



HAL
open science

Évolution de la gestion de l'orthographe en production de textes entre 6 et 15 ans : apports du TAL

Claude Ponton, Catherine Brissaud, Claire Wolfarth

► **To cite this version:**

Claude Ponton, Catherine Brissaud, Claire Wolfarth. Évolution de la gestion de l'orthographe en production de textes entre 6 et 15 ans : apports du TAL. Enseigner et apprendre l'orthographe à l'heure du numérique, université Paris Est Créteil; université Grenoble Alpes, Oct 2021, Créteil, France. pp.01005, 10.1051/shsconf/202214301005 . hal-04818965

HAL Id: hal-04818965

<https://hal.science/hal-04818965v1>

Submitted on 11 Dec 2024

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 Attribution 4.0 International License

Évolution de la gestion de l'orthographe en production de textes entre 6 et 15 ans : apports du TAL

Claude Ponton¹, Catherine Brissaud¹ et Claire Wolfarth¹

¹ Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France

Résumé. Nous présenterons une analyse assistée par des outils informatiques de données recueillies dans le cadre du projet E-Calm, pour les niveaux allant du CP à la 3^e. Nous étudierons l'évolution de la segmentation en mots, du traitement du verbe et de la morphologie de l'adjectif dans des productions narratives. Des pistes didactiques seront proposées.

Abstract. Evolution of spelling management in text production between 6 and 15 years old: contributions of NLP. We will present a computer-assisted analysis of data collected in the framework of the E-Calm project, for levels between grade 1 and 9. We will study the evolution of word segmentation, verb processing and adjective morphology in narrative productions. Didactic approaches will be presented.

Alors que le travail de constitution de corpus scolaires a commencé dès le début des années 2000 (Elalouf, 2005, 2011), la performance orthographique en production d'écrit est encore peu documentée. Or la gestion de l'orthographe y est souvent décrite par les enseignants comme particulièrement défaillante, en regard des performances des élèves dans les exercices proposés dans les séances consacrées à l'orthographe. Par ailleurs, un alarmisme ambiant, français comme francophone, stigmatise le niveau orthographique des élèves et pointe la baisse continue de leur niveau. En France, cette baisse est attestée par des études conduites par le ministère de l'éducation nationale français, notamment en ce qui concerne l'orthographe grammaticale dans des textes de dictée (MEN-DEP, 1996 ; Manesse et Cogis, 2007 ; Andreu et Steinmetz, 2016).

Dans les mises à l'épreuve de dispositifs destinés à améliorer la compétence orthographique des élèves, la qualité orthographique des textes est rarement évaluée (des exceptions cependant avec l'étude de Fisher et Nadeau (2014) et le projet « Twictée pour l'apprentissage de l'orthographe » (TAO) (Brissaud *et al.*, 2019)). Les récentes évaluations nationales conduites en France en classe de sixième ne comportaient pas de production d'écrit.

Nos recherches sur l'évolution des compétences orthographiques en primaire et collège s'inscrivent au sein du projet ANR E-Calm¹ (Écriture scolaire et universitaire : Corpus, Analyses Linguistiques, Modélisations didactiques) dont les principaux objectifs sont :

- la constitution et mise à disposition d'un vaste corpus d'écrits d'élèves et d'étudiants²,
- la caractérisation de ces écrits notamment du point de vue de l'orthographe et de la cohérence,
- l'étude des interventions des enseignants et de leur influence sur les réécritures.

Les analyses présentées ici sont issues en grande partie des travaux menés dans la tâche *Orthographe* de ce projet et s'appuient sur une partie du corpus développé.

Les données de l'étude

Le corpus complet servant de base à cette étude est constitué d'écrits d'élèves du cours préparatoire (première année primaire) à la classe de 3^e (4^e année du secondaire, dernière année du collège français) répartis de la manière suivante :

¹ <http://e-calm.huma-num.fr/le-projet/>

² La version 2 du corpus est disponible sur la plateforme Ortolang : <https://hdl.handle.net/11403/e-calm/v2>

Niveau	Nombre de textes	Nombre de mots	Longueur moyenne des textes en nombre de mots
CP	373	10 186	27,31
CE1	373	25 445	68,22
CE2	369	44 337	120,15
CM1	369	53 727	145,60
CM2	336	60 276	179,39
6e	161	29 109	180,80
3e	90	24 822	275,80
Total	2071	247 902	119,70

Le sous-corpus CP-CM2 correspond au corpus Scoledit³ (Wolfarth *et al.*, 2018a) qui comporte des textes narratifs produits par les mêmes enfants suivis du CP au CM2 (corpus longitudinal). Les sous-corpus de 6^e et de 3^e sont composés d'écrits de collégiens recueillis par les équipes Clesthia dans le projet Ecriscol⁴ (Doquet *et al.*, 2017) et CLLE⁵ (Garcia-Debanç *et al.*, 2018).

Pour assister la description linguistique de ce corpus, une chaîne de traitement semi-automatique (Wolfarth *et al.*, 2018b) a été développée dans le cadre du projet E-Calm. Cette chaîne prend en entrée les transcriptions et une version normée des textes (Ho-Dac *et al.*, 2020). Des comparaisons entre mots transcrits et mots normés sont ensuite opérées à différents niveaux : graphique, phonologique, lexical, morphosyntaxique... Ces comparaisons fournissent ainsi des éléments de description sur les réussites et les difficultés des élèves comme, par exemple, sur la segmentation en mots, la morphologie verbale ou la morphologie adjectivale, que nous développerons tour à tour.

La segmentation en mots au CP

Cette première exploitation⁶ du corpus se concentre sur l'analyse des segmentations non conventionnelles.

Connaitre les correspondances entre graphème et phonème (principe phonographique) ne suffit pas pour écrire le français ; il y a aussi le principe sémiographique qui renvoie au sens et qui inclut la délimitation des mots. Cela implique de se construire une conception de ce qu'est un mot quand on écrit. Ainsi la séparation entre les mots, représentées par des blancs ou espaces, n'est pas naturelle, notamment parce qu'elle ne correspond pas à la segmentation de la parole. Il faudra plusieurs années à l'élève de l'école élémentaire pour en venir à bout en production d'écrit. Cela dit, on a pu observer des progrès très nets à l'échelle d'une année scolaire, entre le début et la fin de l'année de CP : la phrase dictée *Tom joue avec le rat* dans le cadre de l'étude *Lire et écrire au CP* a permis de mesurer les progrès des élèves sur la seule année de CP : « En septembre, 15,2 % des élèves n'ont pas écrit la phrase et 61 % l'écrivent en un seul bloc. 23,8 % des élèves commencent à segmenter la phrase en mots (de 2 à 6 mots). En juin, près de 90 % des élèves segmentent la phrase en mots (4, 5 ou 6 segments) » (Goigoux, 2016, p. 170).

La segmentation dans l'écriture est représentée visuellement par l'insertion d'espaces vides (ou de blancs) entre les mots. L'hyposegmentation consiste à laisser moins d'espaces vides qu'attendus, c'est-à-dire à agglutiner les mots plus que nécessaire (par exemples *senva* pour *s'en va*) ; inversement, l'hypersegmentation consiste à insérer des blancs à l'intérieur des mots et donc à fragmenter les mots graphiques (par exemple *a ve* pour *avait*).

Les deux phénomènes peuvent se combiner, par exemple *par seque* pour *parce que* (hypersegmentation de *parce* et hyposegmentation de *seque*).

Présence de l'hypo et de l'hypersegmentation dans le corpus Scoledit

Un peu plus d'un tiers des 333 textes de CP analysés ne présente aucun problème de segmentation (37,54 %). L'hyposegmentation est le phénomène le plus fréquent dans notre corpus, conformément aux recherches antérieures (Ferreiro et Pontecorvo, 1996 ; Piacente et Querejeta, 2012 ; Ugarte et Argüero, 2017). On en trouve au moins un exemple dans 54,35 % des 333 textes. Les textes présentant au moins un cas d'hypersegmentation sont moins nombreux : ils représentent 23,42 % du corpus.

Enfin, 8,11 % des textes présentent au moins un cas mêlant hypo et hypersegmentation.

³ <http://scoledit.org/scoledition/>

⁴ <http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/>

⁵ <http://redac.univ-tlse2.fr/corpus/resolco.html>

⁶ Ces résultats s'appuient sur un texte publié dans la revue *Langue française* (Ponton *et al.* 2021) qui porte sur 333 des 373 textes de CP maintenant disponibles.

Nombre d'éléments impliqués

S'agissant des cas d'hyposégmentation, les cas composés de deux éléments ou mots graphiques réunis en un seul mot prédominent (83,69 %). Les cas d'hyposégmentation impliquant plus de deux éléments représentent 16,06 % des occurrences.

En ce qui concerne les cas d'hyposégmentation, on constate une nette prévalence d'une hyposégmentation en deux éléments. Ce cas représente 98,15 % des hyposégmentations dans le corpus.

Catégories grammaticales concernées par les phénomènes d'hyposégmentation

Les éléments de type fonctionnel/grammatical (articles, pronoms, prépositions et conjonctions) ont été distingués des éléments de type lexical (noms, adjectifs, verbes et adverbes). Le cas le plus fréquent d'hyposégmentation est composé d'un élément fonctionnel et d'un élément lexical (83,95 %) ; la configuration premier élément = pronom personnel et second = verbe prédomine (30 %), par exemple *la* (l'a), *sest* (s'est), *cerévei* (se réveillent), *selève* (se lève), *onva* (on va). 7 % des cas consistent en une préposition suivie d'un nom, par exemple *partere* (par terre).

L'hyposégmentation : un premier élément de type fonctionnel

Le premier élément est le plus souvent de type fonctionnel : 10,18 % des cas contiennent « a » comme premier élément (*a ve* (avait), *a pre* (après)) ; 8,33 % des cas présentent « c' » comme premier segment (*c'est* (ses)) ; 6,48 % des cas impliquant le segment « et » comme premier (*et te* (était)), et 6,48 % avec le segment « ton », par exemple, *ton be* (tombe).

La segmentation : conclusion

La proportion de mots graphiques correctement séparés par des blancs est révélatrice de la conscience qu'a l'élève de l'organisation spatiale de l'écrit et des problèmes que pose le passage de la chaîne orale à la chaîne écrite. Cette observation a déjà été réalisée dans d'autres langues romanes. Ainsi, ces données semblent confirmer que les segments que les élèves ont tendance à accoler plus qu'ils ne le devraient ou, inversement, à segmenter à l'excès, non seulement ne sont pas aléatoires mais se retrouvent dans diverses langues (Ponton *et al.*, 2021).

Le traitement du verbe

Cette section s'appuie principalement sur des résultats réactualisés parus respectivement dans les revues *Repères* et *Langue Française* (Wolfarth *et al.*, 2018b ; Lavieu-Gwozdz *et al.*, 2021). Le corpus de la présente étude contient 43 265 formes verbales (hors participes passés) représentant 17,45 % des mots de notre corpus. Cette proportion est relativement stable sur l'ensemble des niveaux scolaires. Du fait, en partie, des consignes et du genre narratif des textes, les élèves mobilisent principalement quatre temps ou tiroirs verbaux (présent, passé simple et imparfait de l'indicatif, infinitif présent) qui représentent à eux seuls 95,61 % de l'ensemble. Une première comparaison entre formes produites et formes normées montre une progression régulière dans la réussite de ces formes verbales puisque l'on passe de 25,85 % de formes correctement orthographiées au CP à 83,00 % en 3^e. Toutefois, cette progression est à nuancer selon les tiroirs verbaux. Comme le montre la figure 1, l'imparfait et surtout le passé simple sont ceux pour lesquels les taux de réussite observés sont inférieurs à ceux des autres tiroirs verbaux fréquemment employés par les élèves, quel que soit le niveau de scolarité, à partir du CE1.

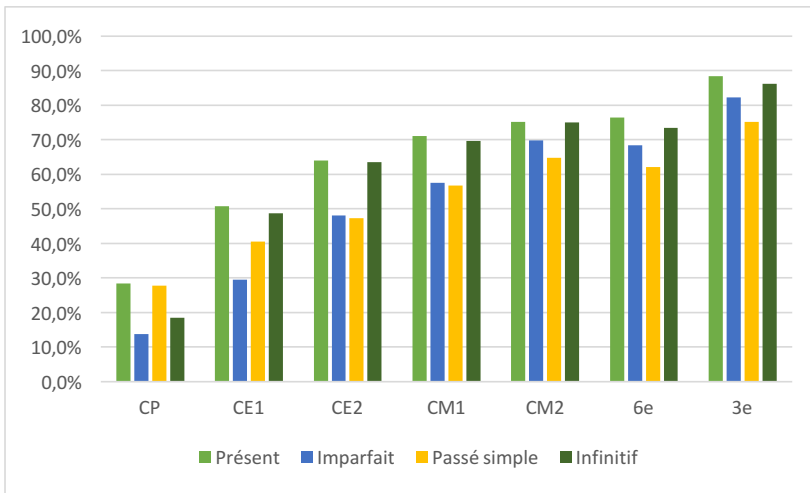


Figure 1. Évolution des réussites selon le tiroir verbal

Pour affiner cette première analyse, nous avons développé un outil permettant de décomposer automatiquement chaque forme verbale en base et désinence (par exemple, la forme « voyait » se découpe en (voy + ait)). Ce découpage est opéré aussi bien sur la forme produite par l'élève que sur la forme attendue (dite forme normée). Nous procédons ensuite à des comparaisons entre (base et désinence produites) et (base et désinence normée) pour déterminer les lieux d'erreurs, à savoir :

- Erreur portant uniquement sur la base lexicale : comme, par exemple, en CM2, la forme produite « eitégnit » (eitégn + it) au lieu de « éteignit » (éteign + it)
- Erreur portant uniquement sur la désinence : comme, par exemple, en CE2, la forme produite « suivre » (suiv + er) au lieu de « suivait » (suiv + ait)
- Erreur portant à la fois sur la base lexicale et sur la désinence : comme, par exemple, en CE1, la forme produite « arivat » (ariv + at) au lieu de « arriva » (arriv + a).

La figure 2 ci-dessous montre l'évolution des lieux d'erreur selon les niveaux de scolarité.

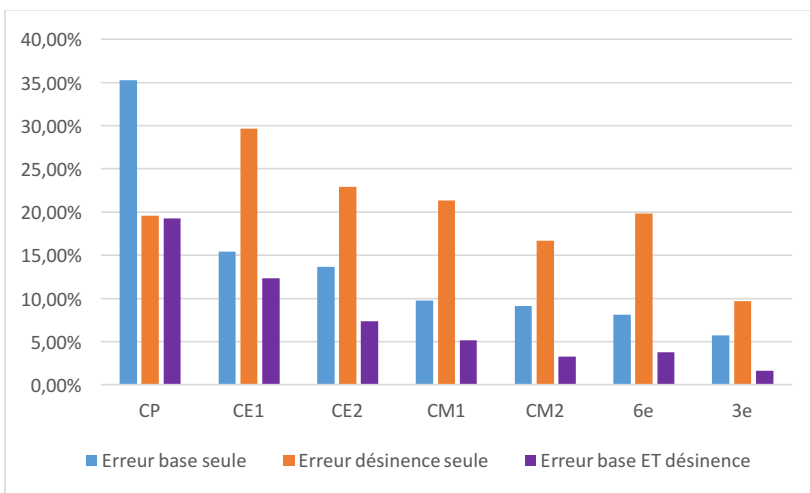


Figure 2. Évolution des erreurs sur bases et désinences

Le taux d'erreur global sur les bases (respectivement sur les désinences) est obtenu en cumulant erreurs sur bases seules avec erreurs sur bases et désinences (respectivement erreurs sur désinences seules avec erreurs sur bases et désinences). Il apparaît alors clairement que si la gestion des bases lexicales s'améliore rapidement (de

54,52 % d'erreurs au CP à 27,72 % en CE1 et seulement 7,38 % en 3^e), la gestion des désinences verbales reste une difficulté persistante puisqu'elle occasionne encore 11,33 % d'erreurs en 3^e. Comme pour les adjectifs, présentés dans la partie suivante, on constate que le passage du CM2 à la 6^e semble accompagné d'une stagnation des compétences voire d'une légère régression comme le montrent les résultats sur les désinences pour lesquels le pourcentage d'erreurs passe de 19,92 % en CM2 à 23,60 % en 6^e.

Notons que, comme le montre la figure 3, la répartition des erreurs selon la base ou la désinence ne se présente pas de la même façon selon les tiroirs verbaux.

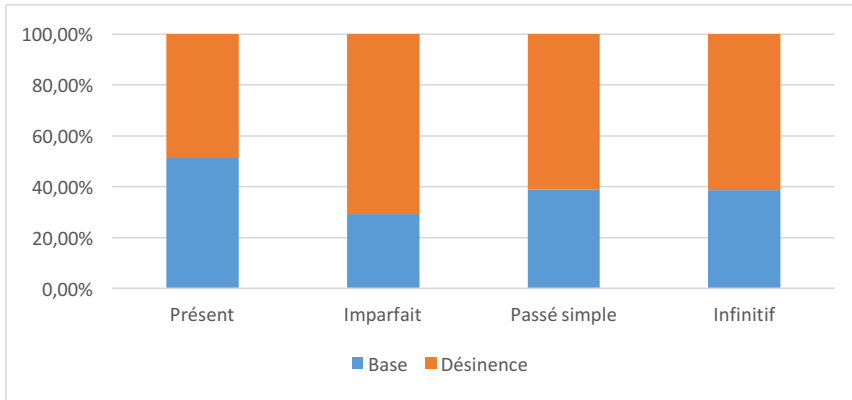


Figure 3. Répartition des lieux d'erreur selon le tiroir verbal

Le présent de l'indicatif est le tiroir pour lequel le taux d'erreur sur les bases est le plus élevé ; il dépasse même légèrement celui sur les désinences (51,24 % contre 48,76 %). Sachant que le corpus comporte essentiellement des verbes à la première et à la troisième personne du singulier (85,21 %), on peut émettre l'hypothèse que ce taux d'erreur sur les bases est lié à une réussite par défaut des désinences. En effet, si les désinences sont relativement régulières au présent, de nombreux verbes utilisent plusieurs bases qui provoquent de fréquentes erreurs chez les élèves (doublement des consonnes - appelle vs appeler -, changement d'accent - cède vs céder -, apparition de la cédille - plaçons vs *placer -, etc.). À l'inverse du présent, l'imparfait est un temps très régulier au niveau des bases mais les problèmes d'homophone-hétérographie avec les finales en /E/ provoquent un taux d'erreur important sur les désinences.

Le traitement de l'adjectif

Avec un total de 9 273 formes, les adjectifs représentent 3,78 % des formes de notre corpus alors qu'à titre de comparaison, les adjectifs représentent 6,96 % des formes de Manulex CP-CM2⁷. Cette catégorie grammaticale semble donc peu mobilisée par les élèves et ce, quel que soit le niveau : d'un minimum de 3,55 % en CE2 à un maximum de 4,51 % au CP. Il est à noter que la consigne du CP demandant aux élèves de raconter l'histoire d'un petit chat implique un suremploi de l'adjectif « petit » (près de 76 % des adjectifs mobilisés) faussant ainsi l'étude comparative. De ce fait, il ne sera pas tenu compte du niveau CP dans cette étude.

La réussite orthographique s'améliore d'un niveau à l'autre passant de 52,17 % de réussite en CE1 à 84,84 % en 3^e. On note toutefois une stagnation des progrès entre le CM2 (75,62 % de réussite) et la 6^e (70,50 %). Il est difficile d'expliquer cette stagnation sans une étude plus précise. Toutefois, on peut rappeler que les textes de CE1-CM2 sont produits à partir d'une même consigne sur la même cohorte d'élèves alors que les textes de 6^e ont été produits à partir de consignes diverses avec des élèves différents.

À l'aide de programmes de traitement automatique des langues, nous avons opéré un découpage des formes adjectivales en base, flexion de genre et flexion de nombre. Ce découpage nous a permis de distinguer les erreurs portant sur la base de celles portant sur les différentes flexions. Par exemple, la forme produite « meilleures » produite en CM2 est décomposée en (meilleur + e + s ; fem/plu) alors que la forme attendue « meilleurs » est décomposée en (meilleur + _ + s ; mas/plu). La comparaison de ces deux découpages indique une erreur de type lexical (bases différentes), pas d'erreur sur le nombre et une erreur sur le genre (masculin attendu).

⁷ <http://www.manulex.org/fr>

Comme le montre la figure 4, si la graphie des bases lexicales des adjectifs s'améliore progressivement, les erreurs restent plus élevées que sur les flexions. En 6^e, par exemple, 16,70 % des adjectifs présentent encore une erreur de type lexical. Notons qu'une nette amélioration s'opère au collège puisqu'en 3^e, les erreurs lexicales ne représentent plus que 7,21 %. La stagnation des résultats entre CM2 et 6^e, observée également sur les verbes, se retrouve sur ce point car les erreurs sur les bases lexicales passent de 14,21 % en CM2 à 16,70 % en 6^e.

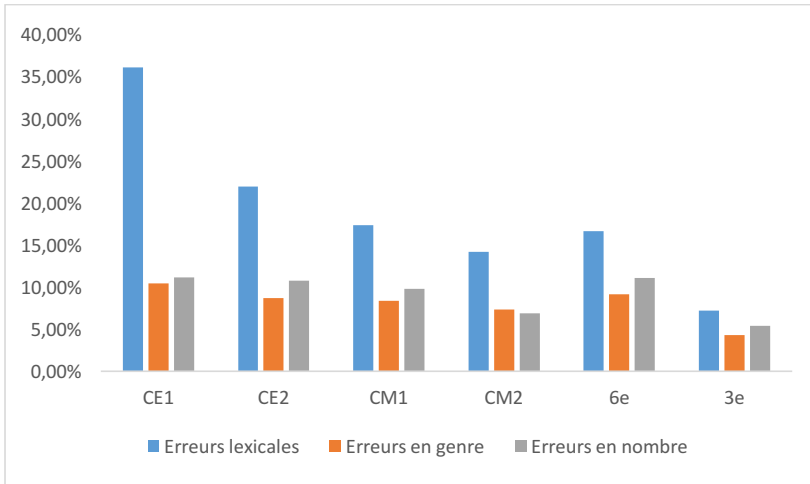


Figure 4. Évolution des types d'erreurs

Concernant les accords en genre ou en nombre, on remarque une réussite supérieure sur les valeurs par défaut (masculin pour le genre, singulier pour le nombre) qui ne nécessitent pas de marques flexionnelles. La figure 5 donne les pourcentages de réussite des adjectifs selon le nombre attendu (singulier ; pluriel ; non marqué, par exemple « gris ») suivant que l'adjectif est masculin, féminin ou épïcène (par exemple « agréable »). Tous niveaux confondus, le pluriel est la marque qui pose le plus de difficulté aux élèves puisque pour les adjectifs attendus au pluriel, seuls 38,13 % sont corrects. Comme le fait apparaître la figure 5, cette difficulté d'accorder les adjectifs au pluriel se retrouve que le genre soit marqué ou non. Les réussites sur les adjectifs attendus au singulier ou non marqués en nombre (ex. « gros ») sont, elles, nettement supérieures (entre 66,60 % et 82,76 %).

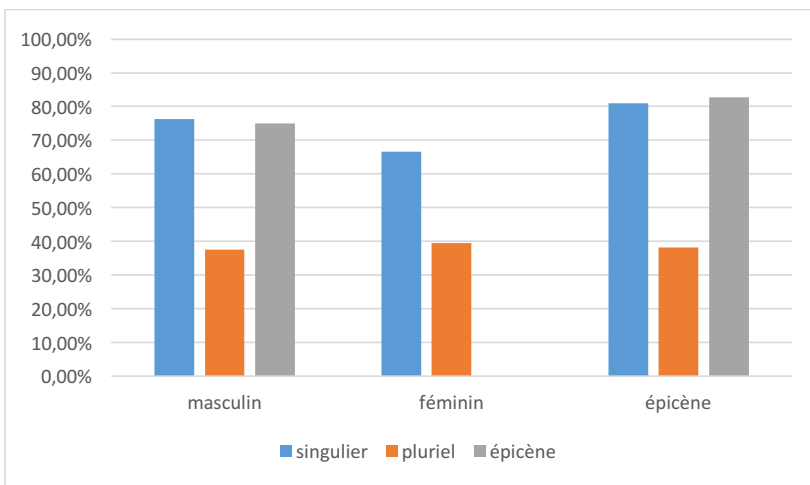


Figure 5. Réussites selon le nombre attendu

Perspectives didactiques

Le corpus de textes présenté permet la mise en évidence d'une part des progrès continus des élèves de l'école élémentaire en orthographe en production d'écrit, d'autre part des zones du système orthographique qui restent longtemps en construction, tel le marquage morphographique des catégories du verbe et de l'adjectif. Ainsi, les résultats obtenus permettent de mieux cibler les tâches proposées aux élèves et d'effectuer des choix sinon rentables du moins raisonnables. Par exemple, concernant la segmentation, la suite *pronom + verbe* serait à travailler en fin de CP et en CE1, si possible en lien avec les activités grammaticales, métalangage à l'appui, afin que ce dernier devienne un point d'appui durable à la réflexion grammaticale, notamment pour les élèves les moins avancés (Jarno *et al.*, 2019 ; Le Levier *et al.*, 2018). Ce travail pourrait avantageusement remplacer les trop nombreux exercices à trous qui portent sur les traditionnels « homophones » (par exemple *la* (l'a), *sest* (s'est), *cerèvei* (se réveillent) (Jarno *et al.*, 2019).

Concernant le verbe, les tiroirs verbaux les plus employés sont en nombre limité et peuvent être décrits à l'aide d'un nombre non moins limité de désinences (un seul modèle d'imparfait fonctionne avec tous les verbes). Si l'on a pu mettre en évidence des progrès pour les quatre tiroirs verbaux les plus employés, l'imparfait et surtout le passé simple sont ceux pour lesquels ces progrès sont les plus lents. Si la gestion des bases lexicales s'améliore rapidement (de 54,51 % d'erreurs au CP à 11,59 % en 6^e), la gestion des désinences verbales reste une difficulté persistante puisqu'elle représente encore 23,98 % d'erreurs en 6^e.

Pour le passé simple, les résultats invitent à travailler avec les élèves non seulement sur la désinence du passé simple mais aussi sur la base verbale, les deux éléments étant clairement impliqués dans la production de l'erreur : ainsi, la mise en évidence du processus de fabrication de ces formes à partir de la base verbale la plus fréquente et de la désinence d'un autre groupe de verbes devrait aider les élèves à mieux comprendre leurs erreurs. Concernant l'imparfait, les résultats obtenus suggèrent de consacrer du temps aux choix graphiques des élèves pour les désinences de ce tiroir verbal et de réfléchir avec les élèves à l'homophonie-hétérographie des formes verbales en /E/ : l'imparfait est un temps très régulier dans sa construction mais il entre dans un jeu d'opposition avec l'infinitif et le participe passé dont la prononciation est très proche dans la francophonie (Brissaud, Negro et Fisher, 2012). L'homophonie-hétérographie appliquée à des unités entières à l'intérieur d'une même classe grammaticale, le verbe, pose des problèmes graphiques redoutables et durables aux élèves comme aux adultes. Il convient donc également, avec de jeunes élèves, de travailler sur le fonctionnement et la reconnaissance du verbe, ainsi que sur la constitution progressive d'un répertoire de désinences verbales.

L'adjectif apparaît comme une catégorie grammaticale peu mobilisée par les élèves. La réussite orthographique s'améliore cependant d'un niveau à l'autre avec des erreurs persistantes sur la base. En ce qui concerne les accords en genre ou en nombre, on remarque une réussite supérieure quand l'adjectif est attendu au masculin ou au singulier et ne nécessite pas de marques flexionnelles (il s'agit ici en quelque sorte d'une réussite par défaut). La difficulté à accorder les adjectifs au pluriel semble bien réelle, que le genre soit clairement le masculin, le féminin, ou que l'adjectif soit épiciène. Cette difficulté a déjà été précisée par Cogis et Brissaud (2019) à l'aide d'un test standardisé proposé en fin de scolarité primaire. Elle perdure en production de texte chez des élèves plus âgés, par exemple les lycéens professionnels (Mout et Brissaud, 2017).

Le travail présenté ici permet aussi de distinguer les phénomènes récurrents à un niveau donné de ceux qui le sont moins et de repérer les élèves les moins avancés, par exemple en observant des difficultés parfois sous-estimées à l'école, comme la segmentation en mots.

Références bibliographiques

- Andreu, S. et Steinmetz, C. (2016). Les performances en orthographe des élèves en fin d'école primaire (1987-2007-2015). *Note d'information*. n° 16.28. MENESR-DEPP.
- Brissaud, C., Negro, I. et Fisher, C. (2012). The relation between spelling and pronunciation: the case of French /E/ verbal endings, *Written Language and Literacy*, 15(1), 46-64. DOI : <https://doi.org/10.1075/wll.15.1.03bri>
- Brissaud, C., Viriot-Goeldel, C. et Ponton, C. (2019). Enseigner et apprendre l'orthographe avec la « Twictée ». Premiers résultats de l'évaluation d'un dispositif innovant d'enseignement de l'orthographe, *Repères*, 60, 107-130. DOI : <https://journals.openedition.org/reperes/2346>
- Cogis, D. et Brissaud, C. (2019). À la poursuite des marques de genre... Dans C. Mortamet (dir.) *L'orthographe : pratiques d'élèves, pratiques d'enseignants, représentations* (p. 43-71). Presses universitaires de Rouen.
- Doquet, C., Enoiu, V., Fleury, S. et Maziotti, S. (2017). Problèmes posés par la transcription et l'annotation d'écrits d'élèves. *Corpus*. 16. DOI : <https://doi.org/10.4000/corpus.2776>
- Elalouf, M.-L. (2005). Écrire entre 10 et 14 ans un corpus, des analyses, des repères pour la formation. Paris : Canopé - CRDP de Versailles.

- Elalouf, M.-L. (2011). Constitution de corpus scolaires et universitaires : Vers un changement d'échelle ? *Pratiques*, 149-150, 56-70. DOI : <https://doi.org/10.4000/pratiques.1702>
- Ferreiro, E. et Pontecorvo, C. (1996). I confini tra le parole. Dans E. Ferreiro *et al* (éd.), *Cappuccetto Rosso impara a scrivere: studi psicolinguistici in tre lingue romanze* (p. 39-77). Firenze: La Nuova Italia.
- Fisher, C. et Nadeau, M. (2014). Usage du métalangage et des manipulations syntaxiques au cours de dictées innovantes dans des classes du primaire. *Repères*, 49, 169-191. DOI : <https://doi.org/10.4000/reperes.742>
- Garcia-Debanc, C., Ho-Dac, L.-M., Bras, M. et Rebeyrolle, J. (2018). Vers l'annotation discursive de textes d'élèves. *Corpus*, 16, 157-184. DOI : <https://doi.org/10.4000/corpus.2783>
- Goigoux, R. (éd.) (2016). *Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages*. Institut Français de l'Éducation. <http://ife.ens-lyon.fr/ife/recherche/lire-ecrire/rapport/rapport-lire-et-ecrire>.
- Ho-Dac, M., Fleury, S. et Ponton, C. (2020). E:Calm Resource: a Resource for Studying Texts Produced by French Pupils and Students. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Marseille. 4327-4332.
- Jarno El Hilali, G., Nadeau, M. et Fisher, C. (2019). L'effet des dictées métacognitives-interactives sur la compétence à orthographier les homophones grammaticaux en rédaction. *Repères*, 60, 45-63. DOI : <https://doi.org/10.4000/reperes.2291>
- Lavieu-Gwozdz, B., Vinel, É., Goossens, V. et Brissaud, C. (2021). Cartographie des usages et des erreurs orthographiques sur les verbes dans des récits écrits par des élèves de 6 ans à 15 ans. *Langue Française*, 211, 51-65.
- Le Levier, H., Brissaud, C. et Huard, C. (2018). Le raisonnement orthographique chez des élèves de troisième : analyse d'un corpus d'entretiens métagraphiques. *Pratiques*, 177-178. DOI : <http://journals.openedition.org/pratiques/4464>
- Manesse, D. et Cogis, D. (2007). *L'orthographe, à qui la faute ?* Paris : ESF éditeur.
- MEN-DEP (1996). Connaissances en français et en calcul des élèves des années vingt et d'aujourd'hui. *Les Dossiers d'Éducation et Formations*. 62.
- Mout, T. et Brissaud, C. (2017). Ce qu'ils savent faire et moins bien faire : le cas de l'orthographe. Dans M.-C. Guernier, C. Barré-De Miniac, C. Brissaud et T. Mout, *Ces lycéens en difficulté avec l'écriture et l'école* (p. 77-114). Grenoble : UGA Éditions.
- Piacente, T. et Querejeta, M. (2012). La separación entre palabras en la escritura infantil. *Revista Neuropsicología Latinoamericana*, 4(1), 1-17.
- Ponton, C., Gutiérrez-Cáceres, R., Teruggi, L., Farina, E., Brissaud, C. et Wolfarth, C. (2021). Scolinter : un corpus trilingue. L'exemple de la segmentation en mots. *Langue Française*, 211, 37-50.
- Querejeta, M., Piacente, T., Guerrero, O. *et al.* (2013), La separación entre palabras en la escritura de niños que inician la escolaridad primaria, *Interdisciplinaria*, 30(1), 45-64. DOI : <http://dx.doi.org/10.16888/interd.2013.30.1.3>
- Ugarte, A. et Díaz Argüero, C. (2017). ¿Se escribe junto o separado? Evolución en el uso de los espacios en blanco en textos de niños que inician su alfabetización. *Revista Signos*, 50(93), 3-25. DOI : [10.4067/S0718-09342017000100001](https://doi.org/10.4067/S0718-09342017000100001)
- Wolfarth, C., Brissaud, C. et Ponton, C. (2018a). Transcrire et normer un corpus scolaire : pour quelles analyses ? *Dyptique*, 36, Presses Universitaires de Namur, 121-145
- Wolfarth, C., Ponton, C. et Brissaud, C. (2018b). Gestion de la morphographie verbale en production d'écrits : que peut nous apprendre un corpus longitudinal *Repères*. 57, 209-226. DOI : <https://doi.org/10.4000/reperes.1576>