Annotation et exploitation numérique d'un corpus d'écrits scolaires corrigés en France et en Italie

Sara Mazziotti

Université Sorbonne Nouvelle – Paris 3, ED 268, CLESTHIA, 4 rue des Irlandais Paris, France
sara.mazziotti@studio.unibo.it

Résumé. À partir d'un corpus de textes d'élèves de CE2 et de CM2 recueillis en France et en Italie, nous investiguerons les pratiques de correction des enseignants qui se confrontent avec deux systèmes linguistiques ayant un taux de correspondance entre graphème et phonème très différent. La numérisation de ces écrits, prenant en compte l'aspect génétique du texte, rend possible, d'une part, l'évaluation du taux de remaniement du brouillon de la part de l'élève et, donc, la prise en compte des différentes corrections apportées par l'enseignant lors de l'étape de réécriture. D'autre part, elle permet le repérage et la classification des modifications et des commentaires des enseignants apportés dans le brouillon. Sur quels aspects linguistiques en particulier se focalisent-ils ? Et quel est le taux de prise en compte des corrections par l'élève lors de l'étape de réécriture ?

Mots-clés. Ecrits scolaires, pratiques de correction, génétique textuelle, transcription, réécriture.

Abstract. Based on a corpus of texts from third and fifth graders collected in France and Italy, we will investigate the correction practices of teachers who are confronted with two linguistic systems with a very different correspondence rate between grapheme and phoneme. The digitization of these writings, taking into account the genetic aspect of the text, makes it possible, on the one hand, to evaluate the student's rate of reworking the draft and, therefore, to take into account the various corrections made by the teacher during the rewriting stage. On the other hand, it allows the identification and classification of changes and teacher comments made in the draft. Which linguistic aspects do they focus on? And what is the rate at which corrections are considered by the student during the rewriting stage?

Keywords. School writings, correction practices, textual genetics, transcription, rewriteing

1 Introduction

Cette étude est le résultat d'une série de questionnements de recherche que nous nous sommes posés pendant une collaboration avec l'équipe ECRISCOL, dirigée par Claire Doquet à l'Université Sorbonne Nouvelle – Paris 3. Le projet ECRISCOL a l'objectif de créer et de mettre à disposition des chercheurs et des enseignants une base de

données accessible en ligne constituée de textes d'élèves de l'école primaire au niveau universitaire. À chaque production écrite est associée une transcription en version TEI qui met en évidence tous les éléments supprimés, ajoutés, déplacés par l'élève ou par l'enseignant. À partir du même protocole de transcription, nous nous intéresserons dans notre thèse aux corrections des enseignants en particulier, afin de dessiner plusieurs « postures de correction »[1] de maitres et maitresses de CE2 et de CM2 en France et en Italie. Notre point de départ est le classement avancé par Jean-Luc Pilorgé qui distingue cinq postures de correction sur la base du type d'intervention de l'enseignant : le « gardien du code » s'intéresse principalement à la grammaire et à l'orthographe ; le « lecteur naïf » commente la « représentation du monde » de l'élève ; la posture de « stimulus réponse » vérifie le « respect de la tâche à accomplir » et le « réinvestissement de savoirs requis par le sujet » ; l'éditeur s'appuie sur le « déjà-écrit » et vise à « l'amélioration du texte » ; le critique commente le texte perçu comme une « construction, un objet de réflexion ». Dans la présentation des exploitations automatiques des transcriptions à l'aide du logiciel de textométrie MkALign[2] et de scripts informatiques conçus pour cette étude, nous nous focaliserons en particulier sur les postures de « gardien du code » et de « lecteur naïf ».

2 Hypothèse et méthodologie

Notre hypothèse est qu'un système linguistique avec une correspondance très faible entre graphème et phonème, comme dans le cas de la langue française, amène les enseignants à corriger plus particulièrement l'aspect orthographique du texte. En revanche, un système linguistique plus régulier, comme dans le cas de l'italien, permet aux enseignants de retrouver moins d'erreurs orthographiques dans les copies et, par conséquent, de proposer plus fréquemment des commentaires qui portent sur la cohérence d'un texte ou sur le contenu. Nous avons proposé dans huit classes de CE2 et de CM2 italiennes et françaises la consigne d'écriture : « Que feras-tu quand tu seras grand ? Raconte une de tes journées ». Les élèves devaient rédiger pendant une heure un premier texte d'une façon individuelle et, après une semaine environ, ils devaient relire le texte corrigé par leur enseignant pour en proposer une version définitive. Les deux versions ont été ensuite transcrites, exploitées automatiquement et relationnées aux réponses fournies par les enseignants à un questionnaire investiguant leurs pratiques de correction. Nous avons recueilli également une série de métadonnées concernant les élèves (langue maternelle, lieu de naissance, catégorie socio-professionnelle des parents, etc.), à travers une fiche que les parents devaient signer pour autoriser l'exploitation des copies anonymisées pour des fins de recherche.

[1] Jean-Luc Pilorgé, « Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves ». In : Pratiques, n° 145-146, 2010

[2] Serge Fleury, Paris 3

2.1 La transcription : un véritable outil d'analyse

La transcription, qui dans le cas de notre recherche est assez chronophage, puisqu'elle n'est pas automatique, constitue en réalité la première phase de notre analyse. Elle nous permet d'évaluer le nombre d'interventions de chaque enseignant et de repérer ses habitudes de correction qui seront ensuite vérifiées pendant l'analyse informatique. Pendant la transcription de la version définitive de l'élève, nous évaluons également le taux de variation du texte lors du passage à la réécriture : en général, un brouillon trop massivement corrigé amène l'élève à ne pas retravailler le contenu, mais à effectuer une mise au propre orthographique de la première version ou à l'abandonner pour en rédiger une différente. Dans la figure 1, nous proposons une mise en comparaison de deux extraits de copies de corrigées, tirées de notre corpus :

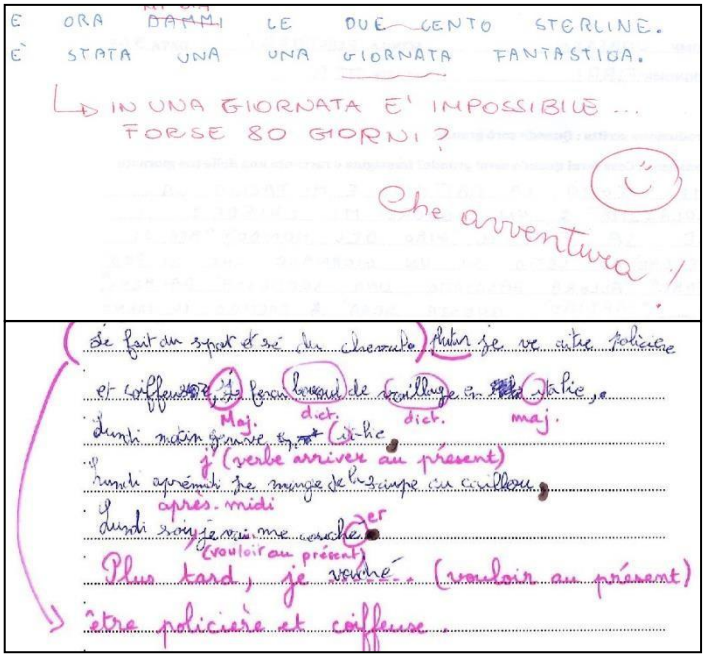

Fig.1 : En haut, un extrait de correction d'un enseignant italien de CE2 qui interpelle la représentation du monde de l'élève et qui adopte la posture de « lecteur naïf » pour commenter « En une seule journée il est impossible…peut-être en 80 jours ? ». L'élève avait raconté une série d'aventures et de voyages qu'il n'aurait pas pu effectuer en une seule journée. En revanche, en bas, un exemple de correction massive apportée par un enseignant français de CM2. Nous signalons la présence d'abréviations qui indiquent le type d'erreurs à l'élève (Maj : majuscule ; dict : dictionnaire...).

Il ne serait pas inutile de présenter alors brièvement les procédures d'écriture les plus fréquentes dans notre protocole, utilisé par le groupe de recherche ECRISCOL également, et que nous signalons avec les balises suivantes :

- <chevrons> si l'élève ou l'enseignant ajoute du contenu (ponctuation, lettre, mot, phrase,...) ;
- [crochets] si l'élève ou l'enseignant supprime du contenu ;
- [crochets]<chevrons> si l'élève ou l'enseignant supprime un premier élément pour le remplacer avec un nouvel élément (ex : [hier]<demain> si "hier" a été remplacé par "demain") ;
- {T2P#accolades} si l'enseignant ajoute un commentaire. "T2P" signale l'intervention de l'enseignant dans un temps successif au premier jet de l'élève et il apparaitra alors indépendamment de la procédure d'écriture qu'il applique ;
- {T2Pσ#accolades et sigma} si l'enseignant souligne un élément (ex : {T2Pσ#magazin} si l'enseignant souligne le mot pour indiquer qu'il y a une erreur d'orthographe.

3 L'analyse automatique des transcriptions

3.1 Le taux d'intervention des enseignants

Ne pouvant pas suivre les enseignants participant au projet pendant toute l'année scolaire, nous avons focalisé notre attention sur un devoir précis, notamment une consigne de réécriture, afin d'investiguer les différentes manières d'intervenir dans le texte d'élève qui devra être réélaboré individuellement. Comme nous l'avons précisé, toutes ces interventions sont signalées dans les transcriptions par le biais de balises qui peuvent donc être calculées automatiquement ou extraites et copiées dans un nouveau fichier. Le script que nous présenterons ici calcule le nombre d'interventions total de l'enseignant et le nombre exact de suppressions, ajouts, remplacements, soulignements et commentaires. Il fonctionne à partir d'un fichier txt qui rassemble l'intégralité des transcriptions des brouillons de ses élèves et il génère automatiquement un nouveau fichier avec les résultats de chaque catégorie (Fig.2). Nous avons ensuite calculé le nombre moyen d'interventions de chaque enseignant, en prenant en considérations le nombre de copies corrigées, afin de diviser de manière équilibrée les enseignants italiens et français en trois niveaux d'intervention : bas, moyen et élevé.

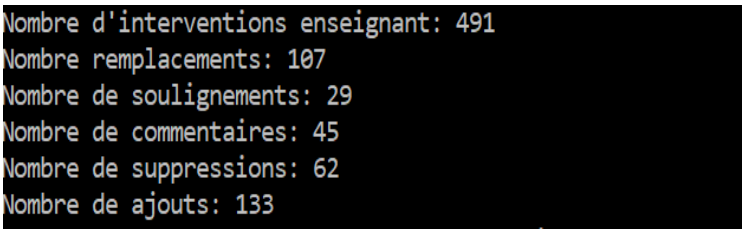

Fig.2 : Cet enseignant italien de CE2 a apporté 491 corrections totales sur 21 copies. Étant dans notre corpus 5,6 le nombre moyen par copie le plus bas (chez un enseignant italien) et 48,2 le plus élevé (chez un enseignant français), le degré d'intervention de cet enseignant est moyen.

Si la plupart des corrections sont locales, donc plus probablement associées à un mot bien précis, alors il est très probable que les enseignants avec un degré d'intervention élevé soient des « gardiens du code », en particulier, ceux qui n'apportent presque pas de commentaires. Nous analyserons successivement les différents types de commentaires chez les enseignants, puisque dans nos transcriptions les abréviations du type « orth, gr, temps », par exemple, et les commentaires sur le contenu sont transcrits de la même façon à l'aide des balises {T2P#accolades}.

3.2 La prise en compte des corrections chez l'élève

Afin d'évaluer le degré de réécriture du texte et d'étudier le type de modifications apportées par l'élève lors du passage du brouillon à la version finale, nous avons utilisé le logiciel MkAlign. Les élèves pouvaient en fait choisir de recopier le texte tel quel, le modifier en partie ou entièrement ou en rédiger un autre complètement différent de la première version. La mise en comparaison des deux textes nous permet de distinguer les éléments ajoutés, supprimés ou modifiés lors du passage à la version définitive et de relationner ces données aux postures de correction des enseignants (Fig. 3).

Fig. 3 : En vert, les éléments ajoutés ans la version définitive ; en rouge les éléments supprimés qui étaient présents dans le brouillon et en bleu les éléments déplacés ou modifiés.

Pour effectuer cette analyse, nous avons d'abord généré un script qui pouvait enlever automatiquement toutes les balises signalant les différentes procédures d'écriture et les corrections des enseignants. Le logiciel aurait sinon signalé en rouge presque tout le texte, puisque les corrections de l'enseignant étaient présentes uniquement dans le brouillon.

4 Conclusions et quelques résultats

À partir de notre classement des enseignants en trois niveaux selon leur degré d'intervention dans la copie, nous constatons que presque la totalité des enseignants français de CE2 et de CM2 applique un nombre élevé d'interventions ; en revanche, la plupart des enseignants italiens a un degré d'intervention bas. Cependant, le nombre

d'interventions locales est, dans les deux cas, plus élevé que les commentaires généraux en-tête, en marge ou en bas de page. Nous avançons alors l'hypothèse que la tendance à corriger surtout au niveau local ne dépende pas de la complexité du système linguistique : même les enseignants-correcteurs italiens se focalisent davantage sur l'aspect formel du texte. Cette volonté de signaler un maximum d'erreurs est-elle due à la perception de la tache de correction elle-même ? Et le degré d'intervention plus bas chez les enseignants italiens est-il motivé uniquement du fait qu'il y ait moins d'erreurs d'orthographe dans les copies ?

Une ultérieure analyse du type de commentaires proposés dans les deux contextes linguistiques nous permettra de répondre à cette dernière question : un nouveau script lancé à partir des transcriptions des brouillons corrigés génère un fichier dans lequel il propose une liste des commentaires de chaque enseignant. Portent-ils sur l'orthographe ou sur le contenu ? Quelle posture pourrions-nous alors attribuer à ces enseignants : toujours "gardien du code" ou "lecteur naïf » ?

Chez un enseignant italien de CE2, nous avons constaté que les deux postures peuvent co-exister au sein de la même copie et d'une façon assez équilibrée. En effet, il essayait de repérer un maximum d'erreurs d'orthographe, mais de proposer aussi en bas de page toujours un commentaire positif sur le contenu du texte, probablement pour communiquer à l'élève qu'il avait lu et apprécié ses idées (Fig. 1). Un « vrai lecteur » plutôt qu'un « lecteur naïf » dans ce cas-là : un « enseignant-lecteur » qui veut s'assurer que ses élèves aient compris qu'il a lu leurs textes.

De toute manière, cette duplicité de postures amène les élèves à ne pas abandonner leur premier texte et à le retravailler. Nous confirmons en effet dans notre étude aussi qu'un nombre très élevé de corrections inhibe la manipulation du texte de la part de l'élève qui passe à une mise au propre du brouillon ou à la rédaction d'un texte *ex novo*. Dans le cas de cet enseignant italien, l'équilibre entre ces deux postures permet aux élèves de rectifier les erreurs, mais au même temps de remanier le contenu. Notre objectif étant d'évaluer les pratiques de correction qui favorisaient davantage la réélaboration de la première version, dans le but d'aboutir à un texte de meilleure qualité.

Références

1. David, J., Guerrouache, C.,: Relire, analyser et réécrire ses écrits au début du primaire, Le français d'aujourd'hui, 29-49 (2018)
2. Doquet, C.: L'Écriture débutante. Pratiques scripturales à l'école élémentaire, Presses universitaires de Rennes, Rennes (2011)
3. Doquet, C., J. David, and S. F. : (Eds). Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement. In Corpus [Online], volume 16 (Special Issue). OpenEdition (2017)
4. Doquet, C., Enoiu, V., Fleury, S. et Mazziotti, S. : Problèmes posés par la transcription et l'annotation d'écrits d'élèves, Corpus [En ligne], 16 | 2017, URL : http://journals.openedition.org/corpus/2776

5. Fabre, C.: Les brouillons d'écoliers ou l'entrée dans l'écriture, Ceditel/L'atelier du texte, Grenoble (1990)
6. Fayol, M., Gombert, J.E., : Le retour de l'auteur sur son texte : Bilan provisoire des recherches psycholinguistiques », Repères, n° 73 (1987)
7. Pilorgé, J-L.: Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves. Thèse de doctorat sous la direction d'Annie Rouxel, Rennes, (2008)
8. Projet ECRISCOL: http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/
9. Logiciel MkAlign (Serge Fleury): http://www.tal.univ-paris3.fr/mkAlign/
10. Logiciel LeTrameur (Serge Fleury): http://www.tal.univ-paris3.fr/trameur/