

Le Traitement automatisé des verbes dans les corpus d'écrits scolaires

Yelle KOULIBALI

EMA, CY Cergy Paris

yelle.koulibali@etu.u-cergy.fr

RÉSUMÉ

Un inventaire des travaux sur l'écriture scolaire montre l'insuffisance des méthodes d'analyse quantitative pour l'étude des corpus d'écrits scolaires. Plusieurs corpus ont été constitués, donnant lieu à des exploitations avec différents outils d'analyse, mais toujours limitées par la difficulté de pratiquer des traitements automatisés. Cette difficulté est liée à la présence de nombreuses variations dans ces corpus non normés.

Ce sont ces variations touchant la graphie, l'orthographe, le lexique et la morphologie selon les genres scolaires, qui rendent ce type de corpus réfractaire à la numérisation, à l'étiquetage et même à l'analyse en traitement automatique du langage. (Elalouf & Boré, 2007).

À partir de l'étude morphologique et lexicale des verbes dans deux corpus de CM2 et de 3^e, nous montrons comment aborder la variation à l'intérieur des corpus d'écrits scolaires et comment la traiter. Ce type de données nécessite des prétraitements avec des règles particulières de transcriptions des manuscrits des élèves que nous allons exposer afin d'aboutir à l'analyse de nos données proprement dites.

***Mots-clés :** Variation – Verbes – Textométrie – Morphologie verbale – Usage – Écriture scolaire.*

ABSTRACT

An inventory of work on school writing shows the inadequacy of quantitative analysis methods of school writing corpuses. Several corpuses have been set up, giving rise to farms with different analytical tools, but still limited by the difficulty of practicing automated treatments. This difficulty is related to the presence of many variations in these non-standard corpus.

It is these variations in spelling, spelling, lexicon and morphology according to school genres that make this type of corpus resistant to digitization, labelling and even analysis in automatic language processing. (Elalouf - Boré, 2007).

Based on the morphological and lexical study of verbs in the fifth and third grade corpuses, we show how to address variation within school writing corpuses and how to treat it. This type of data requires pre-processing with specific rules for transcription of student manuscripts that we will be exhibiting in order to come up with the analysis of our actual data.

Key words : *Variation – Verbs – Textometry – verbal Morphology – Use – School writing.*

INTRODUCTION

Le verbe est la catégorie grammaticale qui présente la plus grande variété de formes, tant pour la base que pour la désinence. En production d'écrit, choisir la forme adéquate exige de la part des élèves des calculs complexes qui mettent en jeu leurs connaissances à différents niveaux : énonciation, morphologie, syntaxe, orthographe (Gomila & Ulma, 2014). La mise en place progressive de ces réglages suscite des erreurs. Pour une forme attendue, les variations non-normées sont nombreuses. Mais sont-elles de la même nature et se répartissent-elles de la même façon dans des textes narratifs produits à la fin de la scolarité primaire et à la fin du collège ? Nous avons cherché des solutions pour les détecter et les catégoriser selon une méthodologie qui combine un traitement manuel et automatisé.

Notre question de recherche est la suivante : comment traiter la variation des formes verbales dans un corpus de textes narratifs d'élèves de CM2 et de 3^e ? Pour la traiter, nous faisons le choix, dans cet article, d'un niveau d'analyse homogène - la morphologie - et d'un niveau qui lui est étroitement lié dans la mesure où on étudie des productions écrites - l'orthographe.

Nous expliquerons d'abord la constitution du corpus et la démarche suivie du prétraitement au traitement de nos données. Nous présenterons ensuite comment se répartissent les variations des formes verbales erronées aux deux niveaux de la scolarité étudiés. Enfin, il s'agira de montrer sur un exemple prototypique comment les variations de formes peuvent être mises en relation avec d'autres niveaux d'analyse. Notre hypothèse est que par-delà une grande variabilité intra et inter individuelle, il existe des constantes qu'une analyse outillée doit permettre de mettre au jour.

1) Méthodologie

1.1 Cadre général

Notre recherche s'inscrit dans le projet ECRISCOL, développé par le laboratoire CLESTHIA, qui a pour but de constituer un corpus de référence d'écrits produits en situation scolaire (Doquet, David et Fleury, 2017) et de permettre l'analyse linguistique systématique de ces écrits à l'aide de logiciels d'analyse textuelle. Les recherches sur l'écriture scolaire ont longtemps privilégié des analyses qualitatives en raison des difficultés d'exploitation quantitative en traitement automatique du langage. Ces difficultés sont principalement dues au caractère spécifique de ces écrits qui les rendait jusqu'à présent réfractaires à la numérisation, à l'étiquetage et même à l'analyse (Elalouf & Boré, 2007). Ce n'est que récemment que des solutions ont été trouvées permettant un traitement automatisé de ces écrits non normés.

1.2 Constitution du corpus

Le corpus sur lequel porte notre étude est un extrait du corpus recueilli dans le cadre du projet ECRISCOL (Doquet, C, David, J. & Fleury, S., 2015).

Le choix des niveaux de classe s'est porté sur deux classes charnières : le CM2, dernière classe de l'école primaire et la 3^e dernière classe du collège.

Au cycle 2 et au cycle 3 de l'école primaire, les régularités de la morphologie verbale écrite sont enseignées à partir de l'étude des verbes les plus fréquents, ainsi que les variations liées à l'accord. À la fin du cycle 4, en 3^e, la morphologie verbale est censée être maîtrisée, les programmes ultérieurs ne prévoyant que des remédiations.

Le corpus étudié est constitué des écrits d'élèves de deux classes de CM2 (49 élèves ayant produit 2 versions dont V1 (le premier jet) et V2 (la réécriture après relecture ou non des remarques du professeur ; âge moyen 10 ans) et d'une classe de 3^e (22 élèves ayant produit une version, âge moyen 15 ans). Les élèves de CM2 ont produit un récit à partir de 6 images issues d'un album de littérature de jeunesse, tandis que les élèves de 3^e ont écrit également un texte narratif en imaginant la suite d'un extrait du roman de Daniel Pennac, *Cabot-Caboche*. La nature du corpus constitué et les finalités de la recherche nous ont conduite à choisir la conception du corpus selon Rastier (2002). Concevoir le corpus comme une archive nous a permis de relier chaque version. Grâce à cette conception du corpus, nous avons pu constater que certaines erreurs subsistent ou que d'autres sont ajoutées malgré les annotations du professeur. Nous pouvons caractériser le corpus étudié de corpus hétérogène parce qu'il comporte les annotations du professeur. Nous avons fait le choix de réduire cette hétérogénéité en nous centrant sur l'étude des formes produites par les élèves. Une fois transcrits, annotés et enrichis manuellement ou semi-automatiquement avec des étiquettes, les corpus ont été inclus dans le corpus de référence d'ECRISCOL.

1.3 Prétraitement et traitement des données

Notre corpus est constitué de copies scannées que nous avons transcrites en texte brut dans Notepad ++. Puis nous avons effectué l'annotation dans un fichier destiné à être lu par un logiciel textométrique. Le choix du logiciel s'est porté sur le Trameur qui permet une analyse automatique de grands corpus textuels avec un système d'étiquetage automatique des catégories grammaticales à l'aide de treetagger. Pour pouvoir faire une étude automatisée de ces verbes, nous avons relié par un codage une forme normée à une forme non normée. Cela permet d'avoir accès au contexte dans lequel apparaît la forme

non normée. Dans une perspective didactique, l'objectif est en effet l'apprentissage des formes normées, mais cela suppose la prise en compte des variantes produites par les élèves.

Exemple : Scans copie

La transcription et l'annotation ont été faites suivant les règles d'ECRISCOL. Nous avons étiqueté les verbes en les faisant précéder d'une étiquette « V :: » dans la transcription et « V@ » dans l'annotation des verbes corrects et de « E :: » et « E@ » pour tous les verbes erronés afin de faciliter leur repérage dans l'ensemble de notre base de données.

Exemple : Transcription

CM2-2017-PAS-D1-E2-V1

Page=1

C'V::est l'histoire d'un petit garçon qui f
s'V::appelle Gabriel.Un jour en partant f
à l'école il E::part{Po#it} sa toupis chinoise.

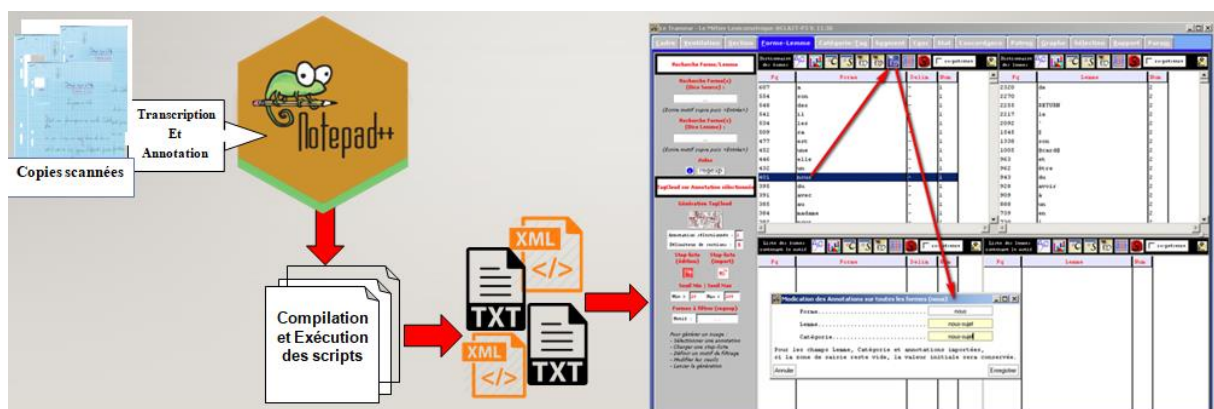
Exemple : Annotation

CM2-2017-PAS-D1-E2-V1

Page=1

C'V@est l'histoire d'un petit garçon qui f
s'V@appelle Gabriel.Un jour en partant f
à l'école il <V@part{Po#it}> <V@Partit> sa <toupis> <toupi

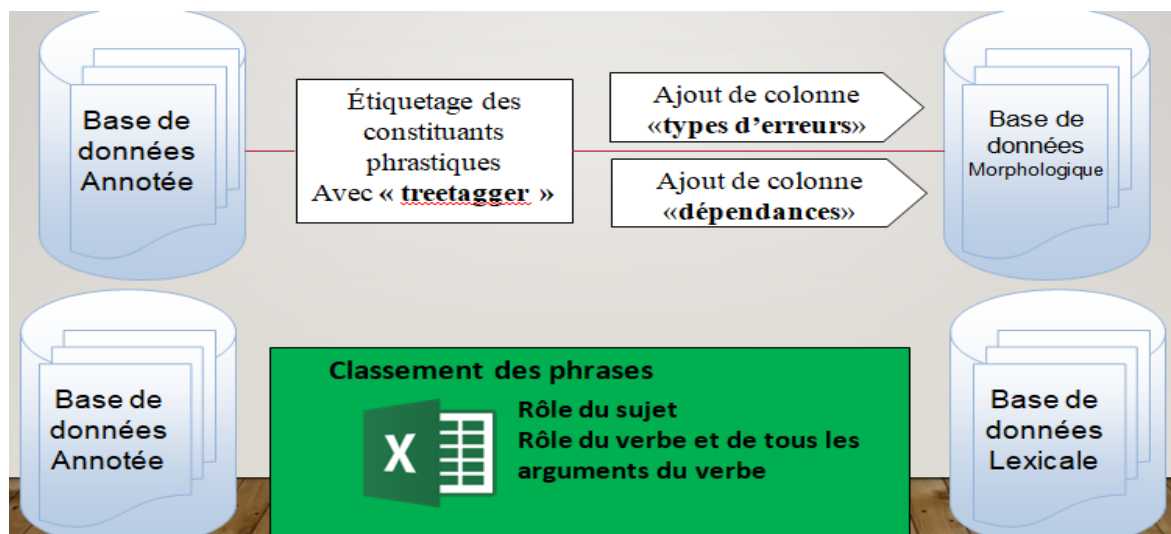
Ce prétraitement a consisté à exécuter des scripts et à étiqueter les constituants phrastiques avec l'étiqueteur morphosyntaxique « treetagger » sur les textes compilés de la base annotée.



Pour les besoins de l'analyse, cette base a été réajustée. Nous avons ajouté des colonnes représentant les types d'erreurs que nous avons identifiées dans les textes d'élèves (voir II-2). Cette démarche permet de constituer la base de données morphologique. Pour constituer la base de données lexicale permettant

de caractériser l'emploi de chaque verbe, nous nous sommes basées sur le contexte syntaxique de la phrase en classant dans un tableau Excel les phrases et en indiquant le rôle de toutes les composantes de la phrase en jeu dans la construction du texte notamment le sujet, le verbe lui-même, les connecteurs, les cadratifs, etc.

Suivant la classification sémantique de Weinrich (1989), nous avons identifié deux registres temporels que sont : le registre temporel du récit et celui du commentaire. C'est selon cette répartition que nous avons analysé nos données.



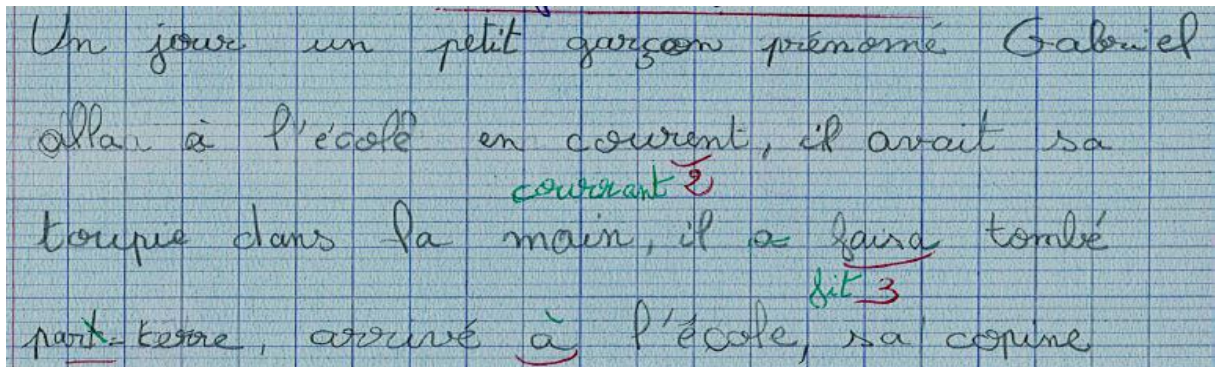
2) Classement des formes verbales erronées

C'est l'exécution de nos tâches de transcription et d'annotation qui nous a permis d'arriver à la conclusion que le verbe, maillon essentiel dans la progression des textes narratifs, présente de nombreuses variations morphologiques. La constitution du corpus permet de les appréhender de deux façons complémentaires : en recensant les formes erronées produites dans l'ensemble des écrits pour dégager des principes de classement, puis en revenant à la singularité de chaque texte pour observer la façon dont les formes verbales justes et erronées s'y répartissent.

2.1 Nature des variations

Les variations observées dans les textes d'élèves empruntent, dans une large mesure, au répertoire des formes verbales existantes, ce qui rejoint le constat effectué par Claire Blanche-Benveniste (2005) : les élèves ont rencontré des formes verbales dans les tableaux de conjugaison au cours de leur scolarité mais la connaissance de ce répertoire ne suffit pas pour effectuer les choix pertinents en contexte.

Exemple en CM :



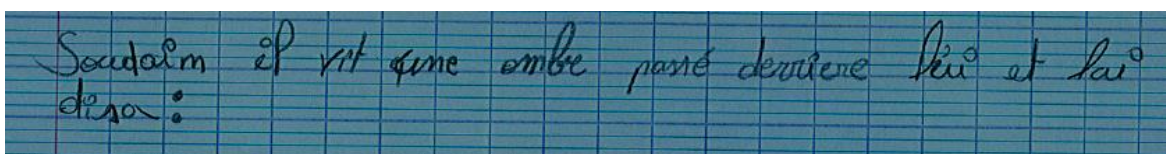
Un jour un petit garçon E::prénomé Gabriel f
V::alla à l'école en E::cour{Pσ#ent} <T2# courant>, il V::avait sa f
toupie dans la main il [T2#a] E::f{Pσ#aisa} <T2#V::fit> E::tombé f
p{Pσ#art}-terre, V::arrivé {Pσ#a} <T2#à> l'école, sa copine f

Exemple en 3° :

Le chiot le <E@scructa>_<V@scructa> puis V@gémit longuement. Ce cri V@exprimait toute f
sa tristesse et Pongot <E@compri[s]>_<V@comprit> que [s]ce miniscule chiot V@était dans f
une détresse sans nom. Ses yeux V@réflétait l'incompréhension. f

Notre hypothèse est que les élèves ont mémorisé des formes dans le cadre de l'enseignement de la conjugaison à l'emploi du temps de leur classe, mais que leur emploi en contexte est source de difficulté. Comme le montrent les travaux de Patrice Gourdet (2009), l'enseignement des formes verbales privilégie à l'école primaire l'apprentissage des tableaux de conjugaison au détriment d'une étude des variations syntaxiques. Quand les élèves produisent des formes non attestées, c'est en mobilisant des connaissances soit sur les bases, soit sur les désinences.

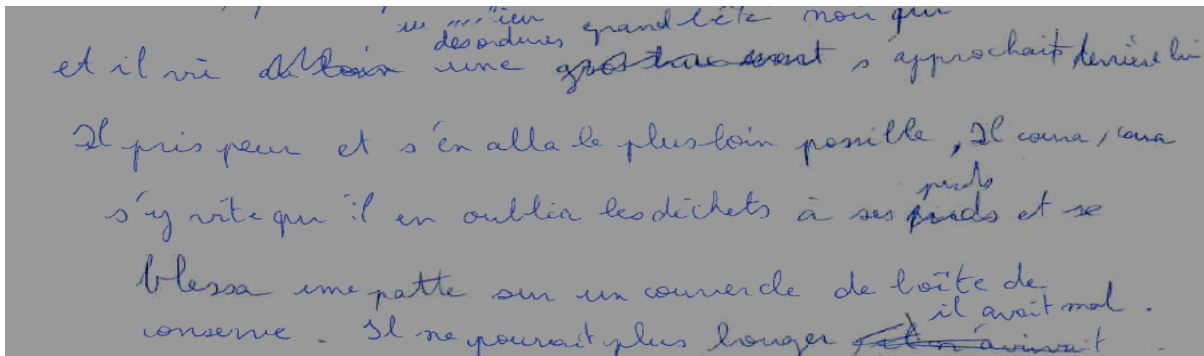
Exemple en CM :



soudain il V@vit une ombre <E@passé>_<V@passer> derrière lui et lui f
<E@disa>_<V@dit>:§

Soudain il V@vit une ombre E@passé derrière lui et lui f
E@disa:§

Exemple en 3^e :



et il <E@vi>_<vit> [au loin] //au milieu des ordures// une [gros truc avant] //grande bête noir qui// s'V@approchait derrière lui
Il <E@pris>_<prit> peur et s'en V@alla le plus loin possible, il <E@coura>_<V@courut>,<E@coura>_<V@courut> <s'y>_<si> vite qu'il en V@oublia les déchets à ses pieds et se f V@blessa une patte sur un couvercle de boîte de conserve. Il ne V@pouvait plus V@bouger [il n'V@arrivait].

et il <E@vi>_<vit> [au loin] //au milieu des ordures// une [gros truc avant] //grande bête noir qui// s'V@approchait derrière lui
Il <E@pris>_<prit> peur et s'en V@alla le plus loin possible, il <E@coura>_<V@courut>,<E@coura>_<V@courut> <s'y>_<si> vite qu'il en V@oublia les déchets à ses pieds et se f V@blessa une patte sur un couvercle de boîte de conserve. Il ne V@pouvait plus V@bouger [il n'V@arrivait].

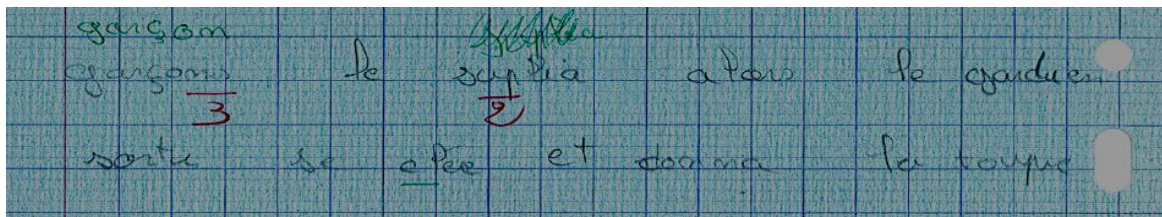
2.2 Typologie des erreurs

Nous avons identifié quatre types d'erreurs que font la plupart des élèves : les erreurs d'homophonie des formes verbales (ER_HO_FV) ; les erreurs sur les désinences orales (ER-DE-RE) ; les erreurs résultant de l'absence de la marque de personne (AB_MARK_PER) ou de l'ajout d'une marque de personne inappropriée (AJ_MARK_PER). Ces erreurs portent essentiellement sur les désinences, ce qui confirme la tendance observée chez des élèves plus jeunes par Wolfarth, Ponton & Brissaud (2018) : « Alors que le nombre des désinences erronées diminue assez peu d'un niveau à l'autre, les bases erronées diminuent fortement du CP au CE2 ».

2.2.1 Les erreurs d'homophonie (ER-HO-FV)

Les élèves ayant majoritairement choisi le registre temporel du récit, les formes verbales homophones les mieux représentées sont celles du passé simple et de la forme adjectivale du participe passé. Elles sont isolées dans cette catégorie. Ce type d'erreur se rencontre dans les deux niveaux de classe et survient sur des verbes traditionnellement classés dans le 3^e groupe. Dans l'exemple suivant, il s'agit du verbe « sortir ».

Exemple : CM2-2017-PAS-D1-E2-V1



**Garçons le E@suplia alors le gardien t
E@sorti se clé et V@donna la toupie t**

Ici, l'élève connaît la forme orale du passé simple attendue, selon la règle d'engendrement :

BASE VERBALE + « I(N) » ou « U »+ Marque de personne

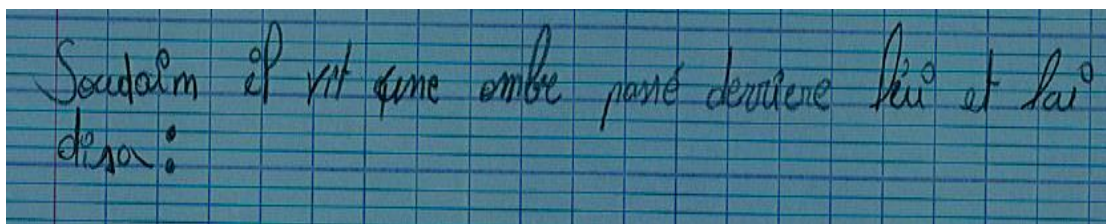
Dans notre exemple, il s'agit d'un verbe à finale « I ». La difficulté réside dans le caractère muet de la marque de personne : le son est transcrit sans prendre en considération la relation morphosyntaxique entre le sujet et le verbe. L'élève a choisi la forme adjective utilisée pour former le passé composé.

On pourrait écrire « Gabriel est sorti » : dans ce cas, le verbe « sortir » ne prendrait pas la marque de personne du sujet « Gabriel » puisqu'au passé composé, c'est l'auxiliaire qui porte cette marque. Par contre dans la phrase produite par l'élève, le verbe « sortir » doit prendre la marque du sujet « le gardien » qui est « t ». Ce type de raisonnement contrastif met en évidence une des sources de difficulté majeure du français : sa morphologie muette.

2.2.2 Les erreurs portant sur la désinence orale des formes verbales (ER_DE_RE)

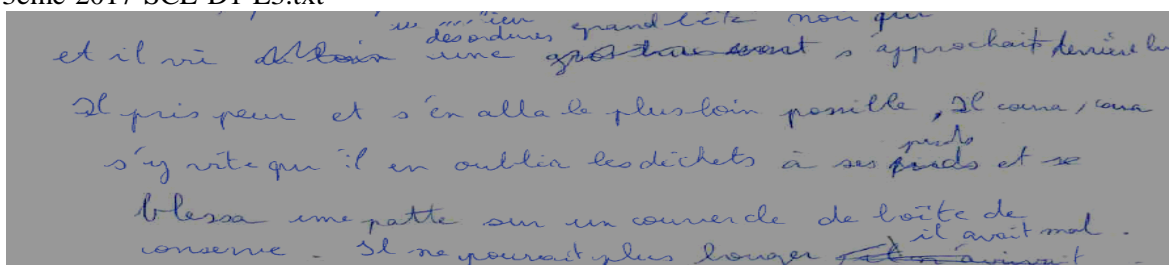
À la différence des précédentes, ces erreurs conduisent à la production de formes non attestées. Elles concernent aussi majoritairement le passé simple, temps de la narration qui n'appartient pas à l'usage oral. On les rencontre fréquemment dans les écrits de cours moyen et elles subsistent dans certains écrits de 3^e, comme l'illustrent les exemples suivants :

Exemple : CM2-2017-PAS-D1-E6-V1



Soudain il V@vit une ombre E@passé derrière lui et
lui £
E@disa:§

3ème-2017-SCL-D1-E3.txt



et il <E@vi>_<vit> [au loin] //au milieu des ordures//
une [gros truc avant] //grande bête noir qui//
s'V@approchait derrière lui£
Il <E@pris>_<prit> peur et s'en V@alla le plus loin
possible, il <E@coura>_<V@courut>, <E@coura>_<V@courut>£

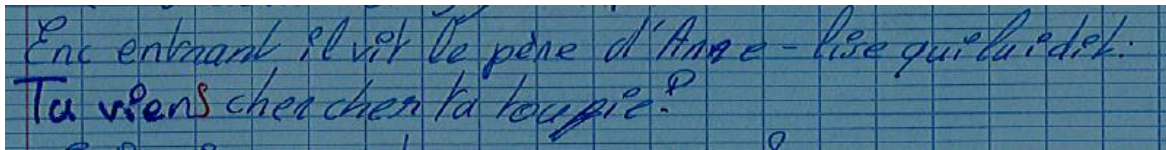
Les désinences du passé simple des verbes traditionnellement classés dans le premier groupe tendent à être généralisées à l'ensemble de la conjugaison. On remarque ainsi que la terminaison « a » est ajoutée à des bases verbales du troisième groupe : « disa » dans l'exemple de l'élève E6 (CM2-2017-PAS-D1-E6-V1), au lieu de « dit », forme identique pour le présent et le passé simple ; « coura » qui est encadré par deux verbes dont la désinence est « a », « s'en alla » et « oublia », ce qui peut expliquer l'attraction dans l'exemple de l'élève E3 (3ème-2017-SCL-D1-E3.txt).

Maîtrisant la formation du passé simple des verbes du 1^{er} groupe, traditionnellement enseignés en premier, les élèves créent par analogie une forme pour des verbes du 3^e groupe. Pour les verbes à plusieurs bases, la base de l'imparfait et du participe présent est privilégiée : « disa », « fesa ». On peut aussi faire l'hypothèse que certains environnements contextuels favorisent cette assimilation.

2.2.3 Absence de marque de personne (AB-MARK-PER)

Lorsqu'elle est muette, il est fréquent que la marque de personne soit absente.

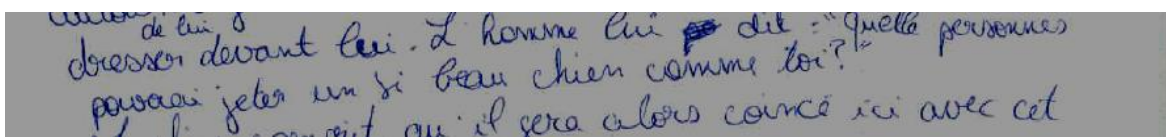
Exemples :
CM2-2017-PAS-D1-E18-V2



En V@rentrant il V@vit le père d'Anne-Lise qui lui V@dit: £
Tu <E@vien//P#s//>_<V@viens> V@chercher ta toupie? §

En V@rentrant il V@vit le père d'Anne-Lise qui lui V@dit: £
Tu <E@vien//P#s//>_<V@viens>V@chercher ta toupie? §

3ème-2017-SCL-D1-E16.txt



L'homme lui V@dit "quelle personnes £
<E@pourrai>_<V@pourrait> V@jeter un si beau chien comme toi? §

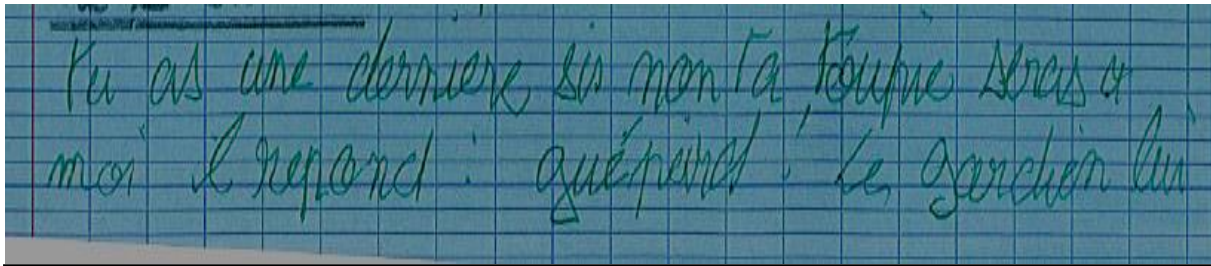
L'homme lui V@dit "quelle personnes £
<E@pourrai>_<V@pourrait>V@jeter un si beau chien comme toi? §

Le choix de la marque muette correspondant au sujet suppose une analyse syntaxique. Dans l'exemple de CM2, la forme sans marque de personne n'est pas une forme verbale attestée, tandis que dans l'exemple de 3^e, l'élève produit un futur (à la 1^e personne du singulier) pour un conditionnel (à la 3^e personne du singulier). Cette hésitation entre futur et conditionnel avait déjà été constatée dans un grand corpus de 724 réponses à l'enquête PISA (Elalouf, 2009). Dans une étude portant sur 903 récits du CP au CE2, Wolfarth, Ponton & Brissaud (2018) montrent que les marques de personnes non nécessaires à la phonographie exigent une attention particulière. Notre étude confirme que la difficulté persiste bien au-delà.

2.2.4 Ajout de marque de personne (AJ_MARK_PER)

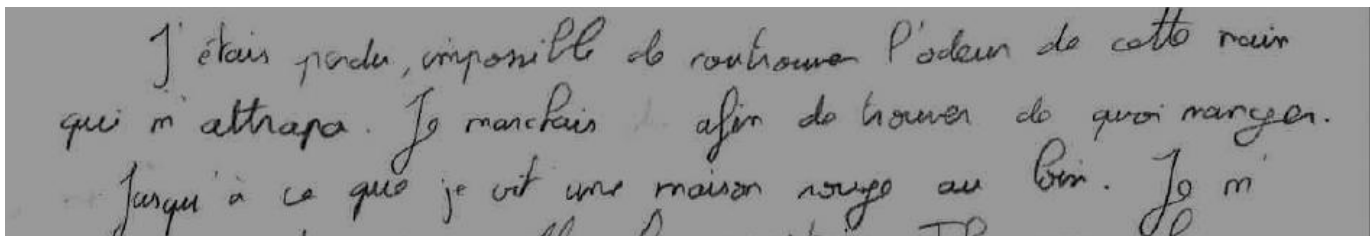
L'ajout d'une marque de personne inappropriée est une autre manifestation des difficultés causées par la morphologie muette du français.

Exemple CM2-2017-PAS-D1-E10-V1



tuV@as une dernière <sis non>_<sinon> ta toupie
<E@seras>_<V@sera> £
moi il V@repond guépard! le gardien lui£

3ème-2017-SCL-D1-E12.txt



J'V@étais V@perdu, impossible de //<E@routrouver>_<V@retrouver>// l'odeur de cette main £
qui m'V@attrapa. Je V@marchais afin de V@trouver de quoi V@manger. £
jusqu'à ce que je <E@vit>_<V@vis> une maison rouge au loin. Je m'£

J'V@étais V@perdu, impossible de //<E@routrouver>_<V@retrouver>// l'odeur de
cette main £ qui m'V@attrapa. Je V@marchais afin de V@trouver de quoi
V@manger. £ jusqu'à ce que je <E@vit>_<V@vis> une maison rouge au loin. Je
m'£

On observe que les élèves utilisent une désinence de personne qui ne correspond pas au sujet du verbe : l'hésitation entre « as » et « a », « s » et « t » montre que les élèves restituent des formes existantes sans toujours être en mesure de choisir celle que le sujet demande. On pourrait supposer que dans l'exemple de CM2, l'élève a fait un lien entre le déterminant possessif « ta » et le sujet du verbe précédent « tu » auquel il coréfère, ce qui justifierait son choix du « s », bien que le groupe nominal « ta toupie » appelle un verbe à la 3^e personne.

Au total, sur les quatre types d'erreurs répertoriés, une seule relève d'une méconnaissance de la forme verbale attendue à l'oral (ER-DE-RE). Les trois autres résultent de la non prise en compte de la relation sujet-verbe dans le choix de la marque de personne muette.

2.3 Fréquence des erreurs

2.3.1 Mode de calcul

Pour connaître les fréquences de ces erreurs, nous avons chargé la base dans le Trameur et sélectionné le type d'erreur que nous voulions compter dans les paramètres pour avoir les fréquences escomptées, comme le montre l'exemple qui suit.

Mode de calcul des fréquences dans iTrameur logiciel en ligne reprenant les fonctionnalités du Trameur

Paramètres +

Chargement Trame Cadre SR/Patron Section Coocs Bi-Texte Dépendance Sélection Export Aide

Création d'une nouvelle base / Importation d'une base

Deux possibilités pour charger des données dans iTrameur :

1. Charger un fichier (nouvelle base) au format TXT brut, encodé en UTF-8, en ayant préalablement partitionné son contenu (cf onglet Aide).
2. Importer une base annotée déjà constituée (cf onglet Aide pour le format de cette base).

Une fois la base chargée, les données textuelles sont représentées sous la forme d'une *Carte des sections* (sections définies via le délimiteur de contexte choisi) qui apparaît au bas de cette page.

1. NOUVELLE BASE Parcourir...

2. DÉLIMITEUR DE CONTEXTE § (si cette zone est vide, contexte=ligne)

3. DÉLIMITEUR(S) ;'~}|&#@='~?%*"\$000_+<=>§\

4. BI-TEXTE (chargement d'un bitexte aligné cf Aide)

5. DÉPENDANCE (chargement d'une base avec annotations en dépendance cf Aide)

iTrameur Analyse textométrique de données

Paramètres -

PARTITION COPIE PARTIES CM2-2017-PAS-D1-E10-V2 LG CONTEXTE 10 FQ MAX 5

GRAPHÉ H 400 SEUIL 5 CO-FREQ 2 INDSPHIN 5

GRAPHÉ L 800 SR LGMAX 12 SR FQMIN 10 NB SÉLECTION SECTION 1

ANNOTATION* 5 ANNOTATION SORTIE 5 RELATION REL ANNOTATION RELATION 4

PÔLE SOURCE ER_HO_FV

Chargement Trame Cadre SR/Patron Section Coocs Bi-Texte Dépendance Sélection Export Aide

Opérations sur la Trame

ANNOTATIONS À FUSIONNER 28,3

Dictionnaire* Concordance* Fusion Annotation

Dictionnaire

Copy CSV Excel PDF Print Recherche :

Item	Fq	Concordance	Ventilation	Carte
-	3327	➔	⬆	⬇
ER_HO_FV	40	➔	⬆	⬇
ERREUR_HOMO_FORM_VERB	1	➔	⬆	⬇

Affichage de 1 à 3 des 3 items Préc. 1 Suiv.

2.3.2 Résultats et comparaison des erreurs morphologiques

L'analyse textométrique permet d'obtenir un panorama des formes verbales erronées et leurs fréquences sur les deux niveaux. Nous avons également compté la fréquence de toutes les formes justes de chaque classe et la fréquence de toutes les formes erronées (tous les verbes confondus).

REPERAGE DES FORMES JUSTES

iTrameur : Analyse textométrique des données dans un fichier...

Item	Fq	Concordance	Ventilation	Carte	Séle
V@était	56				
V@perdu	32				
V@avait	32				
V@est	29				
V@dit	28				
V@as	22				
V@demanda	19				
V@ai	14				
V@suis	11				
V@a	10				
V@retrouver	9				
V@chercher	9				
V@vit	8				
V@arriva	8				
V@retrouva	7				
V@vais	7				
V@entra	7				
V@va	7				
V@veux	6				
V@chercha	6				

REPERAGE DES FORMES ERRONEES

iTrameur : Analyse textométrique des données dans un fichier...

Item	Fq	Concordance	Ventilation	Carte	Sélection
E@disa	3				
E@attenda	3				
E@la	2				
E@veu(soulignement_who=P?)t(/soulignement)	2				
E@vait	2				
E@devra	2				
E@parti	2				
E@fin(soulignement_who=P?)i(/soulignement)(ins)T2#e(/ins)	1				
E@ammène	1				
E@perdu	1				
E@seras	1				
E@tomb(remplacement)(b)(del)T2#er(/del)(/b)(a)T2#é(/a)(/remplacement)	1				
E@nomm{P?#é(ins)T2#s(/ins)	1				
E@fai(soulignement_who=P?)t(/soulignement)	1				
E@perdu(del)T2#(soulignement_who=P?)e(/soulignement)(/del)	1				

TABLEAU RÉCAPITULATIF DES FRÉQUENCES

	ER_HO_FV		ER_DE_RE		AB_MARK_PER		AJ_MARK_PER		Formes justes		Formes erronées	
	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2	V1	V2
CM2 A	46	40	24	13	41	24	27	19	625	622	169	123
CM2 B	16	24	30	1	37	7	22	4	966	956	148	92
3 ^e	11		5		7		7		664		47	

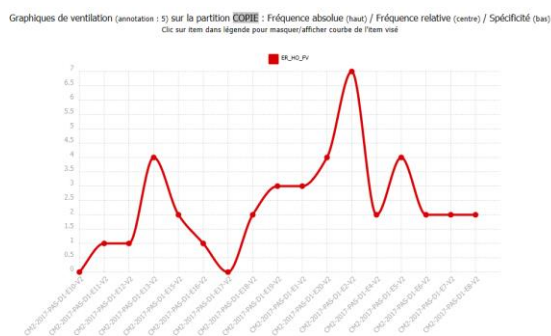
Ce tableau récapitulatif des fréquences présente le poids respectif des différentes catégories d'erreurs selon la classe. On constate que les quatre types d'erreurs présentes en CM2 se retrouvent en 3^e, dans des proportions certes moindres. Un noyau d'erreurs d'homophonie subsiste (6,8% en CM2A, 2%, CM2B et 1,6% en 3^e). Pour la classe de CM2 dans laquelle la proportion de formes erronées est la plus

importante (27% en V1 et 19,8% en V2), les erreurs d’homophonie prédominent, suivies des erreurs d’absence de marque de personne puis viennent les erreurs portant sur le choix de la désinence et de l’ajout de la marque de personne. Dans l’autre CM2, les erreurs d’homophonie sont moins nombreuses en V1 mais augmentent en V2, alors que pour les autres types d’erreurs, des corrections sont faites en V2. Les effets de la révision sont donc contrastés : l’élève peut ajouter des erreurs à la relecture ou ne pas comprendre les indices de correction donnés par l’enseignant sur les formes erronées.

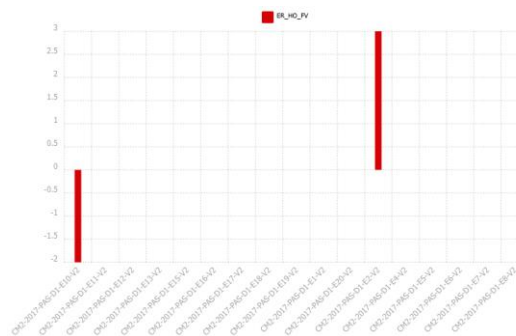
2.3.3 La ventilation des erreurs sur chaque classe

En cliquant sur l’option de calcul statistique de l’iTrameur, nous avons obtenu les graphes des fréquences absolues qui représentent les erreurs sur la classe (CM2A, CM2B ou 3^e), ainsi que les spécificités qui mettent l’accent sur l’élève représentatif de l’erreur dans cette classe.

GRAPHE DE FREQUENCE ABSOLUES



GRAPHE DE SPECIFICITE



Pour les résultats de fréquences, nous avons utilisé les versions V1 des élèves de CM2.

Comme nous avons constaté que les erreurs portant sur l’homophonie restent les plus nombreuses et les plus résistantes au signalement par le professeur en cours moyen, nous avons choisi de faire un focus sur les formes verbales concernées par ER-HO-FV.

Nous montrerons d’abord la répartition de cette erreur dans les textes d’une des classes de CM2 puis nous étudierons une copie que nous avons jugée prototypique de cette erreur.

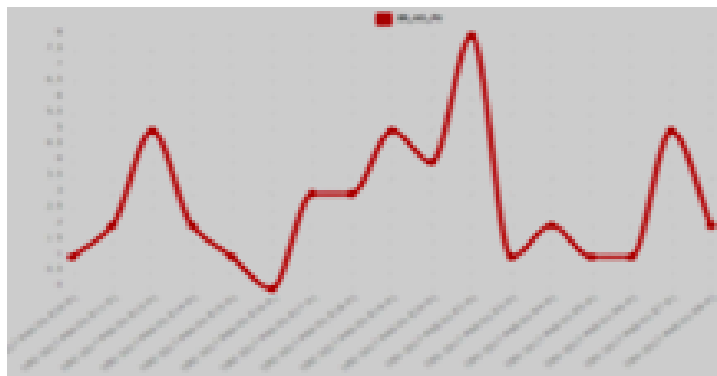
À l’aide du Trameur, nous avons cherché les fréquences des erreurs portant sur l’homophonie.

LES FRÉQUENCES

Item	Fq
-	3469
ER_HO_FY	46

La première ligne du tableau indique l'absence d'erreur par des tirets (-) et la seconde la fréquence de ce type d'erreur dans cette classe selon le iTrameur.

Graphe de fréquence absolue



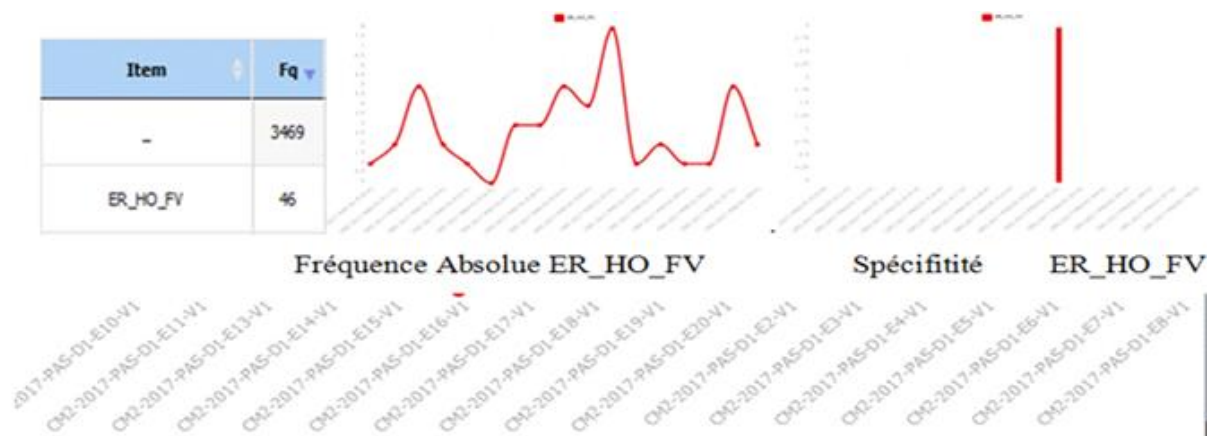
GRAPHE DE SPÉCIFICITÉ



CODE D'IDENTIFICATION DES ÉLÈVES

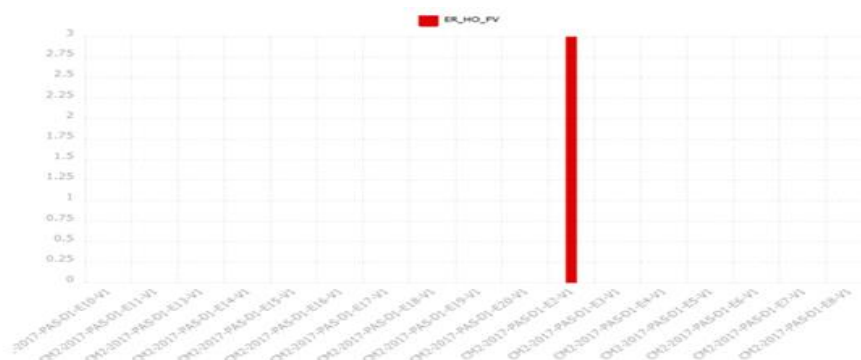


RÉCAPITULATIF DE LA RECHERCHE DANS LE iTRAMEUR



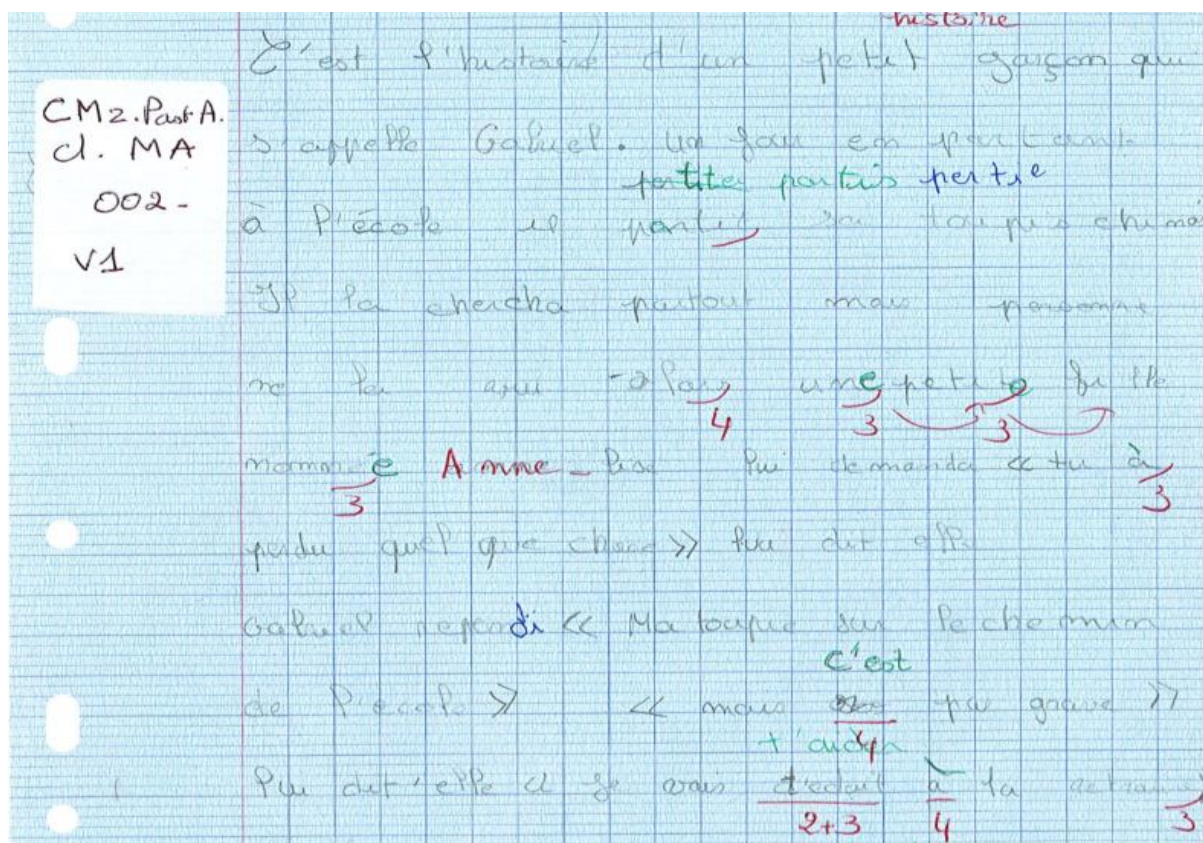
Cette figure nous montre qu'il y a 46 formes verbales erronées dans les écrits de cette classe de CM2 qui sont du type ER_HO_FV dans le corpus de CM2 A V1. Sur le graphe de la fréquence absolue presque tous les élèves font des erreurs de ce type. Lorsqu'il y a une ascendance verticale de la courbe, cela signifie que l'élève a fait une erreur de ce type, mais lorsque la courbe est horizontale, cela veut dire que l'élève n'a fait aucune erreur de ce type. Quant au graphique de spécificité, il indique que l'élève E2 est l'élément représentatif de cette erreur dans cette classe, dans la mesure où son texte comporte le plus grand nombre d'erreurs de ce type (ER-HO-FV).

GRAPHE DE SPÉCIFICITÉ



3. Étude d'un texte prototypique

Le texte de l'élève E2 présente un nombre important d'erreurs d'homophonie, tant dans le registre du commentaire (III-1-1) que dans le registre du récit (III-1-2) (Weinrich, 1989 : 125-126). Nous montrerons que la fréquence d'erreurs d'homophonie s'accompagne de problèmes relevant d'autres niveaux d'analyse. Cela suppose de prendre en considération l'ensemble du texte de l'élève.



CM2-2017-PAS-D1-E2-V1

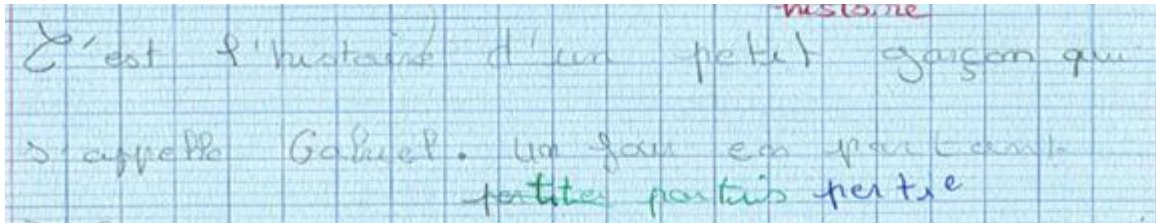
Page=1

C'est l'histoire d'un petit garçon qui s'appelle Gabriel. Un jour en partant à l'école il a perdu sa toupie chinoise. Il l'a cherchée partout mais personne ne l'a vue. Alors une petite fille nommée Anne-Lise lui demanda "tu as perdu quelque chose" lui dit elle Gabriel répondit "Ma toupie sur le chemin de l'école" "mais c'est pas grave" lui dit elle "je vais t'aider à la retrouver et les deux amis commencèrent à chercher mais ils n'ont rien trouvé et Anne-Lise eut l'idée de partir chez le gardien de l'école ils y allèrent chez le gardien de l'école."

3.1 Morphologie et orthographe

3.1.1 « C'est » écrit de deux façons différentes

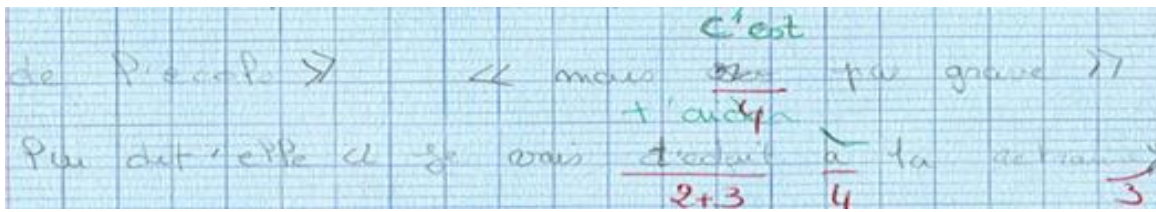
« C'est » est bien orthographié en début de récit, mais n'est pas identifié dans la locution « c'est pas grave » : l'élève écrit « ses » et corrige après le signalement de l'erreur par le professeur.



CM2-2017-PAS-D1-E2-V1

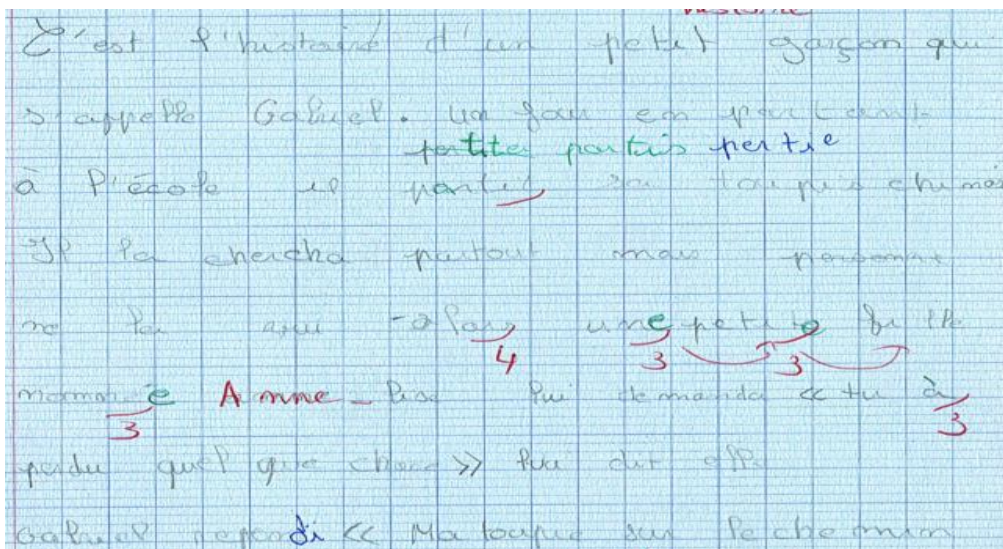
Page=1

C'est l'histoire d'un petit garçon qui s'appelle Gabriel. Un jour en partant à



Gabriel <E@repondi> <V@repondit> "Ma toupie sur le chemin de l'école" "mais <{Po@ses}> <c'V@est> //T2#c'V@est// pas grave" à lui V@dit'elle "je V@vais <{Po@t'E@edait}> <t'aider> //T2#t'V@aider// <{Po@a}> <à> //T2#à// la <E@retrouve> <V@retrouver

3.1.2 Problèmes pour les passés simples en -i ou en -u



C'est l'histoire d'un petit garçon qui s'appelle Gabriel. Un jour en partant à l'école il a perdu sa toupie chinoise. Il la chercha partout mais personne ne l'a vue. Alors une petite fille nommée Anne-Lise lui demanda "tu as perdu quelque chose" lui dit elle Gabriel a répondu "Ma toupie sur le chemin".

Les erreurs portent sur des verbes très fréquents, traditionnellement classés dans le 3^e groupe.

La concurrence entre la forme du participe passé en « u » et celle du passé simple est manifeste dans la graphie « il l'a *vui » qui convoque à la fois « il la vit » et « il l'a vu ». Le choix de la marque de personne fait l'objet de nombreux tâtonnements qui accompagnent une hésitation lexicale entre « perdre » et partir » : après le soulignement du « i » par le professeur, plusieurs désinences sont proposées, toutes erronées : « tes », « s », « e ».

3.2 Les problèmes qui relèvent d'un autre niveau

Nous allons montrer que les erreurs morphologiques s'articulent avec des difficultés relevant de différents niveaux d'analyse, en suivant la démarche adoptée par Paul CAPPEAU et Marie-Noëlle ROUBAUD (2018) : « C'est dans cet aller et retour entre la linguistique et l'analyse des textes que peuvent être dégagées des informations précieuses sur les usages langagiers des enfants, à un moment donné et sur leurs évolutions dans la maîtrise de la langue écrite ».

3.2.1 La segmentation

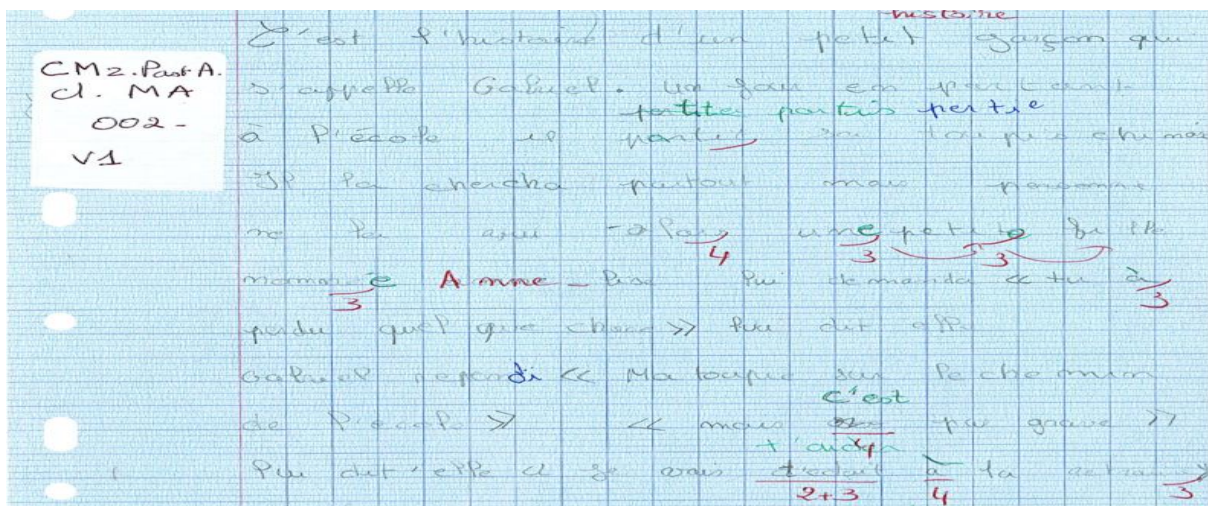


Mais au moment où le gardien def
 l'oublie <{P@#E@ouvert}>_<V@ouvrit> //T2#V@ouvrit// la <{P@#porter}>_<porte> //T2#porter// il <E@s'avait>_<V@savait>f
 que V@s'était pour la toupie mais il f
 ne <E@vois les>_<V@voulait> pas <E@donne> <donner> et le petit s
 <garç{P@#ons}>_<garçon> //T2#garçon// le <E@suplia>_<V@supplia> [T2#E@suplia] alors le gardien f
 <E@sorti>_<V@sorrit> <se>_<ses> <{P@#clée}>_<clés> et V@donna la toupie f

L'élève écrit « vois les » pour « voulait ». Pour Paul Capeau et Marie-Noëlle Roubaud, ce type de variation, appelé sous-segmentation, survient chez les élèves du CP au CE1, donc chez des plus jeunes. Ne sont donc pas résolus chez cet élève des problèmes que l'on rencontre en tout début d'apprentissage.

3.2.2 Le texte est sous-ponctué

La ponctuation joue un rôle essentiel notamment pour la lecture (Capeau et Roubaud, 2018 : 34-36). Dans ce texte, on observe que certains passages narratifs sont sous-ponctués alors que le discours direct est bien encadré par des guillemets.



3.2.3 Le changement de repère énonciatif

CM2-2017-PAS-D1-E2-V1

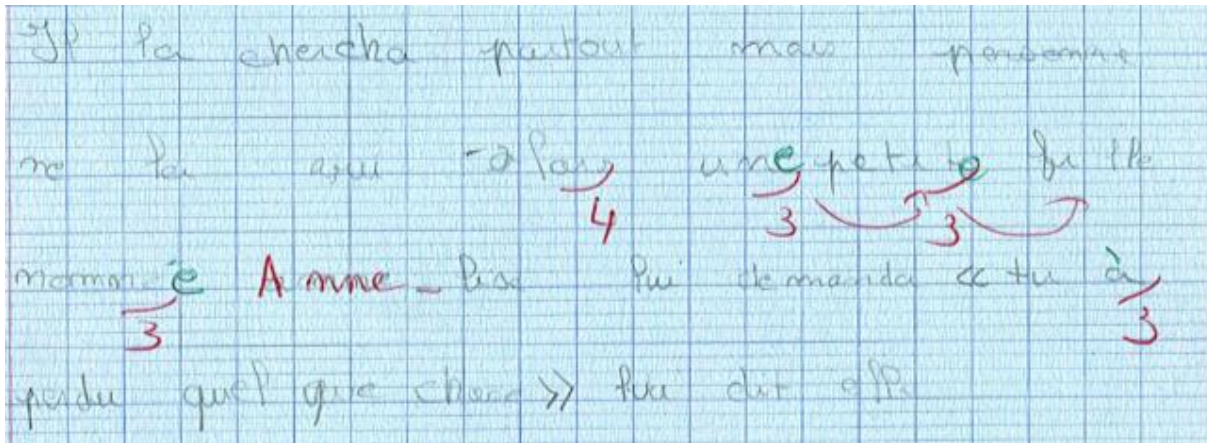
Page=1

C'est l'histoire d'un petit garçon qui s'appelle Gabriel. Un jour en partant à l'école il a perdu sa toupie chinoise. Il la chercha partout mais personne!

On observe un changement de repère énonciatif entre la 1^e et la 2^e phrase, qui accompagne le passage du registre du commentaire à celui du récit : le début du texte s'apparente à un récit fait à l'oral ; à la deuxième phrase l'élève passe du présent à un récit au passé simple, avec comme repère « un jour ».

3.2.4 Introduction du discours direct doublée d'une incise

Cette redondance peut être l'effet d'une écriture pas à pas : au moment où il ferme les guillemets, l'élève semble avoir oublié, sous l'effet d'une surcharge cognitive, qu'il a déjà introduit le discours direct par un verbe de parole.



Il la V@chercha partout mais personne!

ne <E@la> <l'v@a> <E@vui> <V@vu> <al{P@#ar}> <alors> <un//T2#e//> <une> <petit//T2#e//> <petite> fille f
<V@nommé//T2#é//> <nommée> <@[P#anne]//P#Anne//@-lise> <Anne-Lise> lui V@demanda "tu <{P@#@[E@a]//T2#à//@}> <V@as>f
perdu quelque chose" lui V@dit elle §

CONCLUSION

L'analyse outillée d'un corpus d'écrits scolaires a permis de répertorier les variations non normées proposées pour la graphie d'une forme verbale en les rapportant à leur contexte de production. Par-delà une grande variabilité intra et interindividuelle, nous avons constaté que les difficultés strictement morphologiques étaient limitées au regard de celles mettant en jeu un calcul syntaxique pour le choix de la marque de personne, qui subsistaient dans le corpus de 3^e. Ces résultats confirment ceux issus du suivi de 5 classes de CE2 pendant 2 ans, présentés dans la thèse de Patrice Gourdet (2009). L'enseignement de la conjugaison qui privilégie l'apprentissage de tableaux de formes à la mise en évidence des marques de personne régulières et des relations syntaxiques qu'elles balisent n'est pas étranger à la persistance de ces difficultés. Le fait que les verbes traditionnellement classés dans le 3^e groupe concentrent les erreurs, alors même qu'ils sont très fréquents, interroge sur leur présentation scolaire. La réponse à notre interrogation était déjà donnée par Serge Meleuc et Nicole Fauchard (2000) qui proposaient un changement dans les habitudes d'enseignement de la conjugaison encore calquées sur les habitudes d'apprentissage des paradigmes du latin. Convoquée de façon extensible, la notion d'*irrégularité* y masque souvent des erreurs de délimitation entre la base et la désinence qui perturbent la perception des régularités. Enfin, l'intérêt d'une analyse outillée est d'identifier les écrits prototypiques d'un type d'erreur et de les mettre en relation d'autres niveaux d'analyse. Cela ouvre la voie à une approche systémique appréhendant selon différents angles la singularité des écrits scolaires.

BIBLIOGRAPHIE

- Blanche-Benveniste, C. (2005). Structure et exploitation de la conjugaison des verbes en français contemporain. *Le français aujourd'hui* 148, 77-87.
- Boré, C. (2006). L'écriture scolaire de fiction comme rencontre du langage de l'autre. *Repères* 33, 37-60.
- Boré, C., & Elalouf, M.-L. (2017). Deux étapes dans la construction de corpus scolaire : Problèmes récurrents et perspectives nouvelles. In Claire Doquet, Jacques David, & Serge Fleury, (Ed.), *Corpus 16*.
- Doquet, C, David, J. & Fleury, S. (2015). Projet *ECRISCOL*. EA-7345, C. (2007). Consulté le 16 mars 2017, sur <http://www.tal.univ-paris3.fr/trameur/>
- Doquet, C, David, J. & Fleury, S. (Ed.) (2017). Spécificité et contrainte des grands corpus de textes scolaires : Problèmes de transcription, d'annotation et de traitement. *Corpus 16*.
- Doquet, C, Enouï, V., Fleury, S., Mazziotti, S. (2017). Problèmes posés par la transcription et l'annotation d'écrits d'élèves. In Claire Doquet, Jacques David, & Serge Fleury, S. (Ed). *Corpus*, 16.
- Meleuc, S. et Fauchard, N. (2000) : *Didactique de la conjugaison ; le verbe autrement*. Paris : Bertrand-Lacoste.
- Elalouf, M.-L. & Boré, C. (2007). Construction et exploitation de corpus d'écrits scolaires. *Revue Française de Linguistique Appliquée XII-1*, 53-70.
- Elalouf, M.-L. (2014). Les formes verbales complexes en fin de scolarité obligatoire : enjeux pour l'apprentissage du français écrit et d'autres langues ». In Gomila, Corinne & Ulma, Dominique, *Le verbe en toute complexité*, Paris : L'Harmattan, 39-56.
- Elalouf, M.-L. et Roubaud, M-N. (2016). Un couple de verbes antinomiques à l'épreuve de la syntaxe et de l'usage. Le cas des verbes : *se souvenir et oublier* », *Scolagram n°2*, La syntaxe du verbe, usage et enseignement. *En ligne* : En ligne : <http://scolagram.u-cergy.fr>
- Gomila, C. & Ulma, D. (dir.) (2014). *Le verbe en toute complexité*. Paris : L'Harmattan.
- Gourdet, P. (2009). *L'enseignement de la grammaire à l'école élémentaire : le cas du verbe en CE2 Des savoirs à enseigner aux savoirs appris : un facteur efficient, l'enseignant*. Thèse en cotutelle sous la direction de Danielle Leeman (Paris Nanterre) et d'Emma Nardi (Roma Tre).
- Weinrich, H. (1989). *La Grammaire textuelle du français*
- Wolfarth, C., Ponton, C. & Brissaud, C. (2018). Gestion de la morphographie verbale en production d'écrits : que peut nous apprendre un corpus longitudinal ? *Repères* 57, 209-226.