



HAL
open science

Le français écrit : transcription et édition. Le cas des textes scolaires

Marie-Noëlle Roubaud

► **To cite this version:**

Marie-Noëlle Roubaud. Le français écrit : transcription et édition. Le cas des textes scolaires. Corpus, 2017, 16. hal-01727465

HAL Id: hal-01727465

<https://hal.science/hal-01727465>

Submitted on 9 Mar 2018

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.

Le français écrit : transcription et édition. Le cas des textes scolaires

Marie-Noëlle Roubaud



Édition électronique

URL : <http://journals.openedition.org/corpus/2770>

ISSN : 1765-3126

Éditeur

Bases ; corpus et langage - UMR 6039

Édition imprimée

Date de publication : 1 janvier 2017

ISBN : 16638-9808

ISSN : 1638-9808

Référence électronique

Marie-Noëlle Roubaud, « Le français écrit : transcription et édition.

Le cas des textes scolaires », *Corpus* [En ligne], 16 | 2017, mis en ligne le 06 janvier 2018, consulté le 07 janvier 2018. URL : <http://journals.openedition.org/corpus/2770>

Ce document a été généré automatiquement le 7 janvier 2018.

© Tous droits réservés

Le français écrit : transcription et édition.

Le cas des textes scolaires

Marie-Noëlle Roubaud

- 1 Si nous plagions le titre de Blanche-Benveniste & Jeanjean (1987) *Le français parlé. Transcription et édition*, c'est parce que notre intérêt pour les écrits scolaires nous a renvoyée à la situation du *Groupe aixois de recherches en syntaxe* (GARS), qui fut le premier en France à s'intéresser au français parlé et à publier une revue scientifique : *Recherches sur le français parlé*. Nous avons participé à cette entreprise¹ de recueil de données orales, de transcription et d'édition et c'est là que nous avons pris conscience des difficultés présentes tout au long des étapes de la constitution d'un corpus et de l'importance des choix méthodologiques effectués à ce moment-là.
- 2 Dès 1997, nous avons commencé à constituer un corpus de textes scolaires². L'idée de recueillir de telles données était novatrice : le français écrit des élèves n'intéressait pas (ou peu) les linguistes (Elalouf & Boré, 2007). Nous nous retrouvions dans la même situation que les auteurs du français parlé dont le but était de donner une légitimité aux données orales avec la constitution d'un grand corpus de français parlé. L'analyse linguistique que nous avons faite des textes scolaires, dans la même perspective que Béguelin (1989) ou Fabre-Cols (2000), montre la façon dont les élèves utilisent le système de la langue et donne ainsi des repères sur l'acquisition (Cappeau & Roubaud, 2005 ; Roubaud, 2014). Le fait que les productions d'élèves soient des lieux où se matérialisent des interrogations sur la langue donne sans conteste une légitimité aux écrits scolaires.
- 3 L'article voudrait rendre compte de notre questionnement au moment de la constitution d'un corpus de textes scolaires, des choix de transcription qui ont été faits en vue de leur édition et des questions soulevées par la spécificité de tels écrits.

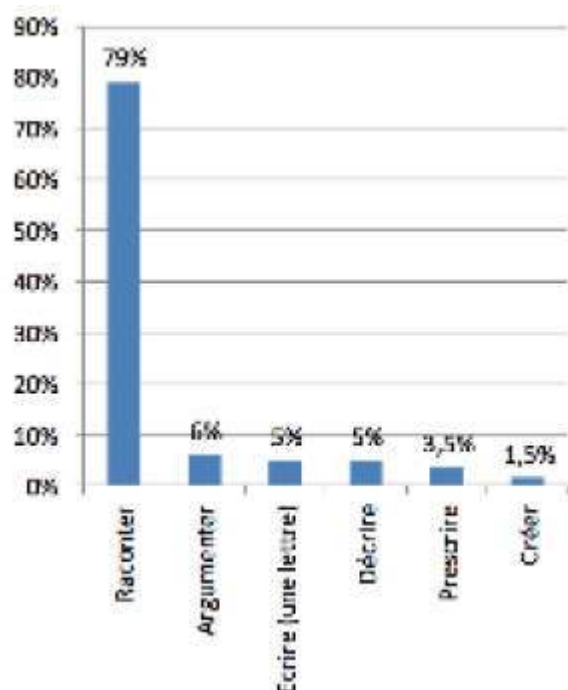
1. Du français parlé au français écrit des élèves

- 4 Jusqu'en 2015³, peu de linguistes se sont intéressés aux écrits scolaires ou à les considérer comme « des réservoirs à erreurs » (Doquet, 2013, p. 120). En 2000, David déclare que : « peu d'ouvrages complets ont été consacrés à l'étude des textes d'enfants » (p. 285). En conséquence, peu de descriptions étroitement linguistiques (lexicales et syntaxiques) leur sont associées si bien que Masseron déplore en 2011 que : « Les travaux précurseurs de *Recherches sur le français parlé* [...] ne se sont étendus aux productions écrites que grâce à la médiation d'acteurs directement informés » (p. 157). Or si l'on veut étudier les faits de langue, il faut s'appuyer sur des données attestées (Cappeau, 2007). Mais le recueil des données linguistiques s'avère complexe et difficile (Culioli, 2002). La méthodologie préconisée pour les corpus oraux (Blanche-Benveniste & Jeanjean, 1987 ; Biber *et al.*, 1999 ; Bilger, 2000 ; Baude, 2006) a servi à notre réflexion pour l'élaboration de notre corpus de textes scolaires.

1.1 La constitution d'un corpus de textes scolaires

- 5 Comme l'avait fait le GARS au moment de choisir les situations de parole, nous avons dû définir la notion d'« écrits scolaires ». Nous l'avons limitée aux productions des élèves à l'école : nous n'avons pas recueilli de textes écrits avec l'enseignant (encore appelés : « dictées à l'adulte »). Notre base comprend des textes bruts, écrits individuellement, non accompagnés de commentaires métagraphiques (Jaffré, 2003) ou de questionnaires⁴. Notre choix a été, comme pour Elalouf & Boré (2007, p. 63) de récolter des textes obtenus non dans des conditions expérimentales, mais attestés par des pratiques ordinaires. Notre démarche a été inductive et exploratoire : elle visait à révéler des textes, à en explorer les régularités, à rassembler les traits linguistiques qui ont tendance à apparaître ensemble.
- 6 Pour constituer notre corpus, nous avons fait varier deux curseurs : la taille et la variation (Habert, 2000). Concernant la taille, l'expérience du français parlé nous avait appris qu'il fallait constituer un corpus ouvert : « Au fur et à mesure que l'étude avance, on voit dans quelles directions il faudrait compléter le corpus. C'est pourquoi il paraît sage, pendant toute une phase du travail, de garder un corpus "ouvert", que l'on peut enrichir quand le besoin s'en fait sentir. » (Blanche-Benveniste & Jeanjean, 1987, p. 89). Nous avons récolté des textes d'enfants de 5 ans à 11 ans, dans des classes et des milieux différents, dont certains sont produits avec la même consigne (par les élèves d'une classe ou de plusieurs classes, de même niveau ou de niveaux différents, de la même école ou de différentes écoles), et dont d'autres correspondent à des productions d'un même élève sur plusieurs moments de l'année ou à différents moments de sa scolarité. Concernant la variation, nous avons pris en compte la notion de genres (Baude, 2006 ; Biber *et al.*, 1999 ; Lee, 2001) et avons tenté de recueillir des textes produits dans des situations très diversifiées, de façon à avoir un échantillonnage représentatif.
- 7 Mais comme pour le français parlé, nous nous sommes heurtées à la question des genres. La constitution du *Corpus de référence du français parlé* (CRFP) avait déjà montré qu'il n'existe pas de recueil idéal : malgré les décisions méthodologiques prises en amont, le CRFP contient 64,5 % de parole privée (récits de vie, de voyage, d'une expérience ou savoir-faire) contre 17 % de parole professionnelle (sur son lieu de travail) et 18,5 % de parole publique (réunions associatives, conférences). Cette distribution inégale des

situations de production s'est retrouvée dans notre corpus de textes scolaires. Sur un total de 1330 productions d'école primaire (élèves de 5 à 11 ans), nous avons 79 % de textes narratifs, comme l'indique le graphique suivant :



Doc. 1. Distribution des situations d'écriture dans un corpus de 1330 textes scolaires

- 8 Au vu de ce graphique, il est indéniable qu'« il existe une répartition des genres scolaires dans l'école et le collège » (Elalouf & Boré, 2007, p. 63). Il est intéressant de constater que même avec une longue expérience de constitution de corpus, les problèmes relatifs à la diversité des données demeurent, si bien qu'actuellement, nous passons du recueil libre au recueil sélectif ou orienté pour pallier les manques. Nous sommes persuadée que la prise en compte plus fine des situations d'écriture et une meilleure approche en genres (Blanche-Benveniste, 2010) sont incontournables si l'on veut dresser un tableau des usages.

1.2 L'organisation de la base de données

- 9 Une fois les écrits scolaires recueillis, nous les avons scannés mais très vite, il est apparu qu'il fallait organiser la base de données qui s'enrichissait sans cesse de nouveaux textes⁵. Comme l'écrit Rastier (2005, p. 31) : « Il [un corpus] est structuré d'une part en fonction d'une typologie des textes, qui se reflète dans leur codage, et d'autre part, dans chaque utilisation, par des sélections raisonnées de sous-corpus ». Notre choix a été de rassembler les textes par niveau scolaire (donc par classe et de ce fait par tranche d'âge) et à l'intérieur, par type de situation d'écriture (narratif, argumentatif, prescriptif, etc.⁶).
- 10 Pour permettre une exploitation du corpus à des fins diverses, il fallait que chaque unité soit « autonomisable » (Habert, 2000, p. 42). Nous avons ainsi mené une réflexion sur l'indexation des textes et avons élaboré un recueil de métadonnées. Nous nous sommes inspirée de la fiche signalétique qui accompagne chaque corpus oral (Blanche-Benveniste,

1996, p. 28) et nous l'avons adaptée aux écrits scolaires. En même temps que nous récoltions des textes, nous rassemblions les informations suivantes :

- 11 - Nom et code du corpus
- 12 - Identification du ou des transcripteur(s)
- 13 - Renseignements sur l'élève (Prénom, sexe, date de naissance, âge au moment de l'écriture du texte, profil de l'élève⁷)
- 14 - Classe (niveau scolaire)
- 15 - Date d'écriture du texte
- 16 - Type de situation d'écriture
- 17 - Contexte scolaire ⁸

Doc. 2. Fiche signalétique du texte
(première partie du recueil des métadonnées)

18 Il nous a fallu également ajouter les conditions de production car connaître la consigne d'écriture donnée à l'élève⁹ ne suffit pas à rendre compte de la situation de production. Nous nous sommes alors procuré les documents (quand il y en avait) relatifs à la démarche de l'enseignant : le travail en amont, les aides apportées aux élèves lors de l'écriture et les ajustements ultérieurs. Nous avons donc joint à la fiche signalétique les éléments suivants :

- 19 - La consigne d'écriture
- 20 - Les documents fournis aux élèves en amont de l'écriture (texte d'auteur donné en modèle, conseils donnés par l'enseignant...)
- 21 - Les aides au moment de l'écriture (grille d'écriture, phrase écrite au tableau, cahier de vocabulaire extrait de la méthode de lecture...)
- 22 - Les ajustements ultérieurs (grille de réécriture, conseils donnés par l'enseignant...)

Doc. 3. Annexes au texte
(deuxième partie du recueil des métadonnées)

23 Comme nous le constatons, la base de données (textes et métadonnées) est bien plus volumineuse que pour un corpus oral car de nombreux documents sont à prendre en compte, en plus de la fiche signalétique¹⁰.

1.3 La transcription des textes d'élèves

24 Les textes scannés des élèves ont constitué notre base à ses débuts mais ces écrits non informatisés empêchaient toutes sortes de requêtes : l'investigation sur les textes reste limitée à cause des images car même si la mémoire pallie les lacunes de la technique, elle ne peut pas le faire sur une grande masse de textes. Dès lors la transcription de ces textes est devenue centrale¹¹ et comme pour l'oral, nous nous retrouvons tiraillés entre deux exigences : la fidélité au texte brut et la lisibilité de son rendu par écrit.

25 Notre expérience de transcripteur du français parlé nous avait maintes fois démontré que la transcription engage une théorie (Cappeau, 2008 ; Caddeo, 2013), que les choix de transcription influent non seulement sur la lecture des textes mais aussi sur notre représentation des locuteurs (Mondada, 2002 ; Gasquet-Cyrus, 2013), ainsi que sur

l'analyse qui en est faite (Masseron, 2011). Par conséquent il nous a fallu élaborer un guide du transcripteur comme pour le français parlé (équipe DELIC, 2004), si nous ne voulions pas que les transcripteurs réécrivent les textes en modifiant ou adaptant des exemples parce qu'ils les jugent délicats à transcrire ou parce qu'ils ne voient pas la cohérence propre au texte de l'élève ou encore parce que tous n'ont pas tous le même rapport à l'écrit et à la norme (Baude & Dugua, 2011). Même si nous savons que le texte informatisé idéal n'existe pas (comme dans le cas du français parlé), il nous fallait adopter la même rigueur scientifique et le même souci de réflexivité que lors de la constitution d'une base de données orales, prendre les mêmes précautions méthodologiques (Gadet, 2000 ; Williams, 2005) et ne pas négliger cette étape d'édition.

- 26 Mais dans le cas des écrits scolaires, la situation n'est pas tout à fait identique à celle du français parlé : les textes sont déjà écrits et comportent de nombreuses « fautes » qui ne sont pas toutes à traiter de la même façon. La difficulté réside en premier lieu dans la lecture des morphèmes qui ne sont pas toujours identifiables : ce ne sont pas des écrits déjà contrôlés comme les textes littéraires. Pour passer de la version brute (le scan du texte) à la version informatisée (la version transcrite), une lecture attentive du texte brut est nécessaire tout en prenant en compte les métadonnées (pour le contexte d'écriture par exemple), ainsi qu'un savoir linguistique de la part du transcripteur. Face à de tels écrits, non normés dans la majorité des cas, un de nos principes majeurs a été de passer par la lecture à voix haute (Fabre-Cols, 2000, p. 40) afin de redonner du sens et de la cohérence au texte de l'élève.

2. Transcrire et éditer des textes scolaires

- 27 Certains choix de transcription de l'oral ont pu être transférés sur les écrits scolaires ou ne pas l'être (cas de liaison, cf. infra) mais d'autres problèmes liés aux usages linguistiques des élèves ont surgi au moment de transcrire. Dans ce dernier cas, de nouveaux choix méthodologiques (qui engagent une théorie) sont à faire. Pour illustrer ces deux points, nous nous appuyons sur des écrits produits par de jeunes enfants : de CP (cours préparatoire, 6-7 ans) et de CE1 (cours élémentaire 1^{re} année, 7-8 ans).

2.1 Du possible transfert des conventions de transcription de l'oral

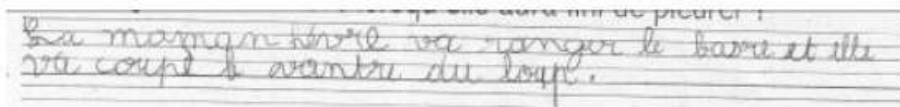
- 28 Les conventions de transcription utilisées pour l'oral (Équipe Délic, 2004, p. 27), concernant les notations graphiques, les amorces et les syllabes incompréhensibles, peuvent être transférées sur les écrits scolaires.

		Oral	Écrits scolaires
notations graphiques	[...]	prononciations particulières	graphies particulières
amorces	...-	amorces phoniques	amorces graphiques

syllabes incompréhensibles	*	syllabe incompréhensible	syllabe indéchiffrable
	***	suite de syllabes incompréhensibles	suite de syllabes indéchiffrables

Doc. 4. Transfert de conventions de transcription de l'oral sur les écrits scolaires

- 29 Nous avons fait le choix de placer la notation derrière le mot orthographié, par souci de lisibilité des textes, comme c'est le cas avec les quatre premiers écrits. Le premier texte a été produit en réponse à la consigne de la méthode de lecture CP¹² : « La chèvre vient d'apprendre que le loup a mangé six de ses cabris. Que va-t-elle faire lorsqu'elle aura fini de pleurer ? ».

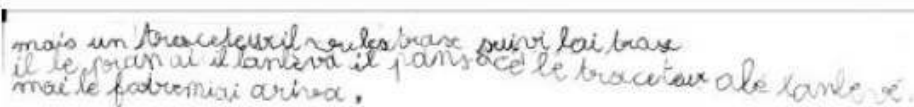


Doc. 5. Récit d'Olivia (CP, mai)

- 30 La maman chèvre [hèvre] va ranger le bazar [basre] et elle
31 va couper [coupè] le ventre [vantre] du loup.

Doc. 5 bis. Transcription du texte

- 32 Pour être en mesure d'identifier le mot « bazar », il faut connaître le contexte d'écriture (ici la production de la suite du conte « Le loup et les sept cabris »), ce qui confirme l'importance des métadonnées.
- 33 Le second texte correspond à la suite de l'histoire de *Porculus*¹³ racontée par l'enseignante. Cette dernière arrête la lecture de l'album sur cette phrase : « Il essaie bien de se soulever mais il ne peut pas remuer », indiquant que Porculus, le petit cochon, est prisonnier du ciment dans lequel il a marché et qu'il ne peut plus revenir à la ferme. L'enseignante questionne les élèves : « Que va-t-il se passer ? » et leur demande d'écrire la fin de l'histoire :



Doc. 6. Récit de Léa (CP, juin)

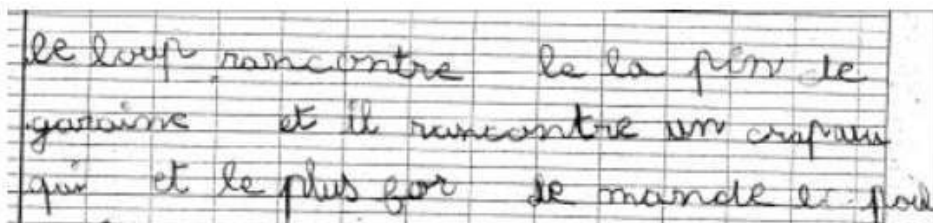
- 34 il le prend [pran] et [ai] il l'enleva [lanlèva] il pensa que [pansace] le tracteur [traceteur] allait [alé] l'enlever [lanlevé].

Doc. 6 bis. Transcription d'un extrait du texte

- 35 La lecture de [il pensace] demande du temps car on voit d'abord un « é » en finale (qui n'est en fin de compte qu'un résidu de lettre mal effacée). Il faut revenir au sens du texte pour lire « il pensa que » et posséder un savoir linguistique : connaître le système

phonologique du français et les difficultés des élèves dans ce domaine (la lettre C prend ici sa valeur de base /k/).

- 36 Le troisième extrait correspond à l'ajout d'une péripétie à l'histoire : *C'est moi le plus fort*¹⁴. L'élève a pour consigne de chercher un personnage non présent dans l'album et de le faire dialoguer avec le loup qui se vante d'être le plus fort :

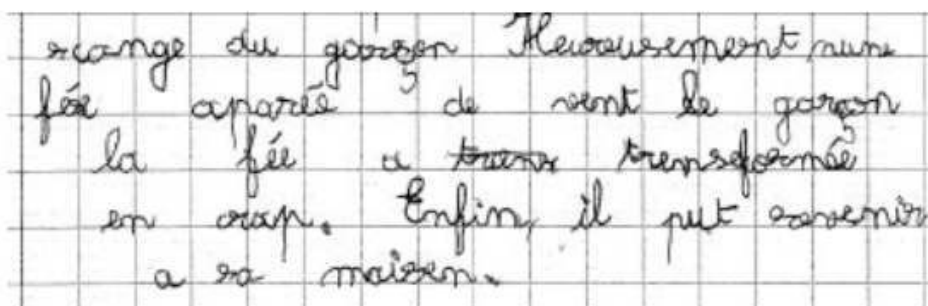


Doc. 7. Extrait du récit de Dounia (CP, mars)

- 37 Le dernier mot est au premier abord impossible à identifier mais une lecture à voix haute ainsi que la connaissance de la situation d'écriture rétablissent le discours direct (c'est le loup qui parle) et le savoir sur les façons d'écrire des élèves aide à déchiffrer le dernier mot, écrit à l'envers :
- 38 qui est [et] le plus fort [for] demande [de mande] le loup [poul]

Doc. 7 bis. Transcription d'une partie de l'extrait

- 39 Comme on peut l'observer sur ces extraits, les jeunes élèves en situation d'écriture, se retrouvent confrontés au système graphique si bien que les transcriptions vont fourmiller de notations touchant la segmentation graphique, la correspondance phonie-graphie, l'orthographe des morphèmes.
- 40 Ce quatrième extrait est un texte libre écrit par Milan sur son cahier de brouillon. Cet élève raconte l'histoire d'un petit garçon perdu dans la forêt, qui va rencontrer un vampire et être sauvé par une fée :

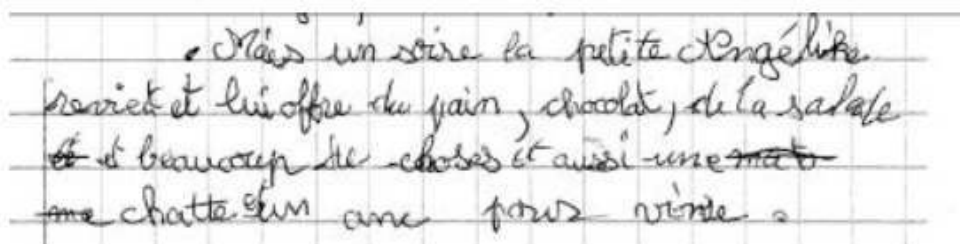


Doc. 8. Extrait du récit de Milan (CE1, septembre)

- 41 C'est la situation d'écriture et la connaissance des contes de fées qui incitent à voir dans « crap » une amorce de « crapaud » :
- 42 la fée a ~~trans~~ transformé [trensformée]
- 43 en crapaud [crap-].

Doc. 8 bis. Transcription d'une partie de l'extrait

- 44 L'extrait suivant illustre la question des syllabes incompréhensibles à l'écrit comme à l'oral. Plusieurs relectures par différents transcrip-teurs montrent que parfois il est possible de les déchiffrer au prix d'un travail long et coûteux. Dans l'exemple ci-dessous, l'élève est invité à écrire la suite de l'histoire *Le diable des rochers*¹⁵, à partir du moment où Angeline, la petite fille, disparaît. Le dernier mot du texte donne lieu à plusieurs interprétations. Nous avons emprunté au GARS la convention des multi-transcriptions (entre barres obliques et séparation par une virgule des différentes interprétations) et l'avons adaptée. À l'image des multi-écoutes, nous avons rétabli des multi-graphies :

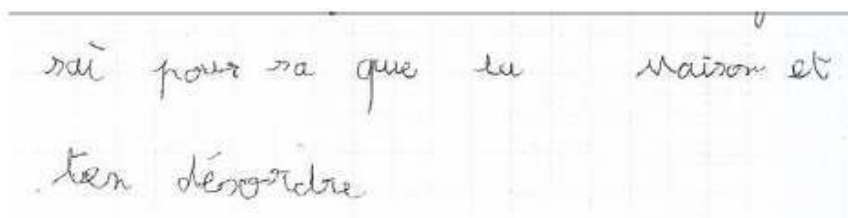


Doc. 9. Extrait du récit de César (CE1, janvier)

- 45 Mais un soir [soire] la petite Angéline
 46 revient et lui offre du pain, chocolat, de la salade
 47 ét et beaucoup de choses et aussi une ~~ma~~
 48 ~~ma~~ chatte et un âne [ane] pour /vivre, venir/.

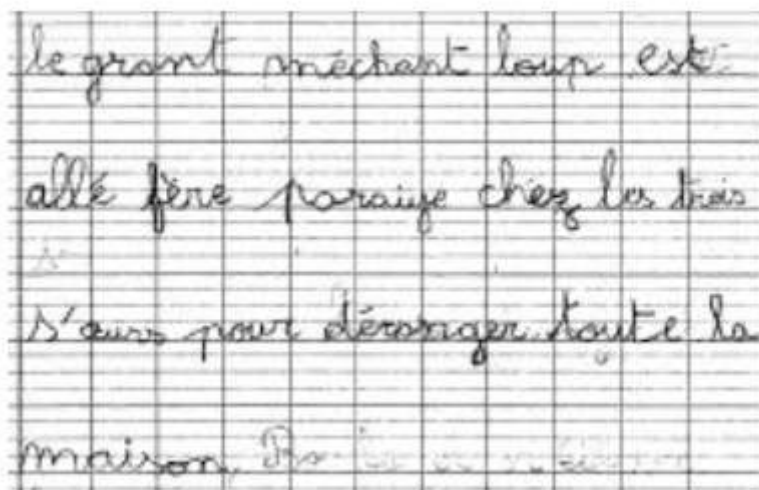
Doc. 9 bis. Transcription de l'extrait

- 49 Pour le chercheur, il est intéressant de voir jusqu'où peuvent aller ces divergences entre transcrip-teurs, comme cela avait été le cas pour le français parlé.
- 50 Mais toutes les conventions de transcription de l'oral ne peuvent être transférées aux écrits scolaires. Prenons le cas de la liaison. Dans les corpus de français parlé, des conventions particulières notent une liaison non standard remarquable (exemple : *on va écouter plutôt les vieux qui =z= ont vécu*) et une absence remarquable de liaison (exemple : *c'étaient trois # amis*). Or dans les textes scolaires, ce sont les élèves qui écrivent eux-mêmes les liaisons (Cappeau & Roubaud, 2005, p. 14) et ils peuvent le faire de différentes manières. Dans la majorité des cas, ils soudent la consonne de liaison au mot qui suit (exemple suivant : *et ten désordre*), comme dans l'extrait suivant qui est une suite d'histoire que l'élève doit écrire en observant une illustration de l'album¹⁶ :



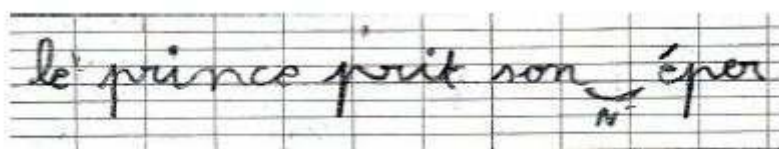
Doc. 10. Extrait du récit de Flavie (CE1, octobre)

- 51 Les élèves inventent aussi d'autres procédés pour marquer la liaison, comme celui qui suit (*les trois s'ours*), l'enfant résumant l'histoire de *La revanche du grand méchant loup*¹⁷ :



Doc. 11. Récit de Margot (CP, mars)

- 52 Mais il est rare qu'un élève de 7 ans ait conscience que cette liaison est celle de l'oral (qu'on entend mais qu'il ne faut pas écrire). C'est pourtant ce que semble indiquer Naïs lorsqu'elle produit un texte libre à la rentrée scolaire ; nous ne donnerons ici que l'extrait dans lequel figure la notation de la liaison :



Doc. 12.

Extrait du texte de Naïs (CE1, octobre)

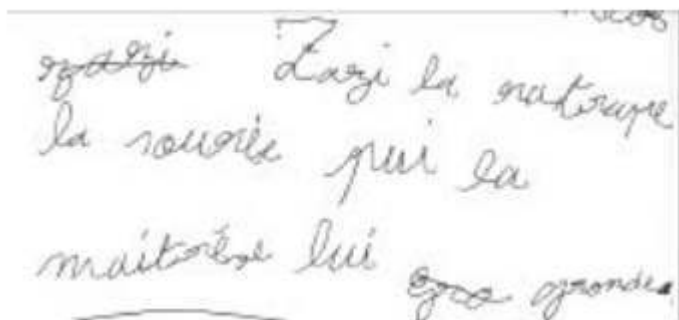
- 53 Adapter la convention utilisée à l'oral pour marquer une liaison (le signe =) n'est pas satisfaisant : *est =t= en désordre, les trois =s'= ours, le prince prit son =n= épée*. Cette convention unique donnerait une homogénéité à ce phénomène alors qu'il correspond à différents procédés de marquage de la liaison à l'écrit. Le guide du transcripteur devra intégrer la spécificité des écrits scolaires et les chercheurs devront trouver des solutions adaptées pour ce type de phénomène.
- 54 Mais il est un domaine plus sensible pour le transcripteur : celui de la morphosyntaxe car les solutions envisagées demandent toute une réflexion méthodologique propre aux écrits scolaires.

2.2 La transcription aux prises avec la morphosyntaxe des écrits scolaires

- 55 Nous voulons soulever ici quelques problèmes qui ont surgi au moment de la transcription des textes et poser des éléments de réflexion face à d'apparentes erreurs morphosyntaxiques.

Le choix des pronoms

- 56 Dans le cas de l'écriture d'une suite de récit ayant comme personnage principal Zazie la sorcière¹⁸, Grégoire emploie le pronom « lui » à la place de « la », ce qui est fréquent chez les enfants avant 8 ans (Cappeau & Roubaud, 2005) :

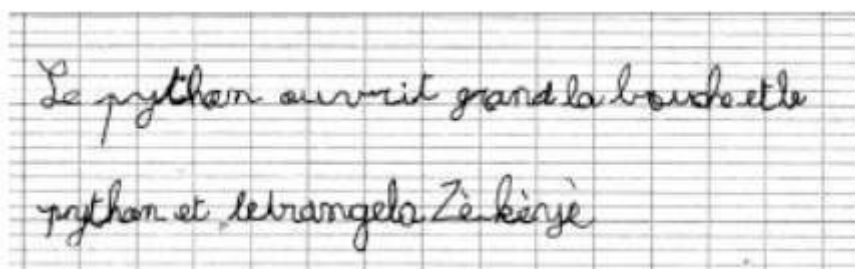


Doc. 13. Extrait du récit de Grégoire (CE1, décembre)

- 57 Si dans la version informatisée le transcritteur rétablit la norme « la grande », alors le chercheur risque de perdre l'information sur les formes typantes. En conservant le pronom « lui », il est possible de dresser un catalogue de ces formes comme pour l'oral (cf. Blanche-Benveniste & Pallaud, 2001). Une grande quantité d'exemples permet en effet de constater que certaines fautes contre la norme, par leur masse même, donnent des indications précieuses sur l'organisation de la grammaire chez les enfants (Roubaud, 2014).

La présence des connecteurs

- 58 L'emploi des connecteurs questionne aussi la transcription. Dans cette suite de récit (« Imaginez ce que va faire Zékéyé face au python¹⁹ ») écrite par Maëva, le connecteur « et » placé devant « le python » est en apparence fautif puisqu'il sépare le sujet « le python » de son verbe « étrangla » :



Doc. 14. Récit de Maeva (CP, mars)

- 59 Si dans la version informatisée le transcripneur supprime le second « et », le chercheur perd alors la trace de la construction des syntagmes sur les deux axes du langage, comme le fait apparaître la mise en grille du texte, procédé élaboré au sein du GARS pour le français parlé :

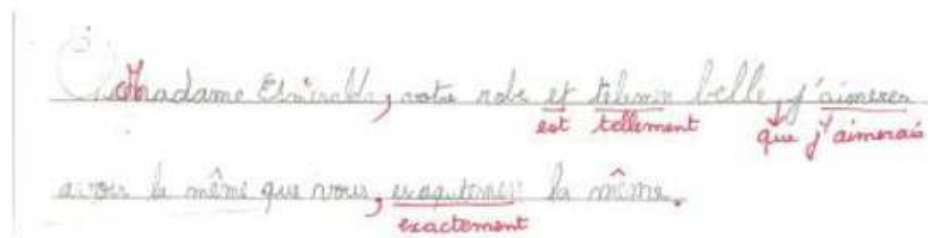
	Le python	ouvrit grand la bouche	
et	le python		
et		l'étrangla	Zékéyé

Doc. 14 bis. Mise en grille du récit
(dont l'orthographe a été corrigée)

- 60 Deux piétinements sont rendus visibles grâce à cet outil : l'un sur l'axe paradigmatique (correspondant à la recherche par l'élève des actions du python), marqué par la reprise de « et » et l'autre sur l'axe syntagmatique, marqué par le double marquage de l'objet (par un pronom « l' » et un nom « Zékéyé »). Comme pour le français parlé, il est important de « conserver trace de toutes ces apparentes scories qui donnent de précieux renseignements sur la construction des syntagmes » (Blanche-Benveniste, 2000, p. 23).

Les cas de la parataxe

- 61 Dans le cas de l'écriture d'une lettre à un personnage célèbre, Laurie écrit ce début de texte :

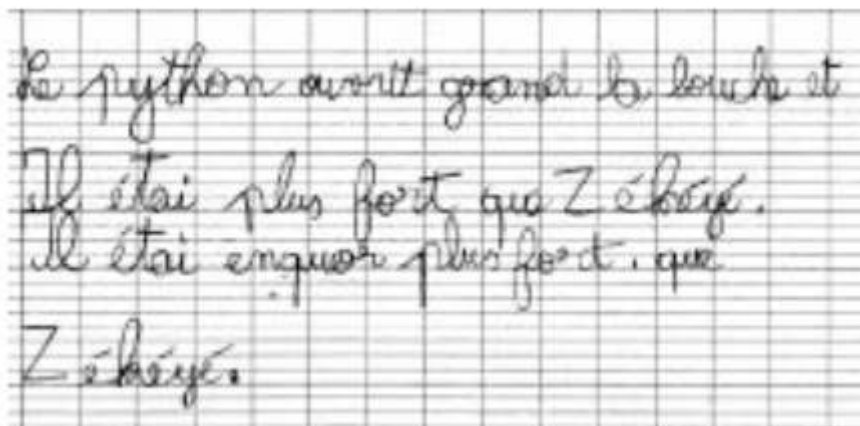


Doc. 15. Extrait de la lettre de Laurie (CE1, février)

- 62 Ce qui nous intéresse ici, c'est la correction de l'enseignant, qui ne se contente pas de corriger l'orthographe des mots et la ponctuation mais rétablit aussi la conjonction *que*, annulant ainsi le cas de parataxe. Avec de tels procédés, nous risquons de perdre toute trace de l'élaboration du discours à l'écrit et c'est la raison pour laquelle dans la version informatisée, nous maintenons ces absences de signes de ponctuation et conservons les cas de parataxe.

Les organisations macrosyntaxiques

- 63 Dans cet autre texte de la suite du récit de *Zékéyé et le python*, Prunile produit deux fois la même construction « Il était plus fort que Zékéyé » mais ne la ponctue pas de la même façon :



Doc. 16. Récit de Prunile (CP, mars)

- 64 Produire une première fois le comparatif « plus fort que Zékéyé » puis une seconde fois en le coupant par un point « plus fort. que Zékéyé » n'est pas fortuit et demande une analyse syntaxique des deux énoncés. Dans le premier cas (*Il était plus fort que Zékéyé.*), nous sommes dans le domaine des éléments régis par le verbe (en microsyntaxe). Dans le second cas (*Il était encore plus fort. que Zékéyé.*), l'analyse se fait en macrosyntaxe : l'élément « que Zékéyé » est un ajout à la construction précédente. Supprimer le point entre « fort » et « Zékéyé » (donc corriger la ponctuation) reviendrait à annuler les configurations discursives que présentent les écrits scolaires.
- 65 Toutes ces apparentes erreurs demandent une véritable réflexion sur les conventions qui « éclairent sur les filtres adoptés et assumés » (Cappeau, 2007, p. 26).

4. Conclusion

- 66 Comme pour le français parlé, le travail de constitution d'un grand corpus de textes scolaires est fructueux à condition que toutes les étapes, de la collecte des données à l'analyse, soient abordées avec la même rigueur et le même souci de réflexivité.
- 67 Les écrits scolaires comportent des spécificités : le recueil de métadonnées est à élaborer avec soin, les erreurs ne sont pas toutes à traiter de la même façon car certaines corrections annuleraient, et par là-même nieraient, les modes d'élaboration du discours par les élèves.
- 68 L'émergence de la linguistique des corpus de textes scolaires comme domaine de recherche doit obligatoirement reposer sur la formation de linguistes et d'informaticiens linguistes. Il y a au-delà du fait de devoir transcrire les textes d'élèves à des fins d'investigation de véritables enjeux linguistiques qui se dessinent.

BIBLIOGRAPHIE

- Baude O. (2006). *Corpus oraux. Guide des bonnes pratiques*. Paris : CNRS Éditions.
- Baude O. & Dugua C. (2011). « (Re)faire le corpus d'Orléans quarante ans après : "quoi de neuf, linguiste ?" », *Varia*, 10 : 99-118.
- [Reichler-]Béguelin M.-J. (1989). « Comment se servir des "erreurs" dans l'enseignement du texte écrit ? » *CARESP* (Journal du Cartel Romand des Associations du corps enseignant secondaire et professionnel), 19 : 4.
- Biber D., Johansson S., Leech G., Conrad S. & Finegan E. (1999). *Longman Grammar of Spoken and Written English*. New-York : Longman.
- Bilger M. (2000). *Corpus : Méthodologie et applications linguistiques*. Paris : Champion.
- Blanche-Benveniste C. (1996). « De l'utilité du corpus linguistique ». *Revue française de linguistique appliquée*, vol. I(2) : 25-42.
- Blanche-Benveniste C. (2000). « Corpus de français parlé », in M. Bilger (éd.) *Corpus : Méthodologie et applications linguistiques*. Paris : Champion, 15-25.
- Blanche-Benveniste C. (2010). *Le français. Usages de la langue parlée*. Leuven-Paris : Peeters.
- Blanche-Benveniste C. & Jeanjean C. (1987). *Le français parlé. Transcription et édition*. Paris : Didier Érudition, Institut national de la Langue française.
- Blanche-Benveniste C. & Pallaud B. (2001). « Le recueil d'énoncés d'enfants : enregistrements et transcription », *Recherches Sur le Français Parlé*, 16 : 11-37.
- Caddéo S. (2013). « La transcription orthographique pour l'analyse des données orales : un fragile compromis », in R. Druetta (éd.) *Claire Blanche-Benveniste. La linguistique à l'école de l'oral*. Sylvain-les-Moulins : Éditions Gerflint, 27-37.
- Cappeau P. (2007). *De la constitution à l'exploitation des données : Réflexions et illustrations*. Habilitation à Diriger des Recherches, Paris : Université Paris X Nanterre [non publié].
- Cappeau P. (2008). « Perception et reconstruction », in M. Bilger (éd.) *Données orales : les enjeux de la transcription*. Perpignan : Presses universitaires de Perpignan, 235-247.
- Cappeau P. & Roubaud M.-N. (2005). *Enseigner les outils de la langue avec les productions d'élèves*. Paris : Bordas.
- Culioli A. (2002). *Entretiens sur la linguistique*. Paris : Klincksieck.
- David J. (2000). « Étudier les textes d'enfants : revue de travaux », in C. Fabre-Cols (éd.) *Apprendre à lire des textes d'enfants*. Bruxelles : De Boeck Duculot, 274-288.
- David J. (2005). « Catégories langagières et cognitives. L'exemple des premiers apprentissages de l'écriture », *Le français aujourd'hui*, 151 : 81-94.
- Doquet C. (2013). « Dépasser l'impossible alliance : quelles inter-actions entre production écrite et maîtrise de la langue ? », *Le français aujourd'hui*, 181 : 119-130.
- Elalouf M.-L. & Boré C. (2007). « Construction et exploitation de corpus d'écrits scolaires », *Revue française de linguistique appliquée*, vol. XII(1) : 53-70.

- Équipe DELIC. (2004). « Présentation du Corpus de référence du français parlé », *Recherches sur le français parlé*, 18 : 11-42.
- Fabre-Cols C. (2000). « De la situation de production à l'interprétation du texte : contexte, matériau, lisibilité », in C. Fabre-Cols (éd.) *Apprendre à lire des textes d'enfants*. Bruxelles : De Boeck-Duculot, 19-55.
- Gadet F. (2000). « Derrière les problèmes méthodologiques du recueil des données », in M. Bilger (éd.) *Linguistique sur corpus. Études et réflexions*. Perpignan : Presses Universitaires de Perpignan.
- Gasquet-Cyrus M. (2013). « Peut-on écrire l'accent marseillais ? Analyse sociolinguistique de l'oral stylisé dans un corpus de littérature contemporaine », *Travaux interdisciplinaires sur la parole et le langage*, 29 [en ligne].
- Habert B. (2000). « Des corpus représentatifs : de quoi, pour quoi, comment ? », in M. Bilger (éd.) *Linguistique sur corpus. Études et réflexions*. Perpignan : Presses Universitaires de Perpignan, 11-58.
- Jaffré J.-P. (2003). « Les commentaires métagraphiques », *Faits de langues*, 22 : 67-76.
- Lee D. (2001). « Genres, registers, text types and styles : clarifying the concepts and navigating a path through the BNC Jungle », *Language Learning and Technology*, vol. 5(3) : 37-72.
- Masseron C. (2011). « L'analyse linguistique des écrits scolaires », *Pratiques*, 149-150 : 129-162.
- Mondada L. (2002). « Pratiques de transcription et effets de catégorisations », *Cahiers de praxématique*, 39 : 45-75.
- Rastier F. (2005). « Enjeux épistémologiques de la linguistique de corpus », in G. Williams (éd.) *La linguistique de corpus*. Rennes : Presses Universitaires de Rennes, 31-45.
- Roubaud M.-N. (2014). *De la description de la langue à son enseignement : Du terrain au laboratoire... du laboratoire au terrain*. Habilitation à diriger des recherches. Grenoble : Université de Stendhal Grenoble 3 [non publié].
- Williams G. (2005). *La Linguistique de corpus*. Rennes : Presses Universitaires de Rennes.

NOTES

1. Au sein du GARS (créé en 1977 autour de Claire Blanche-Benveniste) depuis 1986 puis au sein de l'équipe DELIC (Description linguistique informatisée des corpus) à partir de 2001.
2. Ce corpus fut constitué au départ dans le cadre de mon activité d'enseignante en école primaire.
3. Jusqu'aux journées à Paris en mars et mai 2015 et à Toulouse en juin 2015 qui ont rassemblé des chercheurs (de différentes universités françaises) autour de la problématique de la constitution d'un corpus d'écrits scolaires.
4. Le corpus Charolles (d'écrits scolaires) constitué à Toulouse autour de C. Garcia-Debanco comporte un questionnaire à remplir par l'élève, suite à sa production.
5. À ce jour, notre corpus dépasse les 1500 textes.
6. La réflexion sur la caractérisation des situations d'écriture se poursuit actuellement.
7. Par « profil », nous entendons : langue de l'élève, suivi (orthophoniste ou autre soutien)...
8. Par « contexte », nous entendons : école rurale, de banlieue, dans un quartier résidentiel, HLM, mixte, dans un quartier défavorisé, en REP ainsi que des renseignements sur la classe (milieu homogène, hétérogène).
9. Nous avons aussi des textes que nous avons appelés « textes libres », écrits par les élèves sans aucune consigne et sur un temps libre en classe.

10. L'équipe Ecriscol (Laboratoire Clesthia - Université La Sorbonne nouvelle Paris 3) en est arrivée à la même conclusion (cf. sa base de métadonnées sur le site Ecriscol).
 11. Nous rejoignons les préoccupations de certains auteurs qui ont proposé une version informatisée de manuscrits d'élèves (Fabre-Cols, 2000, p. 21 ; David, 2005, p. 88).
 12. Méthode de lecture *Ribambelle CP*, Hatier.
 13. *Porculus*, Arnold Lobel, L'école des loisirs, 2003.
 14. *C'est moi le plus fort*, Mario Ramos, Pastel-École des loisirs, 2001.
 15. *Le diable des rochers*, Grégoire Solotareff, École des Loisirs, 1993.
 16. *Le loup sentimental*, Geoffroy de Pennart, École des Loisirs, 1999.
 17. *La revanche du grand méchant loup*, Allan Mac Donald, Mijade, 1998.
 18. *Zazie-Pas de magie à l'école*, Rose Impey & Katherine Mc Even, Hatier, 2010.
 19. *Zékéyé et le serpent python*, Nathalie Dieterlé, Hatier, 2004.
-

RÉSUMÉS

Notre expérience de constitution d'une base de données orale (au sein du *Groupe aixois de recherches en syntaxe*) nous a servi à constituer, à la fin des années 90, une base de données d'écrits scolaires manuscrits (élèves de 5 à 11 ans). L'article rend compte des questions inhérentes à la constitution d'un tel corpus : du recueil de données à leur transcription et à leur édition. Si les méthodes d'analyse de l'oral peuvent résoudre certains des problèmes rencontrés, elles ne peuvent le faire sur l'ensemble des textes. En effet, les écrits des élèves comportent des spécificités linguistiques qui interrogent les conventions de transcription et des décisions sont à discuter au sein de la communauté scientifique, comme nous le montrerons en nous appuyant sur des textes de jeunes élèves (6 à 8 ans).

The written French: transcription and editing. The case of school texts

Our experiment of building a database oral (in the *Groupe aixois de recherches en syntaxe*) was used to establish, in the late 90s, a school texts database (5-11 years). The article reports on the issues involved in the constitution of such data collection to their transcription and editing. If the methods of analysis for spoken language can solve some of the problems encountered, however they cannot do it on all texts. Indeed, the writings of children include some linguistic specificities so the conventions of transcript are to be discussed within the scientific community, as we shall show from texts young pupils (6-8 years).

INDEX

Mots-clés : écrits scolaires (5 à 11 ans), linguistique de corpus, base de données, transcription

Keywords : school texts (5-11 years), corpus linguistics, database, transcription

AUTEUR

MARIE-NOËLLE ROUBAUD

Aix-Marseille Université - ADEF EA 4671, Marseille, France