

Évolution des compétences rédactionnelles à l'école primaire : quelques pistes exploratoires à partir de méthodes quantitative et qualitative

Evolution of writing skills at primary school: some exploratory approaches based on quantitative and qualitative methods

Christophe BENZITOUN - Université de Lorraine, CNRS ATILF

Dominique LEGALLOIS - Université Sorbonne Nouvelle, CNRS LATTICE

Résumé : Dans cet article, nous proposons de recourir à des techniques issues de la linguistique de corpus et de la linguistique de l'oral pour observer l'évolution des compétences rédactionnelles d'élèves francophones. Pour ce faire, nous présentons principalement une méthode non-supervisée basée sur des ngrams ainsi que le dispositif de l'analyse en grille (Blanche-Benveniste et al., 1979). Notre étude porte sur le corpus longitudinal Scoledit (Wolfarth et al., 2017).

Mots-Clés : écrits scolaires ; compétences rédactionnelles ; corpus ; analyse en grille ; grammaire des performances

Abstract: In this article, we propose to use techniques from corpus linguistics and oral linguistics to observe the development of writing skills in French-speaking students. To this end, we mainly present an unsupervised method based on ngrams and the grid analysis device (Blanche-Benveniste et al., 1979). Our study focuses on the Scoledit longitudinal corpus (Wolfarth et al., 2017).

Keywords: academic writing; writing skills; corpus; grid analysis; performance grammar

Introduction

L'évaluation des compétences rédactionnelles des élèves représente une question difficile. En effet, comment montrer objectivement, au-delà de l'orthographe, que les aptitudes

littérariques se développent? Jusqu'à présent, on pouvait échafauder des hypothèses et les vérifier sur un petit nombre de copies. Grâce à des projets de mutualisation et de diffusion de copies d'apprenants comme *E-Calm*, la situation est en train de changer. Cela nécessite cependant, en amont, de réfléchir à des dispositifs permettant d'observer l'évolution des compétences rédactionnelles. Dans cet article, nous présentons modestement quelques pistes allant dans ce sens.

Grâce à sa taille et à sa dimension longitudinale, le corpus *Scoledit* (Wolfarth et al., 2017) permet des avancées majeures dans l'observation et l'analyse des performances rédactionnelles des élèves d'écoles primaires. En effet, ce corpus donne aux chercheurs la possibilité d'observer l'évolution des productions des mêmes élèves, productions répondant au même genre textuel et à la même consigne, du moins du CE1 au CM2.

Nous abordons la question de l'évaluation à partir de deux points de vue différents mais complémentaires : l'un global et quantitatif et l'autre plus local et qualitatif. Dans un premier temps, nous présentons le corpus *Scoledit*, puis nous étudions globalement les copies en recourant aux principes et techniques de la linguistique de corpus ; c'est la perspective qui est mise en œuvre dans les troisième et quatrième sections de cet article. Il s'agit d'identifier, de la façon la moins supervisée possible, les évolutions des performances en ciblant quelques phénomènes grammaticaux, lexicaux et textuels. Ainsi, nous présentons l'évolution d'items par niveau, mais sans volonté d'exhaustivité. En effet, définir une « grammaire des performances » par niveau est encore un travail en chantier qui demande que soit affinée et améliorée la méthode que nous présentons. Autrement dit, nous proposons les résultats d'une première analyse exploratoire, qu'il conviendra de développer par la suite.

Dans les parties suivantes, nous procédons à une étude qualitative plus locale : premièrement en analysant l'évolution de la place de l'adverbe *enfin* ; deuxièmement, à l'aide du dispositif de l'analyse en grille, il s'agit de représenter l'architecture syntaxique de chaque copie afin de faciliter les comparaisons et de mettre en lumière l'enrichissement des énoncés. Ce dernier dispositif permet à la fois de visualiser et de comprendre l'évolution de l'architecture de la phrase au cours du CE1-CM2, mais aussi du texte.

1. Corpus d'étude

Nous avons travaillé sur le corpus longitudinal *Scoledit* intégré à la ressource *E-Calm* (Hodac et al., 2020). Nos analyses ont porté sur les transcriptions des copies en orthographe

normalisée. Étant donné que nous nous intéressons aux dimensions syntaxiques et textuelles, les erreurs orthographiques auraient rendu le traitement automatique plus laborieux. En outre, des travaux ont montré que les problèmes d'orthographe ont tendance à masquer les acquis et progrès des élèves aux niveaux syntaxique et textuel (Blanche-Benveniste, 1980 ; Blanche-Benveniste et al., 1979).

Les données ont été collectées au mois de juin entre 2014 et 2018. Il s'agit de courts récits produits par des élèves des académies de Toulouse, Bordeaux, Lyon, Clermont-Ferrand et Grenoble. Les consignes n'étaient pas les mêmes pour le CP (1^{ère} année primaire), d'une part, et les autres niveaux (2^{ème}-5^{ème} année primaire), d'autre part. Pour éviter des difficultés de comparaison liées à la consigne, nous avons décidé d'écarter les productions des élèves de CP.

Il s'agissait pour les élèves de raconter une histoire à l'écrit en choisissant un ou deux personnages parmi ceux qui étaient proposés : une sorcière, un chat, un robot et un loup. Dans la version qui nous a été fournie par C. Ponton – que nous remercions – nous avons 373 copies par niveau (CP-CM1) puis 337 (CM2). Pour travailler sur les mêmes élèves, nous n'avons retenu que les 337 pour lesquels nous disposions de l'ensemble des copies. En guise d'illustration, voici les 4 copies en orthographe normalisée de l'élève 3142 :

- (1) Il était une fois un chat et un robot jouaient dans la forêt. Ils vont faire une course. Et ils vont se perdre. Fin (CE1)
- (2) Il était une fois un robot et une sorcière. Ils étaient amis. un jour il alla à la mer. La sorcière est tombée à l'eau. Le robot a voulu la sortir de l'eau. Mais il n'est pas arrivé à la sortir de l'eau Alors la sorcière lui a jeté un sort qui l'a transformé en bateau Pour la sauver. Fin (CE2)
- (3) Il était une fois une sorcière qui avait un chat qui toujours passait ses journées à dormir. Alors un jour elle demanda à carrefour marché et demanda un livre magique pour transformer le chat en robot. Elle acheta tous les livres de sorcellerie et elle trouva le livre qu'elle voulait et elle rentra à sa maison. arrivée elle prépara la potion, elle la versa dans la bouche du chat et il se transforma en robot et il fit tout le travail. Depuis ce jour-là c'est elle qui resta couchée. Fin (CM1)
- (4) Il était une fois le petit robot rouge qui devait aller rendre visite à sa grand-mère. Mais sur le chemin de la maison de sa grand-mère il croisa la sorcière qui lui dit ne va pas dans ce bois. Il y a ma soeur maléfique. Il dit je n'ai pas peur de ta sœur. Elle va te transformer en chat ou en loup. La voilà, viens te cacher derrière ce rocher. Elle part du bois. On peut aller voir ma mamie. Non ce n'était pas ma soeur trop tard il était déjà dans le bois. Mais où tu vas voir mamie ? Ouf on a réussi à passer le bois de la sorcière. Mais

où est ma copine ? Tiens mamie ce que tu m'as demandé ton petit pot de beurre et de la confiture avec du pain. J'ai réussi. Fin. (CM2)

Grâce aux copies reproduites ci-dessus, on voit tout l'intérêt de la dimension longitudinale pour observer les évolutions des compétences rédactionnelles. Par exemple, on peut voir ici, d'une part, la complexification de la zone préverbale ; d'autre part, l'introduction du discours d'abord narrativisé, puis du discours direct, assez complexe car il constitue le support de la progression narrative.

La composition de notre corpus de travail est la suivante¹ :

Tableau 1 : Composition du corpus de travail

| Classe | Types | Tokens | Phrases | Tokens/phras e |
|--------|-------|--------|---------|-------------------|
| CE1 | 2691 | 27447 | 2385 | 11,5 |
| CE2 | 4270 | 48317 | 3995 | 12,09 |
| CM1 | 5258 | 58864 | 4743 | 12,41 |
| CM2 | 5954 | 65029 | 4893 | 13,29 |

L'élève ayant le moins écrit n'a produit que 141 tokens dans ses 4 copies et celui qui a le plus écrit 1247. La moyenne est de 555,5 tokens par élève et la médiane est de 536 tokens. Les phrases sont des unités délimitées par des ponctuations fortes.

2. Une perspective globale à partir des spécificités de descripteurs par niveau

2.1. Méthode

Dans l'objectif d'identifier automatiquement l'évolution des emplois des formes linguistiques au cours des quatre années, nous avons eu recours à une méthode informatisée. Celle-ci consiste à appliquer une annotation spécifique aux textes normalisés, qui tient compte

¹ Obtenue grâce au paquet R Quanteda (<https://quanteda.io/>).

d'un « grain » ni trop général (comme une simple annotation morphosyntaxique), ni trop spécifique (comme une simple lemmatisation). Par exemple, un passage comme :

(5) Un jour une biche passa. les loups ont essayé de l'attraper mais la biche leur demanda
Pourquoi vous nous attaquez ? vous étiez si gentils avant (1113, CE2)

fait l'objet d'une annotation « mixte » qui consiste à assigner à chaque forme un « descripteur » (c'est-à-dire une étiquette) correspondant soit à un lemme (par ex. *être*, *avoir*), soit à une catégorie morphosyntaxique (ex. ADJ=adjectif, INF=infinitif, VPRES=verbe au présent, VPS=verbe au passé simple, PPAS = participe passé, etc.), soit à une catégorie morphosyntaxique à laquelle est associée la fonction sujet ou objet (par ex. NCs renvoie à « nom commun sujet », NCo à nom commun objet, etc.²), soit une catégorie sémantique (NCT vaut pour les lexèmes exprimant une temporalité : *jours*, *semaine*, *mois*, *minute*, etc.), PHASE vaut pour les adverbes chronologiquement orientés, par ex. : *avant*, *après*, *ensuite*, *d'abord*, etc.)³. Le résultat de l'annotation pour le passage ci-dessus est :

(6) un NCT un NCs VPS. le NCs avoir PPAS de le INF mais le NCs lui VPS pourquoi vous nous VPRES ? vous être si ADJ PHASE.

2.2. Premiers résultats

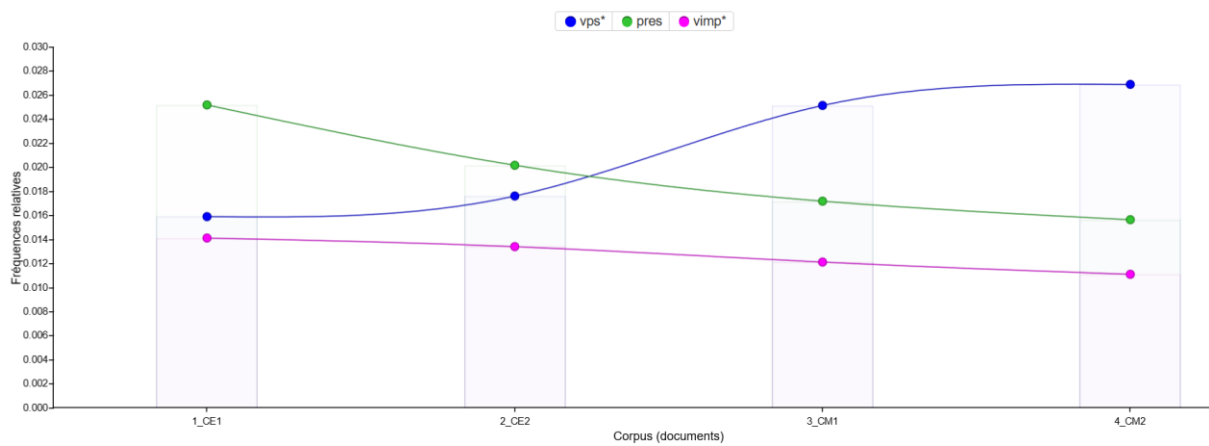
En normalisant les fréquences, on peut observer l'évolution de certains descripteurs. Le premier graphique⁴ ci-dessous atteste d'une évidence pour les enseignants, à savoir, dans les récits, le maintien de l'imparfait, et le « déclin » du présent (noté *pres* dans le graphique), remplacé par le passé simple :

² NC vaut pour les noms qui ne sont ni sujets ni objets.

³ L'annotation a été faite par le paquet R Udpipes (<https://bnosac.github.io/udpipe/docs/doc1.html>) et ensuite enrichie par un script R « maison ».

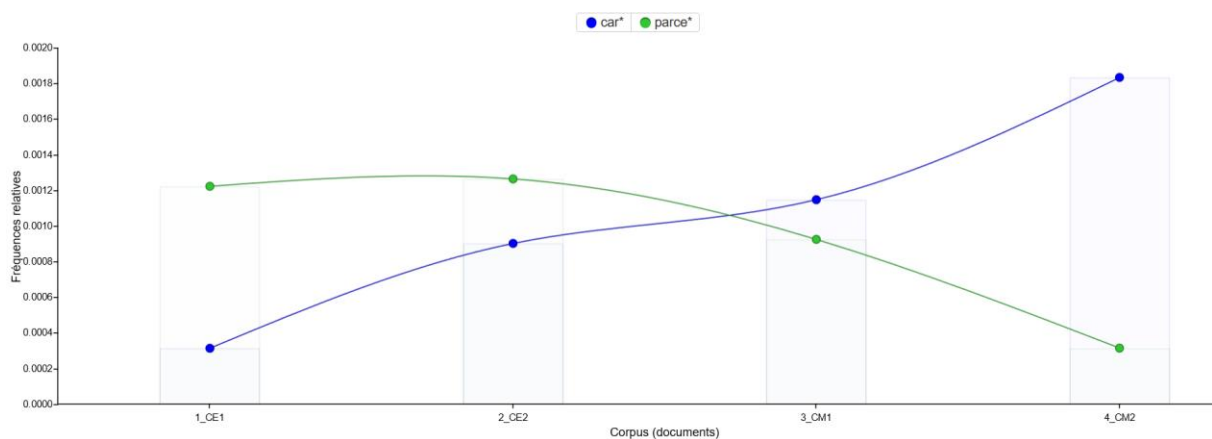
⁴ Fait à partir de l'outil en ligne Voyant Tools (<https://voyant-tools.org/>).

Figure 1 : Évolution de la fréquence des temps verbaux



Certaines évolutions sont plus surprenantes et mériteraient une étude précise ; c'est le cas des connecteurs causaux *car* et *parce que*, dont les courbes d'évolution se croisent :

Figure 2 : Évolution de la fréquence de *car* et *parce que*



Car apparaît 110 fois dans le corpus CM2 et *parce que* seulement 19 fois⁵. Ce rapport montre que la différence entre causalité subjective (c'est-à-dire déterminée par inférence du locuteur) et causalité objective ne tient pas ici (cf. Zufferey et al. 2018).

L'observation de ces évolutions est instructive ; la méthode a le défaut cependant d'être supervisée : l'analyste doit choisir les descripteurs qui lui paraissent intéressants à étudier, au risque de manquer certaines formes tout aussi significatives. Aussi, la méthode non-supervisée que l'on propose par la suite s'appuie-t-elle sur une comparaison des quatre niveaux par des méthodes statistiques. Ceci permet de dégager quelques-unes de leurs caractéristiques ; on applique un calcul des spécificités chronologiques des descripteurs (Salem, 1991) en comparant : 1-le CE2 avec le CE1 ; 2-le CM1 avec le CE1-CE2 ; 3-le CM2 avec le CE1-CE2-CM1.

⁵ La fréquence de *puisque* est négligeable dans le corpus.

La statistique fait « remonter » ainsi les formes caractéristiques d'un niveau⁶. Peuvent être considérés soit les descripteurs simples, soit des suites de descripteurs (des « ngrams » de descripteurs) qui correspondent alors à des unités sinon syntaxiques, en tout cas syntagmatiques (par ex. [se VIMP dans], [et il VPS]).

Dans les résultats qui suivent, nous allons essentiellement nous intéresser aux suites de descripteurs, dans la mesure où notre objectif principal est la caractérisation des compétences rédactionnelles – ce qui implique une observation des combinaisons de formes. Faute de place, nous ne donnons ici que quelques suites de formes par niveau, dont l'indice de spécificité est supérieur à 3.

2.3. Comparaison CE2/CE1

- [NCs] [il]

[NCs] renvoie à une séquence commençant par une virgule, suivie par un nom sujet ; [il] renvoie à une séquence commençant par une virgule, suivie par un pronom de 3^e personne sujet.

Les concordances (ou le retour aux textes) concernant ces séquences indiquent la présence en CE2 de phrases juxtaposées :

(7) Elle vit le robot, l'attrapa, elle a voulu le manger (3011, CE2)

là où en CE1, sont préférées des phrases simples ou une coordination entre phrases – ce que montrent les séquences spécifiques du CE1 :

[. et le], [. et il] [NCo. et]

(8) Le chat voit le loup. Et le chat court dans la maison. (1212, CE1)

La séquentialité des actions trouve donc en CE2 une autre organisation : la phrase graphique est de moins en moins le lieu de l'expression d'une seule action.

- [NCs lui dire]

Les élèves de CE2 introduisent de plus en plus le discours direct (invariablement avec le verbe *dire*), qui peut avoir une fonction structurante comme dans ce texte :

(9) Le loup lui dit ma vieille amie je vais vous dire quelque chose. je ne suis pas votre serviteur la sorcière lui dit mais alors qu'est-ce que vous êtes ? le loup lui répondit en s'énervant. je suis un loup. as-tu compris ? le loup lui dit je vais m'en aller puisque tu n'arrêtes pas de me dire d'aller te chercher des pommes. et moi qu'est ce que je vais faire sans mes pommes ? dit la sorcière. le loup lui dit vas te les chercher toi. (1842, CE2)

⁶ Cela ne signifie pas nécessairement qu'elles apparaissent à ce niveau, mais le plus souvent qu'elles deviennent significativement plus fréquentes.

L'emploi de plus en plus fréquent du discours direct est la raison de la spécificité des pronoms de dialogue et des séquences auxquelles ils participent, comme [je être AD]] (*je* vaut ici pour les pronoms de dialogue *je* et *tu*) :

(10) le loup dit « tu es bête et moi je suis intelligent ». (2442, CE2)

Le dialogue favorise l'émergence d'un discours évaluatif dont l'adjectif attribut constitue l'une des expressions directes.

2.4. Comparaison CM1/CE1-CE2

- [. NP être]

Ces séquences apparaissent en CM1 en suremploi, comme si elles s'étaient propagées du discours direct à la narration même. Elles permettent de caractériser un personnage (désigné par un nom propre).

(11) . François était un président totalement nul. (864, CM1)

(12) . Pomie était très maligne elle savait même assassiner quelqu'un (2887, CM1)

- [le NCmod de]

(13) la sorcière ouvre la porte de la maison du chat et entre. (3032, CM1)

NCmod désigne un nom complément du nom. Du point de vue des compétences syntaxiques, la présence de cette catégorie est à mettre en relation avec la capacité de plus en plus affirmée de construire des groupes nominaux étendus. Cette capacité va s'accroître, comme on le verra, en CM2.

- [NCT plus tard]

L'étiquette NCT (pour les noms communs à valeur temporelle) se montre tout à fait pertinente dans les récits. Les concordances montrent que la séquence est systématiquement employée comme cadratif, en position initiale. Elle témoigne d'une compétence accrue dans le découpage temporel de l'histoire :

(14) Quelques années plus tard Felix trouva une amoureuse et eut beaucoup d'enfants. (1161, CM1)

- [VPS et VPS] [et se VPS] [et il VPS]

Cet ensemble de séquences caractéristiques de la coordination des propositions indique une évolution dans la construction de la phrase complexe ; si la juxtaposition était caractéristique du CE2, en CM1 les élèves usent de la coordination des actions et de la mise en commun du sujet, notamment par l'ellipse ; une raison possible à cela est la relation causale assez fréquente entre les deux propositions (la première proposition étant la cause, ou la circonstance « forte » de la réalisation de la seconde) :

(15) Le loup énervé se releva et bondit sur le chat. (2883, CM1)

(16) Le robot se transforma en géant robot et il cassa la porte et combattit contre la sorcière (1932, CM1)

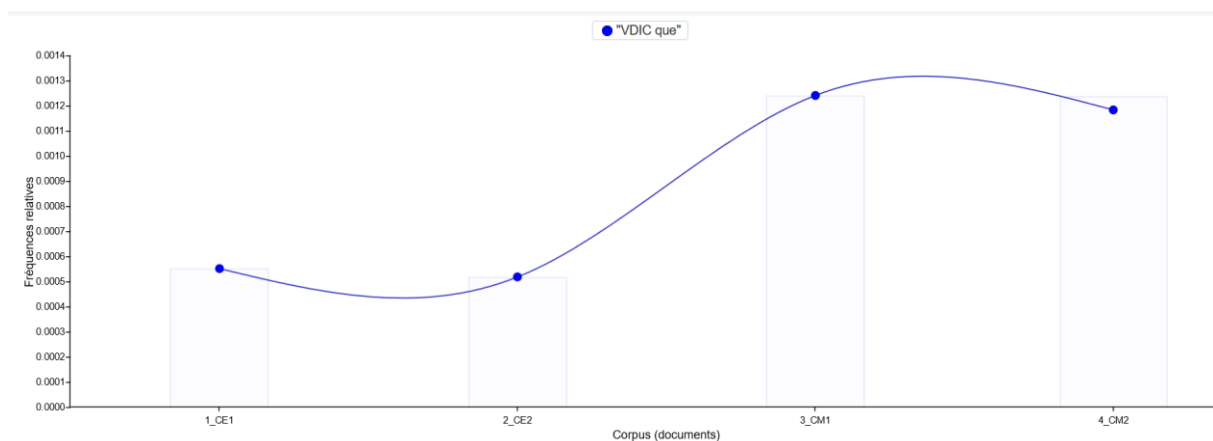
- [il VPS à]

Cette séquence suremployée en CM1 par rapport aux classes antérieures témoigne d'un usage nettement en progression des verbes de mouvements ou de déplacements orientés vers un site :

(17) il sortit dehors par une fenêtre, il courut à la barrière et regarda avec attention la forêt (502, CM1)

Enfin, signalons qu'au CM1, se développe spectaculairement le discours indirect, comme le montre ce graphique (VDIC = verbe introducteur de discours) :

Figure 3 : Évolution de la fréquence des verbes introducteurs de discours



2.5. Comparaison CM2/CE1-CE2-CM1

- [, ce être]

Cette forme n'est évidemment pas homogène ; elle peut aussi bien correspondre à des phrases clivées :

(18) , c'est pour ça qu'elle l'appelait Juliette (1931, CM2)

des phrases pseudo-clivées :

(19) Ce qu'il ne savait pas, c'était qu'un loup habitait pas très loin (1323, CM2)

qu'une phrase venant déterminer un contenu non spécifié :

(20) le chat lui préparait quelque chose, c'était un piège. (1866, CM2)

Un ensemble de structures en [ce être] apparaît donc à ce stade de façon massive, et participe à l'expansion de phrases non canoniques.

- [de le NCmod] [le NCo de] [NCT de NCmod]

Avec ces formes en spécificités positives se développe le constat fait pour le CM1 : l'expansion du groupe nominal, aussi bien en position initiale pour [NCT de NCmod] :

(21) Après plusieurs mois de travail il a réussi (1332, CM2)

(22) Un jour d'hiver glacial le loup avait faim (655, CM2)

qu'en position postverbale pour [le NCo de] :

(23) elle voulut changer les couleurs de la maison pour qu'elle soit plus gaie. (3032, CM2)

- [il VASP à]

VASP vaut pour les verbes aspectuels (*commencer, continuer, etc.*) mais ne se réalise en fait dans les textes des CM2 qu'avec *commencer*, comme dans celui-ci, assez chargé de cette forme :

(24) Elle commençait à se demander si ce n'était pas à cause du chat noir. Elle soulève sa robe trop longue et voit le chat. Elle commence à l'adopter, voir si en faisant ça elle allait redevenir plus belle ? Elle commençait à devenir une sorcière, elle voulait très vite redevenir comme avant mais elle n'y arrivait pas. (1320, CM2)

On peut y voir non pas le début d'une conceptualisation aspectuelle des procès, puisqu'il y a bien des occurrences aux niveaux inférieurs, mais une plus grande systématisation de cette conceptualisation (et plus particulièrement la phase initiale), dont il faudrait, là encore, examiner les raisons. Une des pistes serait le développement d'une compétence narrative à « diviser » aussi bien les procès en termes aspectuels (début, milieu, fin) qu'en sous procès (*elle prit son balai magique et s'envola*).

Notre méthode est encore trop peu avancée pour donner un classement exhaustif de l'évolution des formes syntagmatiques ; nous n'avons ci-dessus relevé que quelques-unes des formes pertinentes extraites par notre logiciel. Si beaucoup reste à faire, nous pensons avoir illustré les possibilités qu'offre notre démarche pour repérer des motifs pertinents. Cela permettra à terme, nous l'espérons, de construire une « grammaire » des performances de rédaction.

Pour montrer d'autres potentialités de notre démarche exploratoire, et l'intérêt d'analyser un corpus longitudinal, on propose, dans la partie suivante, une analyse manuelle de l'adverbe *enfin*.

3. Évolution de la place d'*enfin*

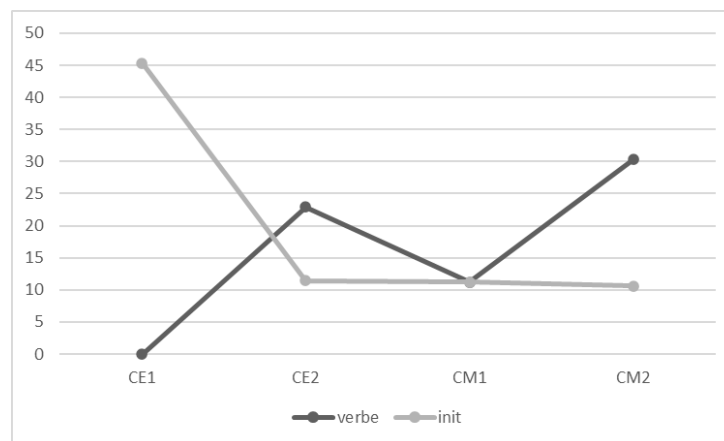
Enfin est un mot relativement fréquent dans nos données (88 occurrences). Il possède une fréquence proche de mots courants comme la conjonction *ou* (92 occurrences) et l'adverbe *encore* (89 occurrences). Il peut marquer une évolution dans la narration (25) ou bien jouer le rôle d'élément métaénonciatif (26).

(25) Un jour où le soleil resplendissait la sorcière réussit **enfin** sa potion magique (1190, CE2)

(26) **enfin** bon c'était les meilleurs amis du monde (828, CM2)

Nous nous sommes tout d'abord concentrés sur la position qu'occupe *enfin* au sein de l'énoncé. Pour ce faire, nous avons distingué deux zones : la zone initiale (init) et la zone entre l'auxiliaire et le participe passé ou directement après le verbe (verbe). Dans cette seconde catégorie, nous avons également inclus la zone entre le verbe modal et le verbe à l'infinitif (*aller enfin dormir, pouvoir enfin parler*). Les résultats sont donnés en nombre d'occurrences pour 100.000 tokens dans le graphique ci-dessous.

Figure 4 : Évolution de la fréquence de la place d'*enfin*



Comme on peut le voir, les élèves de CE1 plébiscitent la zone initiale. La fréquence relative d'*enfin* dans cette position est divisée par 4 entre le CE1 et le CE2. On observe parallèlement une progression marquée de la zone entre auxiliaire/modal et verbe ou juste après le verbe. À noter que la zone postverbale est plus utilisée en CE2 que la zone entre l'auxiliaire/modal et le verbe qui ne compte que 2 exemples dont l'un avec un verbe modal. Il faudrait vérifier si ce fonctionnement est propre à *enfin* ou si, de manière générale, la zone entre auxiliaire et participe passé est peu disponible en CE1-CE2.

Au niveau des exemples remarquables, on peut en citer deux qui ne se retrouvent qu'au CM2. Il s'agit d'un emploi sous forme d'incise et un autre accompagnant un participe passé détaché en tête.

(27) Il était une fois une maison hantée, (**enfin** un palais pour ceux qui habitaient dedans) (2033, CM2)

(28) **Enfin** transformée elle sortit de chez elle (1329, CM2)

Enfin apparaît très rarement comme marqueur de fin de coordination de syntagme (seulement 2 occurrences).

(29) le renard attaquait du côté gauche, le lapin du côté droit, la poule en avant et **enfin** le lion et le tigre derrière (690, CE2)

Il existe très peu d'exemples en fin d'énoncés (seulement 4) et parmi ceux-ci il y en a deux qui ont une valeur métaénonciative :

(30) « calmez-vous **enfin** » (3005, CM1)

(31) Non je ne t'ai pas pris de croquettes **enfin** (2981, CM2)

À partir de cette brève illustration, on voit l'intérêt qu'il peut y avoir à ne pas se limiter à la seule analyse automatique. L'annotation manuelle d'*enfin* dans les rédactions des élèves nous renseigne sur l'utilisation des différentes zones à l'intérieur des énoncés et donc sur une partie du développement grammatical des apprentis scripteurs. Mais il nous manque encore une vision d'ensemble de l'architecture du texte.

Pour y remédier, dans la partie qui suit, nous recourons au mode de visualisation que permettent les grilles syntaxiques.

4. Les grilles syntaxiques

L'analyse en grille est un mode de représentation graphique proposé par Blanche-Benveniste et al. (1979). Ce dispositif a été essentiellement utilisé pour faciliter l'analyse linguistique des transcriptions de français parlé et plus particulièrement leur lecture. En effet, il est difficile, quand on n'en a pas l'habitude, de lire des transcriptions sans mise en page, l'oral n'étant pas destiné à une édition écrite.

Rapidement, l'analyse en grille a été aussi utilisée pour aider à la lecture des copies d'élèves et faire prendre conscience des différences entre les productions écrites et orales de ces mêmes élèves. On peut trouver les prémices d'une telle démarche dans Achery et al. (1979).

L'application aux copies d'élèves a été plus récemment remise au goût du jour par Auriac-Slusarczyk et Blasco-Dulbecco (2010) ainsi que Roubaud (2014).

Le principe est simple : il s'agit de mettre en lumière l'élaboration syntaxique des textes. Pour ce faire, on utilise l'axe syntagmatique (horizontal) et l'axe paradigmatique (vertical). Ainsi, lorsqu'un élève concatène les syntagmes les uns à la suite des autres, on va les faire se succéder, comme dans l'exemple suivant⁷ :

(32) Mais / en fait, / un voleur journaliste / a échangé / la potion magique. (1955, CE2)

Lorsque l'élève instancie à plusieurs reprises une même position syntaxique (sujet, complément d'objet, circonstant, épithète, etc.), on les aligne à la verticale :

(33) il faut continuer de boire du chocolat chaud tout moche
 et laid
 et tout doux
 et moche (3071, CE1)

En définitive, il faut aligner toutes les unités insérées dans les mêmes positions syntaxiques à l'échelle du texte, comme dans l'exemple ci-dessous :

Figure 5 : Analyse en grille de la copie de l'élève 1627 (CM1)

| | | | | |
|----------------------------|-------------|-------------------|--|---------------------------------|
| | | Il était une fois | une sorcière qui s'appelait Fripouille. | |
| | | Elle était | malheureuse et toute seule | |
| Un jour | | elle passait | ses journées | à faire des potions maléfiques. |
| | | elle eut | une idée, | |
| | | elle grimpa | | sur son balai magique |
| | | et s'envola. | | |
| Arrivée au cœur du village | | elle aperçut | l'ANIMALERIE. | |
| | | Elle entra | | dedans, |
| | | elle vit | un chat, un chat noir, un chat mignon tout en étant maléfique. | |
| | Fripouille | fut | ravie | |
| | | elle alla voir | le vendeur, | |
| | | lui demanda | combien il coûtait, comment il s'appelait et patati patata | |
| | Le vendeur | lui dit | qu'il coutait 70 euros et qu'il s'appelait Maléfique. que c'était une fille. | |
| Puis | | il ajouta | | |
| | la sorcière | l'acheta | | |
| puis | | elle rentra. | | |
| | Le chat | fut | très heureux de vivre avec elle | |
| et | | ils vecurent | heureux | jusqu'à la fin de leurs jours. |
| | | FIN | | |

En un coup d'œil, on embrasse l'architecture générale du texte. On voit notamment qu'il y a seulement trois occurrences de connecteurs (*puis* 2 fois, *et* 1 fois), que le champ gauche de

⁷ Les '/' représentent les frontières syntagmatiques.

l'énoncé (*un jour, arrivée au cœur du village*) est moins développé que le champ droit et qu'il y a 4 sujets lexicaux pour 12 sujets pronominaux.

Cela permet aussi de mettre en évidence les contextes où il peut exister des problèmes de rattachements prépositionnels. Dans l'exemple ci-dessous, on peut voir que la préposition utilisée correspond à la première préposition du syntagme précédent alors que c'est la seconde qui devrait être employée :

(34) Le robot était toujours **au** service des humains
 mais jamais **aux** animaux (élève 3045, CE2)

Ce dispositif permet entre autres d'observer que les élèves n'écrivent pas comme ils parlent, même pour celles et ceux qui ont des difficultés avec l'écrit. C'est la conclusion de l'article d'Achery et al. (1979) et c'est aussi ce qui a été observé par Benzitoun (2020) à propos de la forme des sujets grammaticaux. L'analyse en grille permet également d'évaluer le développement des compétences syntaxiques (généralement masquées par l'orthographe) en mettant en évidence les tournures grammaticales qui apparaissent et la complexification progressive. C'est ce que nous allons voir dans la partie suivante.

5. Développement des compétences syntaxiques

Les résultats que nous présentons dans cette section ne sont pas représentatifs de l'ensemble des productions des élèves de notre corpus de travail. Pour l'instant, nous n'avons pu mettre en grille les copies que de quelques élèves. Mais nous avons tout de même observé un grand nombre de copies. Nous commençons par aborder le développement syntaxique pour un élève en particulier, avant de faire une synthèse de nos observations.

Dans ce qui suit, nous reproduisons uniquement quelques passages des récits rédigés par l'élève 1153.

Figure 6 : Extraits de l'analyse en grille de l'élève 1153

| | | | |
|-----|----------------------------------|---|--|
| CE1 | Un jour | elle se fait attraper | par une méchante sorcière qui l'amena dans son repère secret. |
| CE2 | Et, dans cette forêt Un jour, | habitait le grand méchant loup. il vit | la très vieille sorcière et son chat qui se promenaient dans la forêt. |
| CM1 | En réalité, | la sorcière n'était pas elle voulait juste avoir | méchante, quelqu'un à qui parler vu qu'elle était seule. |
| CM2 | Juliasse | Que puis-je faire n'a pas | pour vous ? le temps de répondre que C92 lui dit : - [...] |

En CE1, on voit que cet élève utilise le cadratif *Un jour* puis une tournure au passif en *se fait* et le complément d'agent contient une relative en *qui*. En CE2, apparaissent un sujet

postposé au verbe et une relative en facteur commun branchée sur deux constituants coordonnés. En CM1, on observe un énoncé complexe se rapprochant d'une corrélatrice composée d'une première construction verbale comportant une négation suivie d'une seconde construction verbale incluant l'adverbe *juste*, le tout se terminant par une subordonnée en *vu que*. Dans l'ensemble du corpus, la conjonction *vu que* ne se retrouve qu'à une seule reprise en CE1, 3 en CE2, 4 en CM1 et 7 en CM2. Pour finir, l'élève 1153 utilise en CM2 une interrogative avec inversion et une corrélatrice. Sur 432 interrogatives partielles du corpus, 119 comportent une inversion du sujet pronominal, dont 10 en CE1, 26 en CE2, 48 en CM1 et 35 en CM2.

À propos de l'inversion du sujet, en dehors des interrogatives et des incises, deux verbes sortent du lot. Il s'agit de *vivre* et *habiter* :

(36) Juste à côté de leur maison **habitait** un loup. (3057, CE2)

(37) au village voisin **vivait** une sorcière (3053, CM2)

Une des rares exceptions que nous avons trouvées est la suivante :

(38) Il se voit dans la maison du sorcier avec le téléporteur où dessus **est marqué** en clignotant 1948 (1997, CM2)

D'un point de vue global, voici des repères provisoires que nous pouvons donner à partir de notre fréquentation du corpus de travail. Nous listons ci-dessous quelques étapes mettant en avant la complexification de la syntaxe des textes des élèves.

Dans un premier temps, on observe des énoncés comportant un éventuel connecteur suivi d'un sujet pronominal ou lexical, un verbe et un complément essentiel. À ce stade, la zone initiale peut comporter un circonstant temporel et un circonstant locatif peut apparaître à la droite du verbe :

(39) **à chaque fois qu'elle le tapait** le chat s'enfuyait **au fond de la forêt** (3060, CE1)

Dans un second temps, on voit apparaître des énoncés ayant une mise en facteur commun du sujet :

(40) Et il bondit sur le chat et le mangea. (1153, CE2)

On peut trouver aussi un discours rapporté ou une interrogative indirecte, une incise (*dit-il*), une complétive ou une subordonnée circonstancielle le plus souvent postposée.

Dans un troisième temps, on assiste à un remplissage plus important de la zone initiale (subordonnée circonstancielle de temps, gérondif et participe passé détaché), une montée en puissance des discours rapportés et de la mise en facteur commun du sujet, ainsi que des syntagmes en *pour* + infinitif.

Dans un quatrième temps, on observe des tournures comparatives, une diversification des adverbes connecteurs, des insertions entre le sujet et le verbe et l'apparition de gérondifs à la droite du verbe principal.

Ce travail demanderait bien évidemment une plus grande systématisation pour être représentatif de l'ensemble des copies. Cependant, nous pensons avoir montré les potentialités prometteuses de la présentation sous forme de grille pour évaluer l'évolution des compétences rédactionnelles des élèves.

Conclusion

Notre démarche exploratoire a permis d'illustrer les intérêts potentiels que l'on peut tirer de techniques empruntées à la linguistique de corpus (avec le calcul de spécificités chronologiques de formes syntagmatiques), et à la linguistique de l'oral (avec l'analyse en grille). Nos premiers résultats donnent un aperçu de ce que pourrait être le répertoire par niveau des formes linguistiques relevant de la compétence rédactionnelle. Le développement de la double méthode pourra permettre aux enseignants d'identifier chez les élèves la présence ou l'absence de formes caractéristiques de leur niveau. Il s'agira donc de déterminer un niveau « moyen » de compétences, et de situer les productions par rapport à cette référence. Avant d'arriver à ce résultat, il conviendra, d'une part, d'améliorer encore l'annotation (par exemple avec de la sémantique) pour une meilleure caractérisation des formes évolutives ; et d'autre part, de systématiser et de formaliser l'analyse en grille, afin de constituer un dispositif extrêmement précieux et efficace pour observer les compétences à un niveau syntaxique et textuel. On pourrait ainsi concevoir des modèles de grilles correspondant à des types de production et comparer le texte d'un élève donné à cette grille.

Bibliographie

- Achery, J., et al. (1979) "Chez nous ses pas male" ou "Que cache l'orthographe ?". *Recherches sur le français parlé* 2, 223-252.
- Auriac-Slusarczyk, J. Blasco-Dulbecco, M. (2010) Interpréter des copies : l'intérêt des mises en grille syntaxique. *Synergies Pays scandinaves* 5, 31-48.
- Benzitoun, C. (2020) Pour un français parlé de référence à partir de corpus. In R. Gagnon et V. Bourhis (coord.), *Les corpus oraux et leur didactisation, éla 198*, Klincksieck, 137-148.

- Blanche-Benveniste, C. (1980) La pelle à godets ou les ravages de l'orthographe dans le technique court. *Le français aujourd'hui* 48, 33-39.
- Blanche-Benveniste, C. et al. (1979) Des grilles pour le français parlé. *Recherches sur le français parlé* 2, 163-204.
- Ho-Dac, M. Fleury, S. Ponton, C. (2020) É:Calme Resource: a Resource for Studying Texts Produced by French Pupils and Students. Proceedings of the *12th Conference on Language Resources and Evaluation (LREC 2020)*, 4327–4332.
- Roubaud, M-N. (2014) *De la description de la langue à son enseignement*. Habilitation à diriger les recherches, Université Stendhal Grenoble 3.
- Salem, A. (1991) Les séries textuelles chronologiques. *Histoire & Mesure* VI-1/2, 149-175.
- Wolfarth, C. Ponton, C. Totereau, C. (2017) Apports du TAL à la constitution et à l'exploitation d'un corpus scolaire. In C. Doquet, J. David et S. Fleury, *Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement*, *Corpus* 16, 185-214.
- Zufferey, S. et al. (2018) Usage and processing of the French causal connectives 'car' and 'parce que'. *Journal of French Language Studies*, Volume 28, Issue 1, 85-112.