

## Usage d'un corpus longitudinal, le cas de la morphologie verbale

Wolfarth, C., Ponton, C., Brissaud, C.

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

Claire.Wolfarth@univ-grenoble-alpes.fr,  
Claude.Ponton@univ-grenoble-alpes.fr,  
Catherine.Brissaud@univ-grenoble-alpes.fr

**Résumé :** L'acquisition de la morphologie verbale a fait l'objet de nombreux travaux psycholinguistiques depuis une trentaine d'années qui ont permis de comprendre la dynamique des acquisitions. Cependant, rares sont les travaux qui ont cherché à décrire la gestion des accords verbaux en production d'écrit. À partir de l'étude de la morphologie verbale, nous montrons en quoi les corpus permettent de mieux connaître les compétences réelles et l'évolution des compétences des élèves en classe de CP, CE1 et CE2. Nous exposons la chaîne de traitement qui permet d'aboutir à l'extraction des formes verbales, à partir de productions manuscrites des élèves. Puis, nous analysons toutes les formes verbales obtenues afin de mettre en évidence les progrès des élèves.

**Abstract:** The acquisition of verbal morphology has been extensively researched and documented for thirty years. This work has led to understand acquisition process. Nevertheless, very little research aimed to describe how French students managed verbal inflexions in writing. From the study of verbal morphology, we show how corpora help to learn more about actual skills and how evolved learner skills between first, second and third grade of primary school. We expose the processing chain that allows the verbal forms extract from young students' writings. Then, we analyze every verbal forms to bring out students' progress.

**Mots-clés :** enseignement, école primaire, écriture, grammaire

**Keywords:** teaching, primary school, writing, grammar

Il est maintenant bien établi que l'apprentissage de l'orthographe du français nécessite au moins une dizaine d'années et que la classe des verbes est une catégorie particulièrement difficile à acquérir à l'écrit tant elle est le siège de nombreux phénomènes d'homophonie-hétérographie. Cela pose encore problème à de nombreux adultes, qu'ils soient peu scolarisés ou inscrits dans le supérieur (Blondel, Brissaud et Rinck, 2016 ; Lucci et Millet, 1994 ; David et Laborde-Milaa, 2002). Si les travaux consacrés à la morphologie verbale ont été nombreux depuis une trentaine d'années, tant en psychologie cognitive qu'en linguistique ou en didactique, ces acquisitions ne sont pas complètement décrites. Les études conduites sont en effet consacrées en général à un aspect du système verbal : l'homophonie verbo-nominale (*marche/marches/marchent*) et le coût du contrôle orthographique, les formes verbales en /E/ (*marché, marcher, marchait*, etc.) ou l'accord du participe passé. Nous manquons donc encore d'une vue panoramique des acquisitions, notamment de la façon dont elles sont mobilisées en production de texte, quand les ressources cognitives sont accaparées par d'autres postes que celui de l'orthographe et de son contrôle. Cet article se propose donc de contribuer à une compréhension globale des acquisitions en se centrant sur les débuts de cet apprentissage.

L'enjeu est bien ici de comprendre comment les premiers apprentissages formels se mettent en place, comment les savoirs sont mis en œuvre en production écrite afin de mieux accompagner les apprentissages par des propositions didactiques adaptées.

Cette contribution est focalisée sur les apprentissages orthographiques en production de texte dans le domaine du verbe du CP au CE2. Au regard des spécificités de notre corpus, nous nous concentrerons uniquement sur les tiroirs verbaux<sup>1</sup> suivants : infinitif, présent de l'indicatif, imparfait, passé simple et participe passé. En raison de l'usage d'outils de traitement informatiques, l'unité d'étude adoptée est le mot, ou plus exactement les chaînes de caractères isolées par des espaces ou des signes de ponctuation. Pour l'objet qui nous concerne, à savoir le verbe, nous nous intéresserons donc aux mots verbaux. Les temps composés ne seront pas traités comme des temps à part entière mais comme des combinaisons d'un auxiliaire à un temps simple et d'un participe passé, par exemple pour le passé composé un auxiliaire au présent et un participe passé.

### Quelle théorie du verbe ?

La tradition scolaire, du moins jusqu'aux programmes de 2008, classe les verbes en trois groupes selon leur terminaison infinitive et distingue radical et terminaison. Cette classification, héritée du latin, qui distingue les verbes terminés par -er à l'infinitif (sauf parfois le verbe *aller*), les verbes terminés par -ir qui ont une base longue en -iss- et puis tous les autres verbes, qui ne rentrent pas dans les deux premières catégories, et qui finissent par -ir, -oir, -re. En raison du grand nombre de modèles de conjugaison qu'engendre cette répartition des verbes, elle ne nous paraît pas opérationnelle pour l'étude de notre corpus, et plus particulièrement pour les traitements informatiques que l'on souhaite y appliquer. En effet, une telle conception de la flexion verbale entraîne une parcellisation des verbes en près d'une centaine de modèles de conjugaison, de 82 pour *Le nouveau Bescherelle, l'art de conjuguer* (1980) à 140 pour le *Conjugaison*, Le Robert & Nathan (1996) (Brissaud, 2002).

À la suite de nombreux linguistes (par exemple Martinet, 1979 ; Meleuc et Fauchart, 1999, Blanche-Benveniste, 2002 ; Pellat, 2009), nous optons pour un classement en deux catégories : les verbes en -er (dont l'infinitif se termine par /e/ à l'oral) et les autres verbes (dont l'infinitif se termine par /R/ à l'oral) et nous proposons une entrée via les désinences (ou terminaisons). L'objectif de cette approche est de générer des modèles de désinences opérationnels pour le plus grand nombre de verbes. Selon cette conception, la conjugaison se réduit donc à un nombre restreint de modèles de désinences, résumés pour l'indicatif dans le tableau ci-dessous.

Temps	Types de verbe	P1	P2	P3	P4	P5	P6
Présent de l'indicatif	Verbes en -er	-e	-es	-e	-ons	-ez	-ent
	Autre verbes	-s	-s	-t (ou Ø)			

<sup>1</sup> Un tiroir verbal est l'ensemble des formes d'un verbe à un temps et un mode donné. Par exemple, le présent de l'indicatif est un tiroir verbal.

Imparfait	Tous les verbes	-ais	-ais	-ait	ions	-iez	-aient
Passé simple	Verbes en -er	-ai	-as	-a	- <sup>^</sup> mes	- <sup>^</sup> tes	-rent
	Autre verbes	-(v)s	-(v)s	-(v)t			

Tableau 1 : désinences du présent de l'indicatif, de l'imparfait et du passé simple<sup>2</sup>

Ainsi, pour l'imparfait, nous considérons que tous les verbes à la deuxième personne de l'imparfait de l'indicatif ont pour désinence –ais, et tout ce qui vient avant cette désinence est considéré comme appartenant à sa base (ou radical), un même verbe pouvant posséder plusieurs bases (jusqu'à sept en se fondant sur l'oral, hors passé simple, infinitif et participe passé ; Dubois, 1967). Nous considérons donc qu'il n'y a pas de particularité de désinences pour les verbes du deuxième groupe, contrairement à certaines conceptions issues d'une certaine tradition grammaticale, qui distingue les verbes comme fin-ir / fin-issais des verbes comme ven-ir / ven-ais (par exemple dans le *Bescherelle, la conjugaison pour tous*, 1997). Pour le passé simple, nous considérons qu'il n'y a qu'un seul modèle de conjugaison pour les personnes dites du pluriel, qui se décline selon 4 voyelles données par la base du verbe (notées v entre parenthèses : a, i, u, in). Pour le présent, nous ne considérerons que deux modèles, avec une exception pour la 3<sup>e</sup> personne, où le –t peut être assimilé par la lettre finale de la base –d.

Le tableau 2 précise les différentes marques pour le mode participe passé :

Types de verbe	Masculin Singulier	Féminin Singulier	Masculin Pluriel	Féminin Pluriel
Verbes en -er	-é	-ée	-és	-ées
Verbes en -oir, -oire, -ure, -dre, -aire, -aitre, certains verbes en -ire et -ivre	-u	-ue	-us	-ues
Verbes en -ir, certains verbes en -ire et -ivre	-i	-ie	-is	-ies
Verbes en -indre et un petit nombre de verbes, comme faire, dire, écrire, ouvrir, mourir, cuire, conduire, frire, traire	-t	-te	-ts	-tes
Un petit nombre de verbes, comme <i>assoir, prendre et mettre</i>	-s	-se	-s	-ses

Tableau 2 : les différentes marques du mode participe passé

<sup>2</sup> L'absence de marque pour la troisième personne du présent de l'indicatif ne concerne que certains verbes terminés en -dre.

En français, les participes passés se terminent majoritairement par une voyelle « nue », sans consonne. Les participes passés en -s et -t sont relativement peu nombreux mais ils concernent certains verbes très fréquents comme *faire* et *dire*.

## 1. Méthodologie

### a. Caractérisation du corpus

Le corpus sur lequel porte cette étude a été recueilli sur trois années consécutives du CP au CE2 auprès des mêmes 301 élèves, soit un total de 903 productions. Il est extrait du corpus recueilli dans le cadre du projet *Scoledit*<sup>3</sup> (Wolfarth, Ponton, Totereau, 2017). Au cours de ces trois années de recueil, un même type de consigne a été donné à l'ensemble des élèves à chaque fin d'année scolaire. En CP, les élèves devaient raconter l'histoire d'un petit chat à partir de quatre vignettes où on le voit s'éloigner, tomber, pleurer et être recherché par la maman chat (annexe 1). En CE1 et CE2, les élèves devaient choisir un ou deux personnages parmi les quatre proposés et raconter une histoire (annexe 2).

Les 903 productions recueillies comptent 63 197 mots. Le tableau 3 précise le nombre total de mots rassemblés à chaque niveau de scolarité, la longueur moyenne des textes produits par les élèves ainsi que la taille du texte le plus court et celle du plus long.

	CP	CE1	CE2	Tous niveaux (CP-CE2)
Nombre de productions	301	301	301	903
Nombre de mots	7 877	20 125	35 195	63 197
Longueur moyenne de production	26,17	133,72	233,85	131,25
Nombre de mots maximal	62	422	696	696
Nombre de mots minimal	0	16	46	0

Tableau 3 : Données générales concernant le corpus recueilli

Avant de nous intéresser plus précisément aux verbes, nous pouvons nous pencher rapidement sur la répartition des catégories grammaticales à l'intérieur des corpus. Les fréquences des 8 catégories principales sont résumées dans le tableau 4 ci-dessous.

Catégories	Fréquence en fin de CP	Fréquence en fin de CE1	Fréquence en fin de CE2	Fréquence totale
Nom	22,9%	21,3%	20,7%	21,1%

<sup>3</sup> Ce travail est soutenu par *Démarre SHS !* de l'Institut des Données de Grenoble et a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'avenir » portant la référence ANR-15-IDEX-02.

Verbe	21,3%	20,3%	20,7%	20,6%
Pronom	15,1%	14,8%	15,4%	15,2%
Déterminant	17,2%	16,5%	15,1%	15,8%
Préposition	5,8%	8,8%	8,7%	8,4%
Conjonction	7,7%	6,5%	6,3%	6,5%
Adverbe	3,5%	6,3%	6,1%	5,9%
Adjectif	5,4%	3,8%	4,1%	4,1%
Autres catégories	1,3%	1,7%	2,9%	2,3%

Tableau 4 : Répartition des mots du corpus selon les classes grammaticales

La lecture des deux premières lignes du tableau 4 permet d'observer que les catégories du nom et du verbe sont les plus fréquentes et qu'il y a à peu près autant de noms que de verbes dans ces écrits ; elles sont suivies par celles du pronom et du déterminant (lignes 3 et 4). On remarque enfin la faible fréquence des adjectifs (4,1% au total, avec de faibles variations du CP au CE2). En comparaison, dans le corpus littéraire Frantext, un mot sur quatre est un verbe et plus de la moitié est un verbe en -er (Brunet, 1999).

### Chaîne de traitement

Au vu de l'ampleur du corpus recueilli, l'analyse ne peut se faire que guidée par des outils de traitement automatique. C'est pourquoi les données sur lesquelles reposent les analyses présentées dans cet article résultent d'une chaîne de traitement relativement longue qui a permis leur extraction à partir des productions initiales. Nous proposons d'en établir un bref résumé ici.

Les productions recueillies étant écrites manuellement par les élèves, nous avons dû en faire une transcription. Pour ce faire, nous avons mis en place une convention de transcription guidée par le choix des caractéristiques des productions que nous souhaitons conserver. Nous avons ainsi gardé certains marqueurs visuels, comme les fins de ligne, pour pouvoir restituer une partie de la mise en page de l'élève. Les traces de révisions, ratures, ajouts, etc. ont également été relevées, de même que les passages illisibles. Bien évidemment, les choix orthographiques et typographiques de l'élève ont été reproduits fidèlement.

À l'issue de cette étape, il nous est possible de faire correspondre le scan du texte de l'élève avec la transcription, mais peu de traitements informatiques sont possibles. En effet, les outils informatiques échouent à reconnaître des formes qui ne sont pas orthographiquement normées. Par exemple, dans l'extrait de production « le chas va fére une aventure sapormére aventure [...] » (élève 3051, CP), seules les formes *le*, *va* et *une* pourront être identifiées, c'est-

à-dire reconnues comme des formes du français auxquelles on peut attribuer un lemme<sup>4</sup> et une catégorie grammaticale ; la forme *chas* va également être identifiée comme un nom de la langue française, en référence au nom *chas*, dont la définition est « trou d'une aiguille ».

Pour pallier ce problème, ou tout au moins pour proposer une première piste de solution, nous avons saisi aussi une version « normalisée » de chacun des textes. Cette normalisation représente une version parallèle de la production dans laquelle certains traitements ont été effectués, comme le rétablissement des normes orthographiques et morphologiques. Pour reprendre l'exemple précédent, pour la production de l'élève 3051 en CP « le chas va fére une aventure sapormére avanture [...] », nous aurons la normalisation suivante : « Le chat va faire une aventure sa première aventure [...] ». La comparaison de ces deux versions à l'aide d'un algorithme d'alignement permet d'identifier la forme *chas* comme une variante orthographique de la forme *chat* et la forme *sapormére* comme étant une variante orthographique des formes *sa* et *première*.

L'algorithme d'alignement qui permet d'obtenir ce résultat a été élaboré par nos soins à partir de la comparaison forme à forme de la transcription et de la normalisation. C'est pourquoi, il est nécessaire que les transformations opérées lors de l'étape de normalisation soient minimales.

Les comparaisons sont réalisées à partir d'indices phonologiques. Les productions transcrites et les versions normalisées ont donc été converties en représentations phonologiques, c'est-à-dire transcrites à l'aide de l'alphabet phonétique international (API). Pour ce faire, nous nous sommes appuyés sur l'outil LIA\_PHON développé par F. Béchet (2001) Si nous reprenons l'exemple précédent, nous obtenons les chaînes de phonèmes suivantes : / lə ʃa va fer yn avãtyr sapɔrmer avãtyr / et / lə ʃa va fer yn avãtyr sa prãmjer avãtyr /. Suite à cette étape, nous obtenons pour chaque forme transcrite et normée sa représentation graphique et sa représentation phonologique. Afin d'améliorer encore le résultat de l'alignement, outre la représentation phonologique, nous calculons également une représentation archiphonologique. Nous empruntons cette notion à N. Catach (1980). Dans cette représentation, nous neutralisons les oppositions /ø/, /œ/ et /ɔ/ ; /e/ et /ɛ/ ; /o/ et /ɔ/ ; /ẽ/ et /œ̃/. Pour résumer, pour la forme *faire* nous disposons des trois représentations suivantes :

Représentation graphique	Représentation phonologique	Représentation archiphonologique
faire	/fɛr/	/fɛr/

Tableau 5 : les différentes représentations d'une forme, l'exemple de faire

Ces représentations sont autant d'indices pour l'alignement des transcriptions et des normalisations. En effet, le fonctionnement de l'algorithme d'alignement est basé sur la

---

<sup>4</sup> Forme graphique, considérée par convention comme non fléchie, choisie comme entrée dans un dictionnaire ou dans un lexique.

comparaison des formes transcrites et des formes normées à partir de chacune de ces représentations.

Le résultat de cet alignement nous permet d'établir plusieurs niveaux de réussite (voir Figure 1) selon que les représentations graphique, phonologique ou archiphonologique sont, ou non, identiques. Ainsi, dans le cas où les représentations graphiques sont identiques, comme pour les formes *le*, *va* et *une* de l'exemple précédent, elles sont dites « normées ». Si les représentations phonologiques ou archiphonologiques sont comparables, comme les formes *chas* et *chat* (niveau phonologique) et les formes *fére* et *faire* (niveau archiphonologique), les formes produites par l'élève sont dites « non normées à phonologie normée ». Si aucune de ces représentations ne sont comparables, les formes produites sont dites « non normées ». C'est le cas des formes *pormére* (après resegmentation de *sapormére*) et *première*.

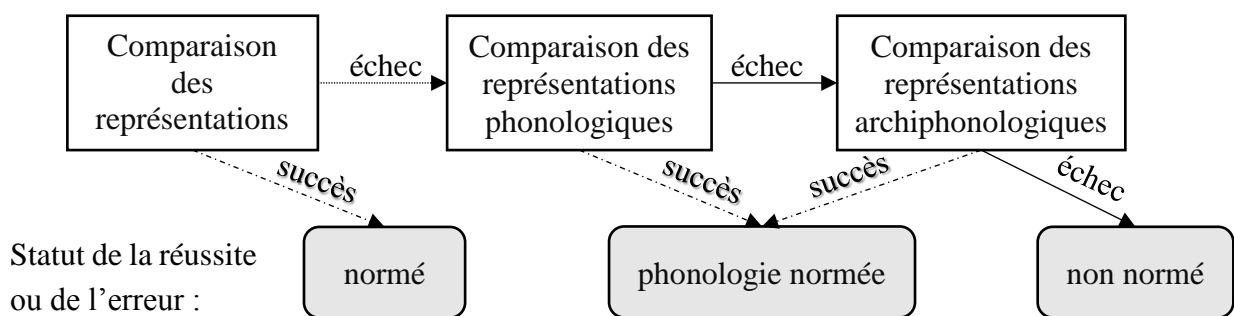


Figure 1 : Procédure de catégorisation des formes selon les échecs et les réussites

À la fin de l'alignement, nous avons utilisé l'étiqueteur morphosyntaxique TreeTagger (Schmid, 1995) pour attribuer une catégorie grammaticale à chaque forme du corpus. Cette étape nous a permis d'extraire les verbes aux modes et aux temps souhaités pour l'étude présentée ici.

Chaque traitement, informatique ou humain, engendre un certain nombre d'erreurs. La chaîne de traitements qui, à partir d'une production manuscrite amène à l'identification des verbes, de leur radical et de leur désinence, de leur personne, temps et mode, ainsi que du type d'échecs ou de réussite qu'ils présentent, est longue. Chaque étape est une nouvelle source d'erreurs. Afin d'exploiter les données qui en résultent, il est donc nécessaire d'évaluer cette marge d'erreur sur l'ensemble de la chaîne de traitement. Une telle évaluation est longue et très coûteuse. Nous l'avons amorcée mais nous ne sommes pas encore en mesure d'en fournir les résultats définitifs. C'est pour cela que dans cet article nos analyses ne s'appuient que sur des grandes tendances qui restent relativement fiables.

## Analyses

Comme nous l'avons mentionné précédemment, parmi tous les résultats et pistes d'analyses offertes par la chaîne de traitement décrite précédemment, nous avons choisi de nous pencher sur l'usage des verbes en début d'apprentissage de l'écriture.

Étant donné la faible fréquence de certains tiroirs verbaux et du fort taux d'erreurs qu'ils engendraient via les traitements automatiques, nous avons choisi de nous concentrer sur les cinq



tiroirs verbaux les plus fréquents : imparfait de l'indicatif, infinitif, participe passé, présent de l'indicatif et passé simple.

#### Répartition des tiroirs verbaux

La figure 2, ci-dessous, précise la répartition des temps des verbes rencontrés selon le niveau de scolarité.

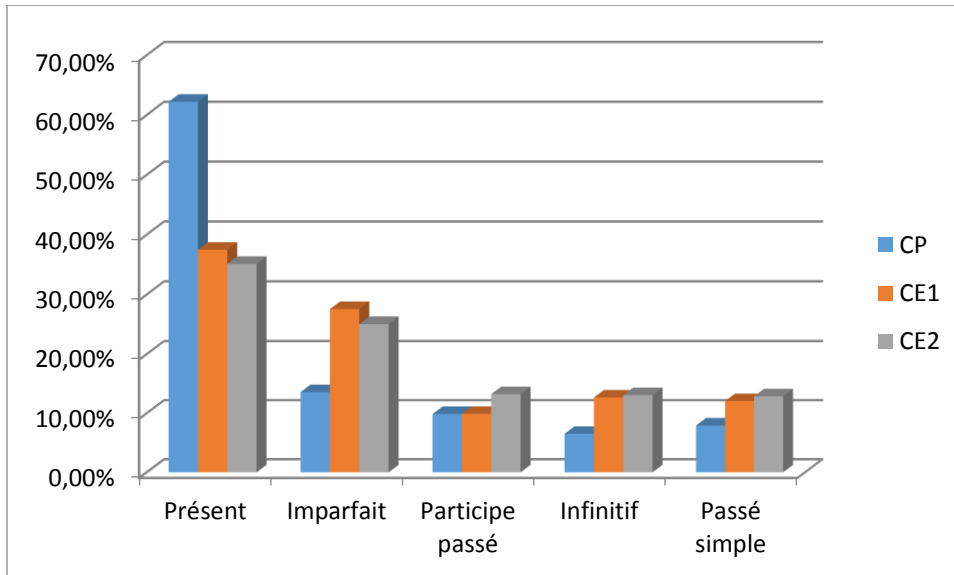


Figure 2: Répartitions des temps et modes verbaux selon le niveau de scolarité

La figure 2 permet de mettre en lumière la répartition des temps verbaux que nous avons choisi d'étudier selon l'année de recueil. Nous noterons cependant que les consignes de production étaient différentes l'année de CP et les années suivantes ; il est possible que la consigne donnée en CP ait amené une prédominance de verbes au présent, représenté à 62% contre 35% et 37% respectivement pour le CE1 et le CE2, ce qui conduit, pour ces deux niveaux, à une plus faible représentation des autres temps.

Au-delà de ces deux tiroirs verbaux particuliers, la répartition semble être relativement constante au cours des années, avec une prédominance de l'imparfait et du présent. On constate une présence plus importante de verbes au participe passé au CE2 qui peut laisser penser, nos données se basant sur les mots verbaux et non sur les formes verbales, qu'il s'agit en réalité de passé composé. Ces données convergent avec celles de Savelli et al. (2002) concernant la fréquence d'usage des tiroirs verbaux : dans leur étude sur le passé simple conduite auprès d'élèves de 3-6 ans (à l'oral) et de classe de 6<sup>e</sup> (à l'écrit), les tiroirs verbaux les plus utilisés dans le récit sont le passé simple (4,5% vs 34,2%), l'imparfait (13,8% vs 19,2%), le présent (61,5% vs 27,8%), et le passé composé (12,3% vs 9,1%).

#### Répartition des erreurs

Comme nous l'avons évoqué précédemment, les indices utilisés dans la phase d'alignement nous permettent de répartir les formes en trois catégories : normées (formes correctes), non normées à phonologie normée (forme incorrecte mais qui permet de restituer la forme sonore), non normées (forme incorrecte ne permettant pas de restituer la forme sonore). En appliquant cette classification aux verbes, nous obtenons les résultats suivants :



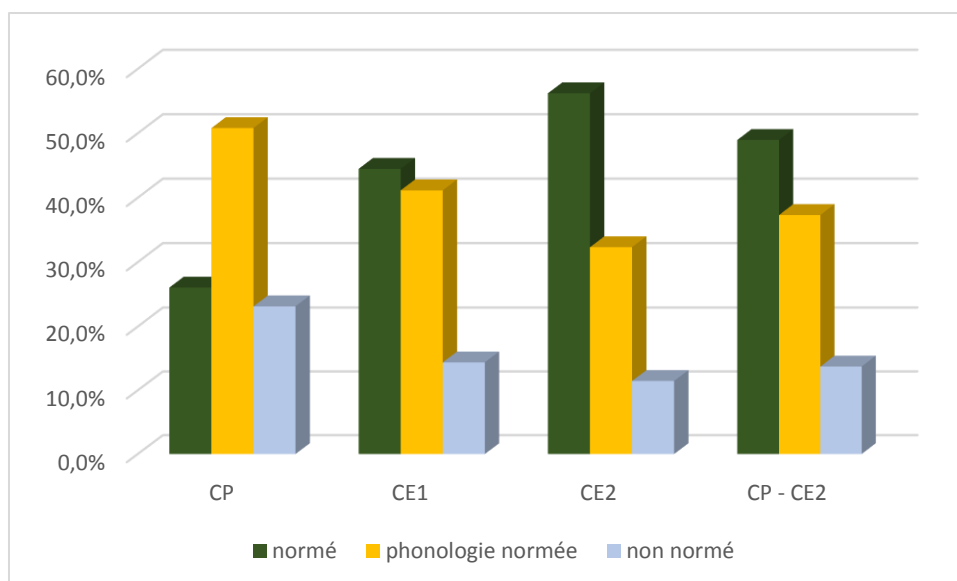


Figure 3. Répartition des erreurs selon l'année d'apprentissage

Dès la fin du CP, on note une maîtrise importante des formes phonologiques des verbes puisque seules 23% des formes verbales écrites ne respectent pas la phonologie. En revanche, l'orthographe des verbes est encore peu maîtrisée dans la mesure où seuls 26% des verbes sont correctement orthographiés. On note d'importants progrès en CE1 et en CE2 sur ce plan-là, puisque en fin de CE2 la majorité des verbes (56%) sont correctement orthographiés. De plus, en fin de CE2, les formes non normées représentent moins de 12% des verbes. Il y a donc une forte progression du CP au CE2, visible dans la figure 3 dans l'évolution des tailles des colonnes du CP au CE2.

Pour plus de détails, il nous est également possible d'obtenir une classification des verbes des cinq tiroirs verbaux qui nous intéressent. Les premières lignes du tableau 6 reprennent les données exposées dans la figure précédente (figure 3).

Niveau		CP	CE1	CE2
<b>Nb formes verbales</b>		<b>921</b>	<b>2792</b>	<b>4638</b>
<b>Ensemble des verbes</b>	normé	26,05%	44,48%	56,21%
	phonologie normée	50,81%	41,14%	32,32%
	non normé	23,14%	14,38%	11,47%
<b>PRESENT</b>	normé	30,30%	52,86%	64,64%
	phonologie normée	49,04%	37,40%	27,38%
	non normé	20,67%	9,45%	7,98%
<b>INFINITIF</b>	normé	30,00%	49,29%	63,20%
	phonologie normée	41,67%	34,28%	25,08%
	non normé	28,33%	16,43%	11,72%
<b>PASSE SIMPLE</b>	normé	26,76%	45,24%	50,25%
	phonologie normée	54,93%	36,61%	29,98%
	non normé	18,31%	18,15%	19,77%
<b>PART. PASSE</b>	normé	18,68%	35,14%	43,32%

	phonologie normée	54,95%	44,20%	42,18%
	non normé	26,37%	20,65%	14,50%
<b>IMPARFAIT</b>	normé	11,38%	34,07%	50,82%
	phonologie normée	60,98%	50,07%	39,21%
	non normé	27,64%	15,86%	9,97%

Tableau 6 : répartition des échecs et des réussites pour chaque tiroir verbal étudié<sup>5</sup>

On note une amélioration substantielle du CP au CE2 pour l'ensemble des tiroirs verbaux étudiés. Pour chacun d'eux, les élèves progressent clairement. En CP, les deux temps les mieux réussis sont le présent et l'infinitif, pour lesquels on atteint des scores de formes normées de 30% ; en CE2, ce sont encore le présent et l'infinitif qui se détachent avec des scores dépassant respectivement 63% et 64%. Pour l'imparfait et le passé simple, près de la moitié des verbes présentent encore des erreurs orthographiques ou phonologiques, le participe passé étant un peu en retrait (43,3% de formes normées).

Au niveau de la progression du CP au CE2, l'imparfait se dégage nettement avec une marge de progression de +40% des formes normées, contre une progression de 30% en moyenne pour les autres temps. À l'inverse, le participe passé et le passé simple progressent de manière moins conséquente, avec une évolution respectivement de 24,6% et 23,5%.

#### Distinction des erreurs sur la base et sur la désinence

Enfin, il nous a paru intéressant de distinguer les erreurs qui portent sur les bases verbales des erreurs qui portent sur les désinences. La figure 4 permet de visualiser la proportion de formes normées (en gris), la proportion de formes présentant une erreur seulement sur la base, celle présentant une erreur et sur la base et sur la désinence et enfin celles avec une erreur seulement sur la désinence.

---

<sup>5</sup> Les formes normées correspondent aux formes correctement orthographiées.

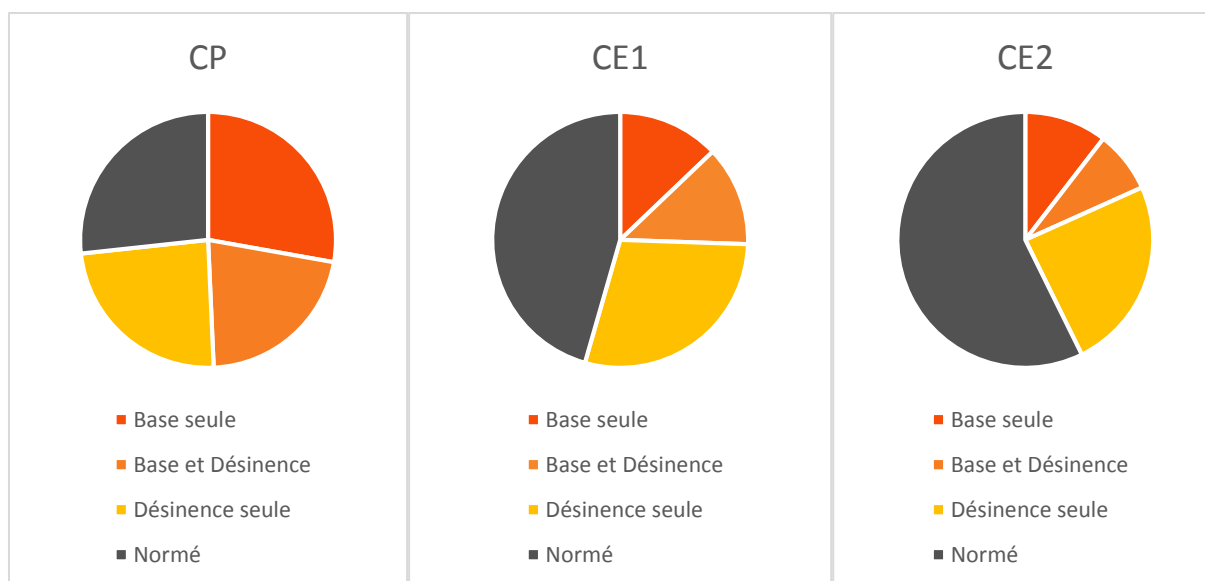


Figure 4 : Répartition des erreurs sur les bases et les désinences des formes verbales<sup>6</sup>

Il apparaît sur ces graphiques qu'en classe de CP, les pourcentages de formes produites comportant une erreur sur les bases et les pourcentages de formes produites comportant une erreur sur les désinences sont proches. En effet, on observe 49,3% d'erreurs sur les bases (donnée calculée en additionnant le pourcentage de formes comportant des erreurs sur leur base uniquement et le pourcentage de formes présentant à la fois des erreurs sur leur base et sur leur désinence), alors qu'on observe 45,5% d'erreurs sur les désinences (donnée calculée de la même manière). Ce n'est plus le cas en classe de CE2, où le pourcentage d'erreurs sur les bases a sensiblement diminué pour n'être plus que de 18,3%, tandis que le pourcentage d'erreurs sur les désinences est de 32,3%. Les erreurs sur les bases diminuent donc plus fortement que les erreurs sur les désinences.

#### Le cas de l'imparfait

Nous proposons de nous pencher plus spécifiquement sur l'imparfait de l'indicatif (figure 5), tiroir verbal quelque peu atypique dans la mesure où il est peu réussi en CP (11,4%) et réussi à environ 50% en CE2.

---

<sup>6</sup> Lecture des graphiques : le nombre d'erreurs sur les bases est donné par l'addition des erreurs sur la base seule et des erreurs à la fois sur la base et la désinence.

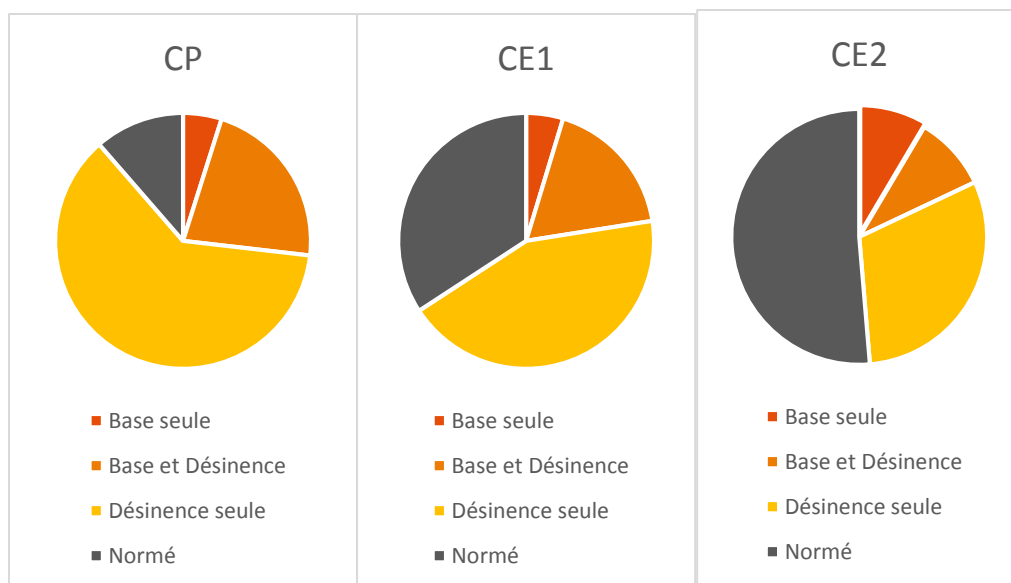


Figure 5 : Répartition des erreurs sur les bases et les désinences des formes verbales à l'imparfait

Globalement, ce tiroir verbal fait l'objet d'un grand nombre de formes présentant un problème de désinence et d'un nombre plus restreint de problèmes de base. En CP, par exemple, environ une forme sur quatre comporte une erreur sur la base (« alait » à la place de « allait » alors que l'ensemble des formes verbales en CP est affectée à près de 50% (figure 5). Cette tendance est moins nette en CE2, les verbes à l'imparfait présentant des proportions de formes avec problème sur la base proches de la moyenne des tiroirs verbaux.

L'imparfait est un tiroir verbal particulièrement notable du point de vue des erreurs de désinence. Alors que le pourcentage de mots verbaux comportant au moins une erreur de désinence au CP est de 45,5%, il est de 83,7% pour les verbes à l'imparfait (*voulé* à la place de *voulait*), ce qui représente un écart conséquent. Cet écart a tendance à s'amenuiser dans les années suivantes, puisqu'en CE2, le pourcentage de mots verbaux comportant au moins une erreur est 32,3% pour l'ensemble des tiroirs verbaux et de 40,2% pour l'imparfait. Les désinences représentent donc un enjeu particulier dans l'apprentissage de l'imparfait.

#### Le cas du passé simple

Penchons-nous à présent sur un autre tiroir verbal moyennement réussi en CE2, le passé simple (figure 6).

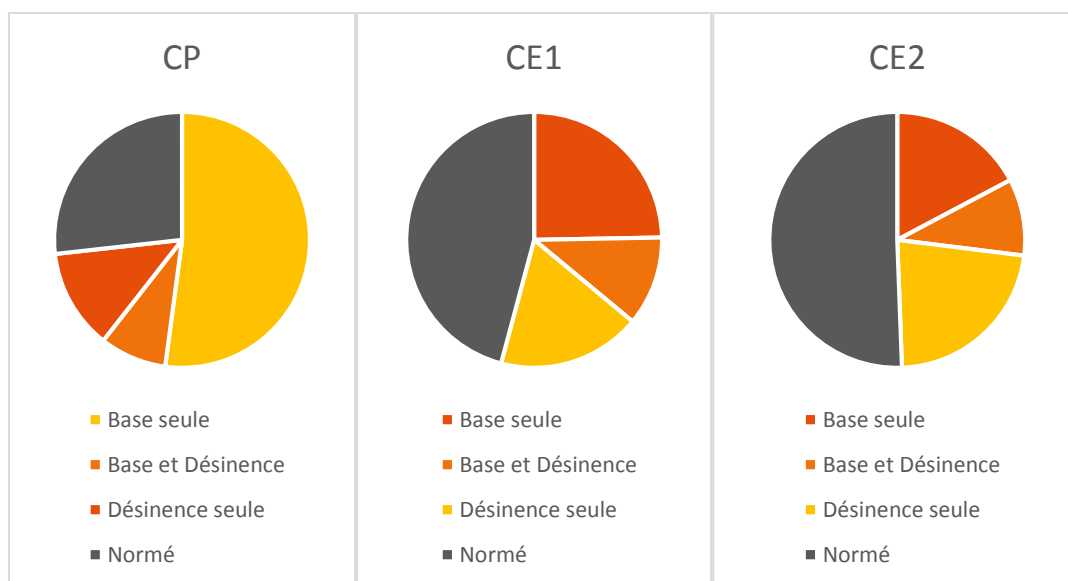


Figure 6 : Répartition des erreurs sur les bases et les désinences des formes verbales au passé simple

Alors que l'imparfait présentait un grand nombre d'erreurs de désinence, le passé simple est un temps qui concentre plus d'erreurs sur les bases, par exemple « rolacha » à la place de « relâcha » (élève 864) ou « prena » pour la forme attendue « prit » (élève 584, CP) qui comporte à la fois une erreur sur la base et sur la désinence. Au CP, 60,6% des verbes au passé simple présentent des erreurs sur leur base. Par comparaison, on en dénombre 49,3% sur l'ensemble des verbes aux temps considérés.

Il peut être intéressant de noter également que la progression des réussites est moins conséquente pour ce tiroir verbal que pour les autres tiroirs verbaux. En effet, sur l'ensemble des tiroirs verbaux considérés, on note une progression moyenne de plus de 30%, de 26,7% en CP à 57,3% en CE2, tandis que la progression pour les verbes au passé simple est inférieure à 24%, de 26,8% en CP à 50,6% en CE2.

Au vu de ces résultats, il semble que, plus que pour tout autre tiroir verbal, la formation de la base des verbes au passé simple soit particulièrement délicate pour les apprenants.

### Perspectives

Les résultats présentés ici, issus d'un premier traitement outillé par le TAL<sup>7</sup> des formes verbales dans 903 textes d'élèves du CP au CE2, nous renseignent sur les zones de réussite et les points d'achoppement lors de l'apprentissage de la morphologie verbale. Du point de vue du traitement automatique des langues, grâce auxquels les outils utilisés ici ont été produits, ils renseignent sur la façon dont ce domaine peut outiller l'exploration linguistique et didactique de l'apprentissage de l'écriture.

Ces premiers résultats permettent également de dégager plusieurs pistes didactiques.

---

<sup>7</sup> Traitement Automatique des Langues

La première concerne les tiroirs verbaux à travailler en priorité si l'on demande aux élèves de produire des récits. En CP, mais aussi en CE1 et en CE2, c'est le présent qui domine dans les usages examinés. Si 62% des verbes utilisés en CP sont au présent (et cela peut être l'effet de la consigne, l'histoire du petit chat pouvant se transformer en une espèce de description des quatre images séquentielles proposées aux élèves), ce sont 35% et 37% des verbes qui sont respectivement au présent pour le CE1 et le CE2. Un premier examen des erreurs tend à montrer que les marques de personne non nécessaires à la phonographie (le e de *pleure* vs le e de *chante*) pourraient faire l'objet d'un travail approfondi, en même temps que les règles phonographiques continuent à être travaillées au CP et au CE. Les verbes en -er comme *pleurer* ou *tomber* sont en effet très employés dans les productions de CP de notre corpus.

En CE1 et en CE2, les élèves utilisent davantage les temps du passé pour raconter leur histoire. L'imparfait est le 2<sup>e</sup> tiroir verbal le plus employé, loin devant le passé simple. Il fait dès le CP, mais aussi en CE1 et en CE2, l'objet d'un grand nombre d'erreurs de désinence, qu'il conviendrait de sérier. Dans quel(s) jeu(x) d'oppositions homophoniques-hétérographiques entre l'imparfait en CP en production de texte ? et en CE2 ? Qu'est-il raisonnable, ou rentable, de travailler en CP et au CE de ce secteur des formes verbales en /E/ ? Des investigations plus poussées devraient permettre de répondre à cette question et ainsi de mieux utiliser le temps scolaire, de façon adaptée.

Parallèlement à l'augmentation du taux de réussite du CP au CE2, le nombre d'erreurs de transcription de la chaîne sonore (formes non normées) diminue du CP au CE2, ainsi que celles qui respectent la phonologie de la forme attendue. Mais la proportion de formes non normées est à un niveau élevé au CP (figure 3). La piste didactique qui en découle est l'appui sur l'oral, pour ne pas dire la correction orale, au moment où les élèves écrivent.

En CE2, après le présent et l'imparfait, on enregistre autant d'occurrences d'infinitifs, de passés simples et participes passés. Globalement, au passé simple près, qui n'apparaît pas comme surmobilisé par les élèves, les tiroirs verbaux les plus utilisés à l'écrit sont aussi ceux mis en évidence par Claire Blanche Benveniste dans les corpus oraux (1999). Il y a sans doute là matière à réflexion.

Si dans la dernière partie nous nous sommes concentrés uniquement sur l'imparfait et le passé simple, il est possible de poursuivre cette étude sur l'ensemble des tiroirs verbaux. De plus, nous envisageons de préciser encore les erreurs présentes, en examinant les erreurs et les réussites pour les verbes les plus employés en concevant des outils de traitement permettant de prendre pour unité le graphème et ainsi percevoir plus précisément les points de tension, sources d'erreurs.

## Références

Béchet, F. (2001). LIA PHON: un système complet de phonétisation de textes. *Traitement automatique des langues*, 42(1), 47-67.

Blondel, C., Brissaud, C., Rinck, F. (2016). Description des pratiques orthographiques de scripteurs ordinaires et de scripteurs en difficulté avec l'écrit à partir de l'analyse d'une dictée, CMLF 2016, Tours. Disponible à <[http://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf\\_cmlf2016\\_07003.pdf](http://www.shs-conferences.org/articles/shsconf/pdf/2016/05/shsconf_cmlf2016_07003.pdf)>

- Blanche-Benveniste, C. (1999). La conjugaison des verbes : virtuelle, attestée, défective, *Recherches sur le français parlé*, 15, 87-112. Blanche-Benveniste C. (2002). , Structure et exploitation de la conjugaison des verbes en français contemporain, *Le français aujourd'hui* 2002/4 (n° 139), p. 11-22. DOI 10.3917/lfa.139.0011
- Brissaud, C. (2002). Etudier le verbe au collège, *Le français aujourd'hui*, 139, 59-67.
- Brunet, E. (1999). La langue française au XX<sup>e</sup> siècle, II. Ce que disent les chiffres, in J. Chaurand (dir.), 675-727.
- Catach, N. (1980). *L'orthographe française. Traité théorique et pratique*, Paris, Nathan.
- David, J. & Laborde-Milaa, I. (2002). Présentation. *Le français aujourd'hui*, 139,(4), 3-9. doi:10.3917/lfa.139.0003.
- Dubois, J. 1967). *Grammaire structurale du français : le verbe*, Paris, Larousse.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods*, 36(1), 156-166.
- Lucci, V., Millet, A., dir. (1994). *L'orthographe de tous les jours. Enquête sur les pratiques orthographiques des Français*, Paris, Champion.
- Martinet, A. (1979). *Grammaire fonctionnelle du français*, Paris, Crédif-Didier.
- Meleuc, S., Fauchart, N. (1999). *Didactique de la conjugaison. Le verbe autrement*, Bertrand-Lacoste / CRDP Midi Pyrénées.
- Pellat, J.-C. (2009). *Quelle grammaire enseigner ?* Paris, Hatier.
- Savelli, M., Brissaud, C., Chevrot, J.-P. & Gounon, V. (2002). L'apprentissage d'un temps peu enseigné : le passé simple, *Le français aujourd'hui*, 139, 39-48.
- Schmid, H. (1995). Treetagger| a language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*, 43, 28.
- Wolfarth, C., Ponton, C., Totereau, C. (2017). Apports du TAL à la constitution et à l'exploitation d'un corpus scolaire au travers du développement d'un outil d'annotation orthographique. *Corpus* 16, janv. 2017, 185-214.

