

Problèmes posés par la transcription et l'annotation d'écrits d'élèves

Claire Doquet, Vanda Enoiu, Sara Mazziotti
ECRISCOL, Clesthia, Paris 3 – Sorbonne
Nouvelle

**8^e Journées Internationales de
Linguistique de Corpus
02 septembre 2015**





I. Contexte

- Rôle dans l'équipe
- Présentation du groupe de recherche
- Objectifs
- Inscription de la démarche dans la génétique textuelle
- Méthodologie
- Caractéristiques du corpus

- ## II. Cas particuliers
- transcription
 - annotation



I. ECRISCOL

Objectifs:

- Constituer un corpus de référence qui permet de mettre en évidence les spécificités de l'écrit en contexte scolaire
- Le mettre à disposition des autres chercheurs au fur et à mesure de sa constitution (création d'une plateforme d'accès *OpenAccess* → *passage sur Ortolang*)
- Observer quelles sont les compétences des élèves en fonction de différentes variables
- Permettre l'analyse à l'aide des outils textométriques (*LeTrameur*)



Pourquoi les *écrits scolaires*?

- Quels sont les problèmes spécifiques aux écrits scolaires?
- Comment les traiter dans les transcriptions et les annotations?
- Dans quelle mesure la textométrie peut nous aider à investiguer ces spécificités?

Écrits scolaires → Transcription-annotation → Textométrie



de se pendre au ~~arbre~~ avec elle. Et après et Charles
et Régine ~~ou~~ ^{vont} ~~font~~ dans le ~~parc~~ kiosque avec la chienne
victoria et le chien albert ~~dans~~ il vont danser.
pendant ce temps le papa de Régine ~~demande~~ ^{demande} en
mariage la maman de Charles.

de trouver un endroit pour discuter dans
tout le parc ils trouvent une vieille maison
abandonnée. Ensuite Charles décide de jouer
à cache cache avec Victoria et
Albert à la fin de la partie tout le
monde fait une pause la maman de
Charles dit
Et vient Charles on rentre à la maison?
Et Charles dit au revoir à Régine et
il s'en va.



De joie à se balancer et à monter dans les arbres
et Charles ~~de~~ ^{Pie} voulut lui montrer comment il
se débrouillait pour monter dans les arbres et ils
jouèrent toute la journée au héros et la mère de Charles
l'appela ^{l'appel} et il retourna chez lui et Prégisse fit
la même chose et Prégisse prépara une tasse de thé
pour son père. ^{elle}



Un cadre théorique

- Génétique textuelle
 - Manuscrits d'écrivains (ITEM, Lebrave et Grésillon)
 - Écriture comme processus : **transcription diplomatique**
- Analyse linguistique de l'écriture ancrée dans la linguistique de l'énonciation
- Prise en compte des opérations survenues au cours de l'écriture repertoriés par l'ITEM:
 - « -suppression = $A \rightarrow 0$
 - ajout = $0 \rightarrow A$
 - substitution = $A \rightarrow A'$
 - déplacement = $ABCDE \rightarrow BCDAE$ » (Grésillon & Lebrave, 1982 : 136)



tps. lex

tps

tps. ac. tps

tps.

gr. gr. ac

tps

lex.

lequel?

lex.

tps

ac

? Terminaison

ac.

Ils ^{vont} alla voir ! Et c'était un monstre tout ^{sub} salle
 il casser tout - il mangé tout. Ça veut dire que
 il ~~il n'y avait~~ plus rien. Les enfant alla prévenir
 leur parents les parents ^{arrivent} avertir de monstre est
partit décort les parents ont ent que c'était les
enfants. Les parents m étaient leurs enfants ^{étaient} dans leur
 chambre et les punir. Le monstre mange les
parents et les enfants appelle leur mami. La mami
 tue le monstre et les enfants ^{des} prend des gigants
ciseaux et coupe le centre du monstre et la
mami prend les parents et les gigants.
 Les parents se réveillent et font de gas calins
 au enfant et maintenant à chaque fois qu'il y a
 un monstre ^{qui} il les croie ! Tout le monde va mieux
 et les autres personne sont vivants.
 fin

bras, songeant à son père, à chaque corps
~~enterrés~~ mis sous terre. ^{et regardant} ~~et voyant~~ avec le fil de
jour ce que le ciel y donnait. ¶

Aussi, Il avait ^{mis} ~~mis~~ les deux alliances dans une chaîne et
accroché ^{de} celle-ci autour de son cou. C'était sa
façon de garder un peu de sa famille avec lui.

^{Cependant} ~~De plus~~, malgré la mort de ^{ce} dernier ~~de sa famille~~
et // horrible sentiments d'abandon, il restait optimiste
face à la difficulté de la vie. Attendant de voir
ce qu'allait lui donner le ciel.



État actuel du corpus

- 935 copies + brouillons correspondants
- Niveaux : CE1, CM2, 3ème-2^{nde}, entrée université
- Production : suite donnée à un texte littéraire



Étapes

- Recueil des copies de différents niveaux (primaire, collège, lycée, universitaire)
- Transcription (à l'aide d'un logiciel de traitement de texte)
- Annotation (préparation pour un logiciel d'analyse textuelle) – correction orthographique
 - deux niveaux de traitement distincts
 - assurer le passage d'une étape à l'autre



Défis méthodologiques

- Créer des conventions adaptées au type d'écrit que nous transcrivons
- Fidélité par rapport à la copie d'origine – transcription diplomatique
- Résoudre le problème de la subjectivité du transcripteur
- Pour les deux niveaux: mettre en évidence les opérations effectuées par le scripteur
- Choix de recodage
- Créer une version lisible par les logiciels de textométrie

Génétique textuelle et écrits scolaires

Génétique
textuelle

Textométrie

Opérations
scripturales

Conventions
de
transcription
/annotation

Analyses et
résultats



Métadonnées

- Élève: prénom, n° dans la classe, date de naissance , âge, classe, sexe, parcours scolaire (redoublement, changement d'école), statut particulier (allophone, handicap, accompagnement), langues personnelles, parlée(s), lu(es), écrite(s), CSP des parents
- Établissement: urbain, rural, ZEP, département, banlieue, taille établissement, privé, public
- Enseignant: M/F, âge, ancienneté, titres, formation
- Tâche: consigne, temps d'écriture, contexte, outils d'écriture
- Texte: longueur, indications de l'enseignant, commentaires de l'enseignant, gloses de l'élève, soulignement, annotations

Identification du texte: EC-CM2-2015-SDPB-D1-E23-V1



II. Transcription et annotation

- Défis:

- Réussir à annoter des occurrences spécifiques aux écrits scolaires
- Créer des conventions de transcription et d'annotation qui reflètent les opérations
- Rendre accessible la lecture des caractéristiques linguistiques des écrits scolaires aux logiciels de textométrie
- Changement et adaptation des conventions au fur et à mesure de l'activité
- Activité interprétative du transcripateur: l'interprétation d'un choix orthographique, d'une opération postérieure et de la restitution de la chronologie scripturale



Conventions de base

Transcription

<> - ajouts

[] – suppressions

#XXX# - segments illisibles

§ - retour volontaire à la ligne

μ - intervention de l'enseignant (autre que soulignement)

σ – soulignement de la part de l'enseignant

Annotation

® – remplacement quand il y a une équivalence syntaxique

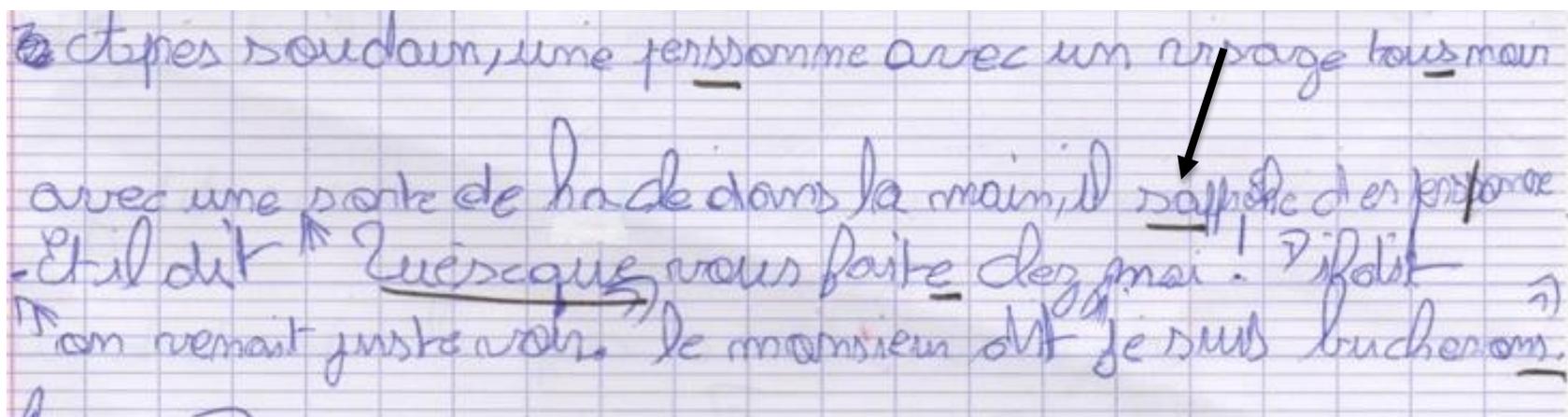
< >- correction orthographique

// // - ajout





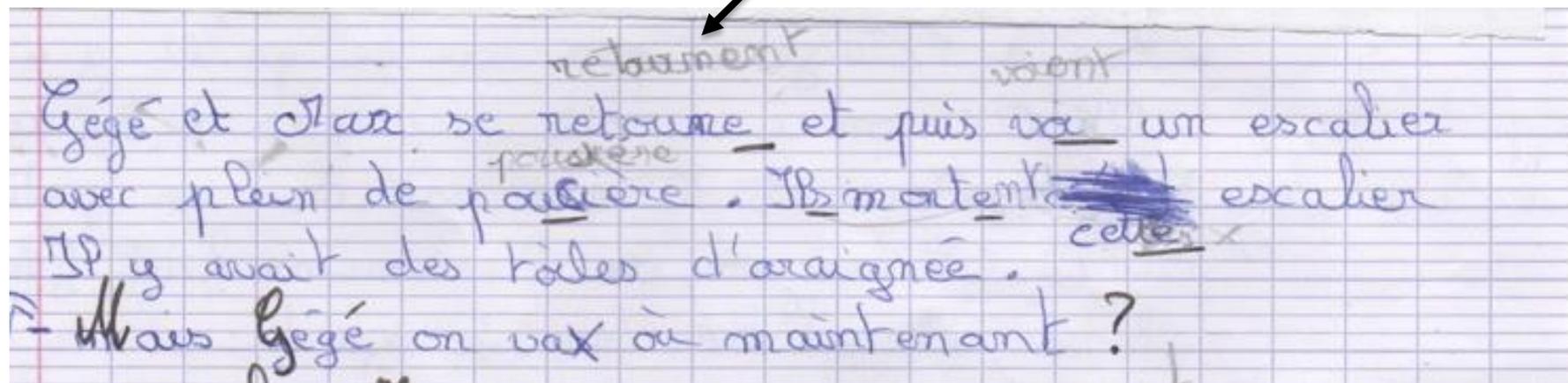
- La présence des interventions de l'enseignant qui influencent la rédaction de la version définitive :



Transcription: {Pσ#sa}pproche
EC-CM2-2015-SDPB-D1-E21-V1



- Dans le cas où l'élève corrige après soulignement
- Marquage de différentes interventions (élève, enseignant)
- Et du temps de l'écriture



Transcription: retourn {Pσ#e} <T2#retournent>
EC-CM2-2015-SDPB-D1-E21-V1



- Plusieurs scripteurs: corrections de l'enseignant
- Transcrites seulement quand il y a une version ultérieure du texte (une V2)

Mais gégé on va où maintenant ?
- ... euh je ne sais pas
- Bravo tu voulais y allais et maintenant on y est !!!
- ~~Allez on y va c'est bon j'ai trouvée~~
une sortie.
- Mais où sont les VTT ? à l'éloge ?
- C'est étrange sa. Je crois on va repartir par où on est venue.

V1

22 jkt L'oncle Jo a disparu lui appelle l'oncle Jo plusieurs fois mais il ne répond pas. Lui appelle lui Denise de répond comme d'habitude ou mais l'oncle Jo n'est pas lui. Lui appelle Denise

V2



Membre fondateur Sorbonne Paris Cité



- Annotation
- Décider s'il s'agit d'un remplacement et le noter comme tel:
®[mot_raturé]//mot_replacé//®
- Choix du récodage:
 - orthographique
 - ponctuation
 - majuscules



- Le respect des segmentations des mots entre sous-segmentation ou hyper-segmentation influencées ou non par l'oral:

Il n'était pas granchasse dit Gégé c'est un animal
et puis Mass dit: ^{vient} vient on va visiter la maison
puis un autre craquement ^{vient} venant de l'étage Mass et
Gégé dit en même temps là!!! ~~on va voir~~
~~ce petit Gégé. Ils montèrent~~ ^{vient} tous les deux et tous ses
craquements étaient faits par un animal qui était
nécessaire comme avant ~~d'un homme s'endormit plus~~ puis
ils s'ennalèrent et ils revinrent tous les mercredis
et ils rentrèrent chez eux. ~~Il n'avait plus tardé puis ils~~
~~s'installèrent 1 an plus tard.~~
Que font-ils ?

S'ennalere → s'en allèrent



DIFFÉRENCES ENTRE UN SEGMENT TRANSCRIT ET UN SEGMENT ANNOTÉ :

Transcription: $\text{fai} \{P\sigma\#s\} \langle T2\#t \rangle$

Annotation: $\langle \text{fai} \{P\sigma\#s\} // T2\#fait // \rangle _ \langle \text{faits} \rangle$



TRANSCRIPTION

EC-CM2-2015-SDPB-D1-E9-V1

μ[L]eμ μ<C>eμ ét[é] <était> pas granchos<se> di[s]<dit> Gégé c'est un animal
 [E] <et> puis Max dit vie{Pσ#n} <vient> on va visit{Pσ#é}<T2#visiter> la mais{Pσ#s}on
 puis un autre craquement μ[venant]μ μ<vient>μ de l'étage Max et
 Gégé [dit en même tant là ! Vienμ<s>μ [#XXX#] <on> va voir qui c'est
 dit Gégé. Ils {Pσ#menterre} tou[t] tou<s> les deux et tou[t] tou<s> {Pσ#ses}
 craquamen{Pσ#e}<T2#craquamenes> ét{Pσ#é} <T2#ait> fai{Pσ#s}<T2#t> par un animal qui ét{Pσ#é}<T2#ait>
 [coisé] [coinsé] {Pσ#acasse} [d'un humain s'endoute] [plu] μ[p]uisμ μ<P>uisμ
 ils s'{Pσ#ennalere}<T2#ennalèreres> et ils rev{Pσ#ené} <T2#revenés> tous les mercredis
 et ils rentr{Pσ#é} <T2#rentrés> chez eu<x>. μ<Une>μ [heureμ[s]μ plus tarres puis ils
 sinstalere 1 an[s] plus tard].§
 FIN§

ANNOTATION

μ@[L]e//Ce//@μ @ét[é]//était//@ pas <@[granchos]//granchosse//@>_<grand-chose> @di[s]//dit//@ Gégé c'est un animal
 [E] //et// puis Max dit vie{Pσ#n} //vient// on va visit{Pσ#é} //T2#visiter// la mais{Pσ#s}on
 puis un autre craquement μ@[venant]//vient//@μ de l'étage Max et
 Gégé [dit_en_même_tant_là_!_Viens_#[XXX#]_//on//_va_voir_qui_c'est£
 dit_Gégé._Ils_{Pσ#menterre} @tou[t]//tous//@ les deux et @tou[t]//tous//@ {Pσ#ses}
 <craquamen{Pσ#e}//T2#craquamenes//>_<craquements> ét{Pσ#é} //était// <fai{Pσ#s}//T2#fait//>_<faits> par un
 animal qui ét{Pσ#é} //T2#était//
 [coisé] [coinsé] <{Pσ#acasse}>_<à_cause> [d'un_humain_s'endoute] [plu] μ@[p]uis//Puis//@μ
 ils <s'{Pσ#ennalere}//T2#ennalèreres//>_<s'en_allèrent> et ils <rev{Pσ#ené}//T2#revenés//>_<revenaient> tous les
 mercredis
 et ils <rentr{Pσ#é}//T2#rentrés//>_<rentraient> chez @[eu]//eux//@. [μ//Une//μ [heureμ[s]μ_plus_tarres_puis_ils£
 sinstalere_1_an[s]_plus_tard].§
 FIN§





- Boré, C. (2007, dir.). Construire et exploiter des corpus de genres scolaires, Namur (Belgique) : Presses Universitaires de Namur, collection Diptyque.
- Bonnet C. (2000). Lire des textes d'élèves: une approche longitudinale centrée sur les qualités des textes. Revue Tranel n° 33, 171-181. https://doc.rero.ch/record/19654/files/14_Bonnet.pdf
- Cappeau, P. & Roubaud, M.-N. (2005). Enseigner les outils de la langue avec les productions d'élèves. Paris : Bordas, coll. « Pratique Pédagogique ».
- Dancel, B. (2001). Un Siècle de rédactions. Écrits d'écoliers et de collégiens. C.R.D.P. de l'académie de Grenoble.
- David, J. & Morin, M.-F. (dir.) (2013). « Premières pratiques d'écriture : état des recherches francophones », Repères, 47. Lyon : École normale supérieure & Institut français d'éducation.
- Doquet, C. (2011) L'Écriture débutante. Pratiques scripturales à l'école élémentaire. Rennes, PUR.
- Elalouf M.-L. (dir.) (2005). Ecrire entre 10 et 14 ans, un corpus, des analyses, des repères pour la formation, CRDP Versailles. (Avec un cédérom pour les corpus et une arborescence pour la navigation).
- Fabre-Cols, C. (2000) Apprendre à lire des textes d'enfants, Bruxelles, De Boeck Duculot.
- Garcia-Debanc C. (2010) Segmentation, connexion et indexation dans des productions écrites d'élèves de 9 à 13 ans de deux genres textuels Synergies Pays Scandinaves n° pp. 81-96
- Grésillon, A. (1994). Éléments de critique génétique. Lire les manuscrits modernes. Paris : Presses universitaires de France.
- Roubaud, Marie-Noëlle (dir.), (2013), Langue et enseignement. Une sélection de 22 manuscrits de Claire Blanche-Benveniste (1976 -2008), revue TRANEL, 58.

• **ECRISCOL** : <http://www.univ-paris3.fr/ecriscol>