

Protocole pour l'élaboration d'une version normée de productions de texte, à partir de leur transcription

De la transcription à sa normalisation

1. Éléments retenus

Ne sont retenus dans la normalisation que les éléments textuels produits par l'enfant : les mots, ou groupes de lettres, et la ponctuation.

2. Éléments non retenus

Sont absents de la normalisation les balises et autres marques spécifiques à la transcription, tels que <revision/>, /, //#. La mention de dessin ou tout autre élément extratextuel n'est pas conservée dans la normalisation.

Principes de normalisation

1. Segmentation en mots, orthographe et accords

L'orthographe est normée (*était* ➡ *était* et *foie* ➡ *fois* dans l'exemple (1)).

(1) Il était une foie (568, CP)

[normalisation] Il était une fois

En cas de mot concerné par les rectifications orthographiques de 1990, le choix de l'apprenant est respecté.

(2) Ses maitres était très gentille. (1557, CE1)

[normalisation] Ses maitres étaient très gentils.

(3) c'est son père quil fait le maître. (2937, CE1)

[normalisation] c'est son père qui fait le maître.

L'orthographe des interjections les plus courantes est normée (exemple (4)).

(4) comme par surprise BOOM le loup le mangea !!! (825, CM1)

[normalisation] comme par surprise BOUM le loup le mangea !!!

La segmentation en mots est rétablie (*lepren* ➡ *le prend* dans l'exemple (5)).

(5) il le pren (563, CP)
[normalisation] il le prend

Les accords en genre, en nombre (*les chaton* ➡ *les chatons* dans l'exemple (6)) et en personne (*tu est* ➡ *tu es* dans l'exemple (7)) sont rétablis.

(6) et les chaton aussi (1143, CP)
[normalisation] et les chatons aussi

(7) pourquoi tu est venu (1866, CE2)
[normalisation] pourquoi tu es venu

3. Verbes et conjugaison

Les temps des verbes sont conservés, le critère oral primant sur le critère écrit : dans l'exemple (8) on rétablit l'imparfait que l'élève a orthographié *er*.

(8) il tombe et il pleurer. (586, CP)
[normalisation] Il tombe et il pleurait.

La cohérence des temps n'est pas rétablie (exemple (8) et exemple (9)).

(9) qan stest ler dalé fair la siest le pe ti chaton ne voulest pas fair la siet (1706, CP)
[normalisation] quand c'est l'heure d'aller faire la sieste <segmentation/> le petit chaton ne voulait pas faire la sieste

La morphologie normée est rétablie (exemple (10) et exemple (11)).

(10) il entendu (94, CE2)
[normalisation] il entendit

(11) elle a ouvris (70, CE2)
[normalisation] elle a ouvert

2. Sur le plan lexical

Lorsqu'un mot est omis :

a. S'il est identifiable de manière non ambiguë, il est rétabli. C'est le cas, par exemple, du *ne* de négation (exemple (12)) ou du déterminant *la* dans l'exemple (13)).

(12) mai il a pas vus le trautoir (562, CP)

[normalisation] Mais il n'a pas vu le trottoir.

(13) la maman le pran par bouche (200, CP)

[normalisation] la maman le prend par la bouche

b. S'il est non identifiable ou que les possibilités sont multiples, mais que la catégorie du mot absent est identifiable sans ambiguïté, il est marqué par une balise <omission type="TYPE"/>, où TYPE est remplacé par la catégorie du mot absent (exemple (14)).

(14) il été t'une foi un petit chat se promené acoté de sa litière (1055, CP)

[normalisation] Il était une fois un petit chat <omission type="pronom"/> se promenait à côté de sa litière

Dans cet exemple, une reprise pronominale est nécessaire pour lier la première proposition « Il était une fois un petit chat » et la deuxième proposition « se promenait à côté de sa litière ». Plusieurs possibilités peuvent être envisagées :

- Il était une fois un petit chat *qui* se promenait à côté de sa litière
- Il était une fois un petit chat, *il* se promenait à côté de sa litière

Pour ne pas avoir à choisir entre ces différentes possibilités, on indique la nécessité de la reprise pronominale par la balise <omission type="pronom"/>.

Les catégories retenues pour cette balise sont les suivantes : verbe, nom, pronom, adjectif, adverbe, préposition, déterminant.

Lorsqu'un choix lexical est jugé incorrect de manière non ambiguë, le mot concerné est remplacé (*les/la* dans l'exemple (15), *voie/voix* dans l'exemple (16)).

(15) Il était une fois dans un village perdu une sorcière qui travailler dur, elle se lever à 3h00 du matin et finir à 15h00 de l'après-midi. Un jour elle eu une idée pour les renplasser de son travaille (1111, CM1)

[normalisation] Il était une fois dans un village perdu une sorcière qui travaillait dur, elle se levait à 3h00 du matin et finissait à 15h00 de l'après-midi. Un jour elle eut une idée pour la remplacer de son travail

(16) il manquer juste de l'abiller comme elle est lui mettre sa voie... (1111, CM1)

[normalisation] il manquait juste de l'habiller comme elle et lui mettre sa voix...

Lorsqu'un mot est identifié comme surnuméraire de manière non ambiguë et de manière non stylistique, il est supprimé (exemple (17)).

(17) le chat paniquai le chat le chat dit « c'est pour quoi » (63, CE2)

[normalisation] le chat paniquait le chat dit « c'est pourquoi »

Lorsque la répétition de la forme « très » est vue comme un procédé stylistique d'accentuation (exemple (16)), elle est conservée.

(18) le loup avait faim très très faim (1155, CM1)

[normalisation] le loup avait faim très très faim

Les constructions issues de l'oral, telles que :

- GN{1} + Pronom{2} + Verbe + GN{2} (exemple (12)) où {2} renvoie à une entité unique,
- GN {1} + Pronom{1} + Verbe (exemple (13)) où {1} renvoie à une entité unique

sont conservées.

(19) Il le mange le chat (exemple non attesté)

[normalisation] Il le mange le chat

(20) Le cha il mache (61, CP)

[normalisation] Le chat il marche

4. Organisation des mots dans la proposition

En cas d'inversion de mots, l'ordre est rétabli.

(21) Et le pour remercier le chat (67, CE2)

[normalisation] Et pour le remercier le chat

(22) (le 30 juin) 2100 elle pue l'emené au travail (1111, CM1)

[normalisation] (le 30 juin 2100) elle put l'emmener au travail

5. Ponctuation et segmentation en propositions

La majuscule en début de production et le point en fin de production sont rétablis.

Lorsqu'un point de fin de phrase est présent sans majuscule, la majuscule est rétablie.

(23) le cha et tonbe sur le tapi. le cha pler (48, CP)
[normalisation] le chat est tombé sur le tapis. Le chat pleure

Lorsqu'une ponctuation est manquante et que le mot suivant contient une majuscule, le point est rétabli.

(24) Le petit chat sest fé male Le peti chat a male (1531, CP)
[normalisation] Le petit chat s'est fait mal. Le petit chat a mal

En cas d'interrogation directe, le point d'interrogation est rétabli (exemple (25)).

(25) - bon jour la sorsiere commens allé vous. (1927, CE2)
[normalisation] <dialogue> - Bonjour la sorcière comment allez-vous ? </dialogue>

En cas d'absence de segmentation en propositions, celle-ci est marquée par la balise <segmentation/> (exemple (26)). Cette balise peut remplacer un signe de ponctuation ou un connecteur.

(26) il cur il tonbe partere il pleur il saute sur / un notre chat (1173, CP)
[normalisation] Il court <segmentation/> il tombe par terre <segmentation/> il pleure
<segmentation/> il saute sur un autre chat.

Dans cet exemple, plusieurs versions normées pourraient être envisagées, en voici quelques exemples :

- Il court. Il tombe par terre. Il pleure. Il saute sur un autre chat.
- Il court, il tombe par terre, il pleure et il saute sur un autre chat.
- Il court et il tombe par terre. Il pleure et il saute sur un autre chat.

Pour marquer ce manque de segmentation en propositions sans choisir un marqueur parmi les multiples possibilités, on utilise la balise <segmentation/>.

Les connecteurs valent pour marque de segmentation, la balise <segmentation/> ne sera donc pas ajoutée.

(27) Le chat sailoigne de sa maman est des bébé pan dans que sa maman dor ansuit
il tonbe est le bébé pleur est sa maman lanten pui sa maman ramas le chton. (1665, CP)

[normalisation] *Le chat s'éloigne de sa maman et des bébés pendant que sa maman dort ensuite il tombe et le bébé pleure et sa maman l'entend puis sa maman ramasse le chaton.*

6. Le cas de la virgule

En cas d'énumération, on rétablit la ponctuation usuelle, à savoir : tous les items sont séparés par des virgules, sauf l'avant-dernier et le dernier qui sont séparés par le connecteur « et ».

(28) alima gata et paulète alèrent se coucher (2978, CE1)

[normalisation] Alima, Gata et Paulète allèrent se couche

(29) mes mon chat qui été dans la marmite a eu telleman pere que il est sorti de la marmote a monter lechelle et courer dans tous sanse. (2978, CE1)

[normalisation] mais mon chat qui était dans la marmite a eu tellement peur qu'il est sorti de la marmite, a monté l'échelle et courait dans tous les sens.

(30) Il étai une foix dans la fôré un loup qui mangé tout le monde qui avai mange / troi cochon (1189, CE1)

[normalisation] Il était une fois dans la forêt un loup qui mangeait tout le monde et qui avait mangé trois cochons

L'énumération est le seul cas où une virgule peut être ajoutée.

En cas de virgule surnuméraire (exemple (26)) ou non attendue (exemple (27)), elle est supprimée.

(31) - C'est vrais aucun mal ?, vous ne manger plus les animaux ? (2978, CE1)

[normalisation] - C'est vrai aucun mal ? Vous ne mangez plus les animaux ?

(32) le chat noir c'est, enfuit de sa cage. (70, CE2)

[normalisation] le chat noir s'est enfui de sa cage.

En cas de virgule marquée par un signe de ponctuation non valide (exemple (33)), elle est rétablie.

(33) Et un beau jour comme dans les dessins animés, Les robots se sont rebellé. (1129, CE2)

[normalisation] Et un beau jour comme dans les dessins animés, les robots se sont rebellés.

Dans tous les autres cas, on n'ajoute aucun signe de ponctuation.

7. Dialogues

Les propos des personnages sont encadrés par les balises <dialogue> (début) </dialogue> (fin). Chaque nouveau tour de parole est encadré par ces balises.

(34) [...] le chat dit : Bonjour méchante et orible sorciere j'ai appris que tu allais dominer le monde. Oui c'est sa et comment t'appelle tu je m'appelle jean - pomme de terre. Sa te dit con fase équipe ensemble [...] (1336, CM1)

[normalisation] [...] le chat dit : <dialogue> Bonjour méchante et horrible sorcière j'ai appris que tu allais dominer le monde. </dialogue> <dialogue> Oui c'est ça et comment t'appelles-tu ? </dialogue> <dialogue> Je m'appelle Jean-Pomme de terre. </dialogue> <dialogue> Ça te dit qu'on fasse équipe ensemble ? </dialogue> [...]

Les ponctuations de dialogue existantes sont conservées, mais les ponctuations manquantes, comme la virgule, le point, le double-point, les guillemets ou encore les tirets, ne sont pas rétablies, sauf lorsqu'une marque double (ex. guillemets) est incomplète ou mal positionnée (exemple (33)), la deuxième marque est rétablie.

(35) il a dit « je vais te manger. Il prend (exemple non attesté)

[normalisation] il a dit <dialogue> « je vais te manger » </dialogue>. Il prend

(36) elle a dit «» je vait le raiquipérait. est anfin elle a réquipérait con chat. (937, CE1)

[normalisation] elle a dit <dialogue> « je vais le récupérer » </dialogue>. Et enfin elle a récupéré son chat.

En cas de proposition incise à l'intérieur d'un dialogue, la balise <dialogue> est fermée puis rouverte.

(37) - Mais quesque tu fais la dedans dit le loup sort vites ! (exemple non attesté)

[normalisation] <dialogue> - Mais qu'est-ce que tu fais là-dedans </dialogue> dit le loup <dialogue> sors vite ! </dialogue>

Les pensées ou idées des personnages sont également encadrées de balises.

(38) - elle eu une idée et si je me transphorma en chat !!! (1111, CM1)

[normalisation] Elle eut une idée <dialogue> et si je me transformais en chat !!! </dialogue>

La balise <dialogue> implique une segmentation, la balise <segmentation> est donc inutile dans le voisinage de la balise <dialogue>.

(39) et lui dit tu ma manqué (1337, CM1)

[normalisation] et lui dit <dialogue> tu m'as manqué </dialogue>

La majuscule est rétablie après un tiret marquant un nouveau tour de parole dans un dialogue.

(40) - ben je vais voir si le crocette son meilleur chez les voisin ! (1337, CM1)

[normalisation] Ben je vais voir si les croquettes sont meilleures chez les voisins !

8. Notations spécifiques

Lorsque un mot ou une phrase est non achevé en fin de production, on le ou la marque à l'aide de la balise <nonfini/>.

(41) et di maou maou qui r (1228, CP)

[normalisation] et dit « miaou miaou » qui r<nonfini/>

Les segments dont le sens ne peut être compris sont marqués à l'aide de la balise <incomprehensible/>.

(42) l ti chat pas c ese bébé cha (1363, CP)

[normalisation] le petit chat <incomprehensible/> bébé chat

Les titres en début de production sont encadrés par la balise <titre>.

(43) Le chat Il était une fois un chat (1845, CE2)

[normalisation] <titre>Le chat</titre> Il était une fois un chat

9. En cas de doute

En cas de doute du normalisateur, le passage concerné peut être encadré de la balise <unsure> ... </unsure>.

(44) il marche sur au bor de la plake. (1352, CP)

[normalisation] Il marche <unsure> au bord </unsure> de la plaque.

Point en réflexion : orthographe => avant ou après 1990 ?