | 2018, mis en ligne le 30 juin 2018, consulté le 18 mai 2019. URL : <http://journals.openedition.org/reperes/1444>.
- DAVID, Jacques, DOQUET, Claire & RINCK, Fanny (2018). « *Que faire avec les écrits des élèves ?* », *Le français aujourd'hui*, n°203. <http://www.revues.armand-colin.com/lettres-langue/francais-aujourd'hui>.
- DE BENEDETTI, Andrea (2009). *Val piu la pratica. Piccola grammatica immorale della lingua italiana*, Bari : Editori Laterza.
- DOQUET, Claire (2011). L'Écriture débutante. Pratiques scripturales à l'école élémentaire. Rennes, PUR.

DOQUET, Claire, ENOIU, Vanda, FLEURY, Serge & MAZZIOTTI, Sara. (2017) Problèmes posés par la transcription et l'annotation d'écrits d'élèves. *Corpus*, Bases, Corpus, Langage - UMR 7320, Spécificités et contraintes des grands corpus de textes scolaire : problèmes de transcription, d'annotation et de traitement. Corpus. hal-01659752.

PILORGE, Jean-Luc (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 : pages 85-103.

WOLFARTH, Claire (2015). Apport du TAL à la constitution et l'exploitation d'un corpus scolaire de cours préparatoire. Linguistique. 2015. dumas-01167286.

Références aux logiciels

FLEURY, Serge, Le Trameur, Téléchargement et description, <http://www.tal.univ-paris3.fr/trameur/>

FLEURY, Serge. iTrameur <http://www.tal.univ-paris3.fr/trameur/iTrameur/>

FLEURY, Serge, mKAlign, Téléchargement et description, <http://www.tal.univ-paris3.fr/mkAlign/>

Références aux sites Internet

Projet E-CALM, ANR, <http://e-calm.huma-num.fr/le-projet/>

Projet ECRISCOL, Sorbonne Nouvelle-Paris 3, <http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/>