LIDILEM

UNIVERSITÉ Grenoble Alpes

Mise à disposition d'un corpus d'apprenants : quelle méthodologie pour quels enjeux ?

Claire Wolfarth, Catherine Brissaud, Claude Ponton

Lidilem

Université Grenoble Alpes

I – Recueil et numérisation

Le projet Scoledit

Contexte

- Travaux en écriture moins fréquents qu'en lecture (Boscolo, 2008)
- Développement des recherches sur l'écriture
- Nécessité de grands corpus pour l'étude de l'apprentissage de l'écriture (Elalouf, 2009, Garcia-Debanc, 2013)
- Corpus souvent restreints ou spécifiques
- Emergence de projets de constitution de grands corpus (Elalouf et al., 2005 ; Gunnarsson-Largy et Auriac-Slusarczyk, 2013 ; Garcia-Debanc, 2014 ; David et Doquet, 2016)
- *Enquête « Lire – Écrire CP » de l'Institut Français de l'Éducation (2013-2015)*

Objectif

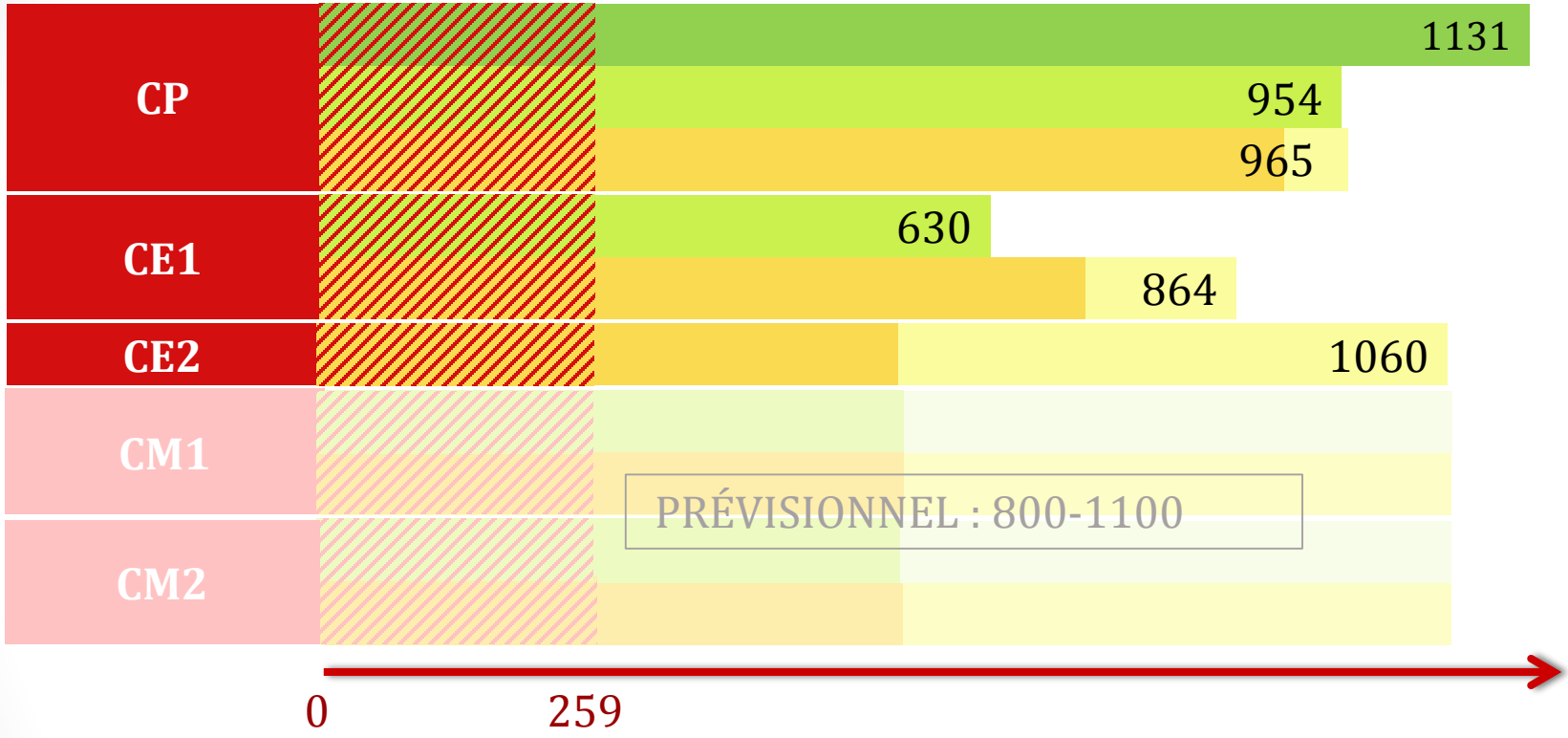
- Constitution d'un corpus scolaire longitudinal (CP-CM2) en complétant le recueil de l'enquête « Lire – Écrire CP » et mise à disposition de la communauté enseignante et scientifique

Le projet Scoledit

Enjeux

- Pour la linguistique :
  - Description en synchronie et en diachronie des caractéristiques des écrits d'apprenants de 6 à 11 ans
  - Meilleure connaissance des phénomènes d'acquisition de l'écriture (évolution de l'acquisition de la syntaxe, de l'orthographe, de la ponctuation, …)
- Pour la didactique:
  - Développement de séquences et de dispositifs à partir de ces connaissances
- Pour le traitement automatique des langues (TAL) :
  - Type de textes encore peu étudiés

Structure du corpus



Recueil du corpus

Recueil CP

Dictées

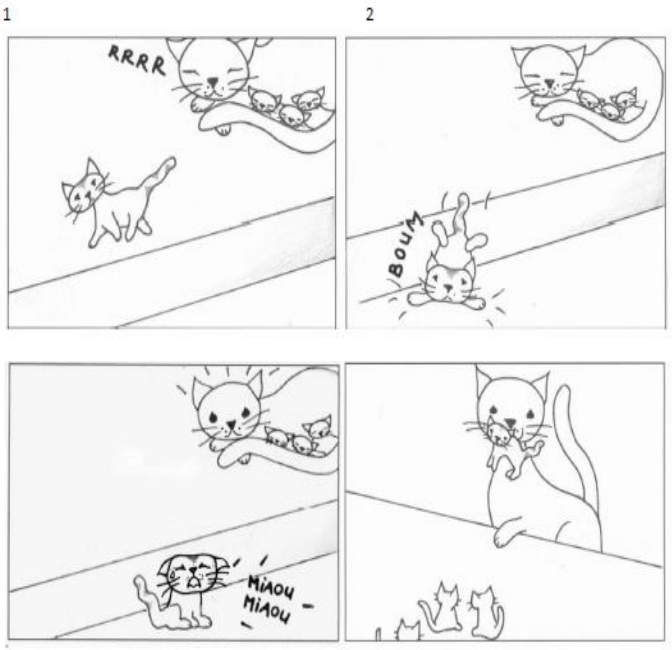
Septembre

lapin
rat
éléphant
Tom joue avec le rat.

lapin
rat
éléphant
Tom joue avec le rat.
Les lapins courent vite.

Juin

Production de texte (15 min)



Recueil du corpus

Recueil CE1 et CE2

Dictées (CE1)

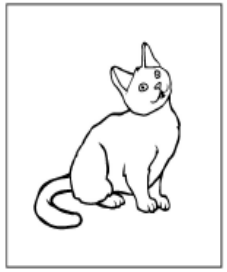
patin	récréation
pâtisson	charitable
capuchon	magnifique

En été, les salades vertes poussent dans les jardins.
Les jeunes canetons picorent le blé avec la poule noire.

Productions de textes
(CE1 : 20 min)
(CE2 : 30 min)

1

2



3

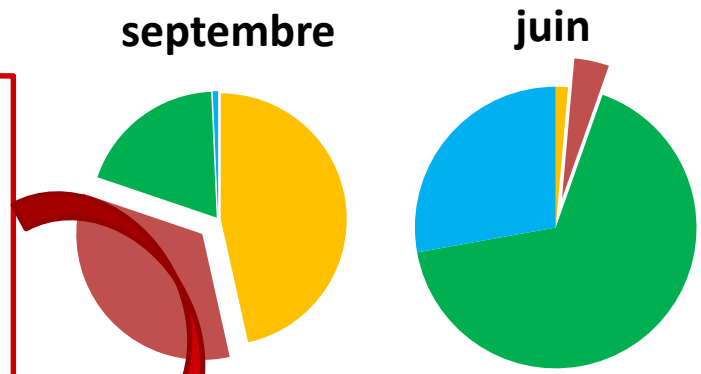
4

Données recueillies (CP)

(ex. élève 1558)

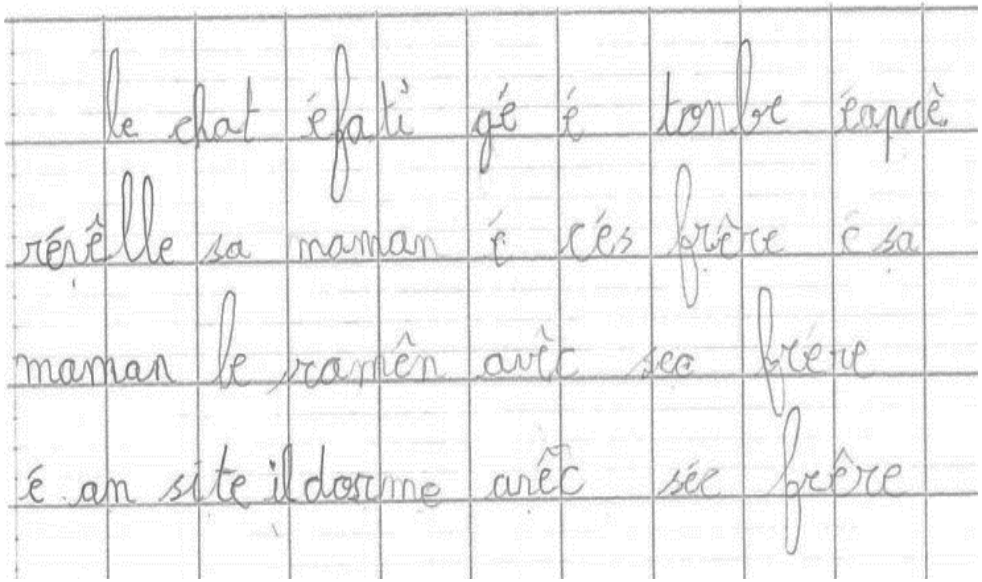
I – Recueil et numérisation
II – Méthodologie de l'analyse
III – Perspectives

Septembre 2013 (début CP)



- 0 à 3, écriture pré-syllabique
- 4 et 5, écriture syllabico-alphabétique
- 6 et 7, écriture alphabétique + ou – conforme à la forme orale
- 8, écriture orthographique

Juin 2014 (fin CP)



Données recueillies (CE1)

(ex. élève 1558)

Juin 2015 (fin CE1)

1/ patimp.
2/ patison.
3/ capuchons.
4/ récréations.
5/ charitable.
6/ manifice

~~Les~~ En été, les salade verte pouse dans les jardins. Les jaunes canetons picore le blé avec la poule noir.

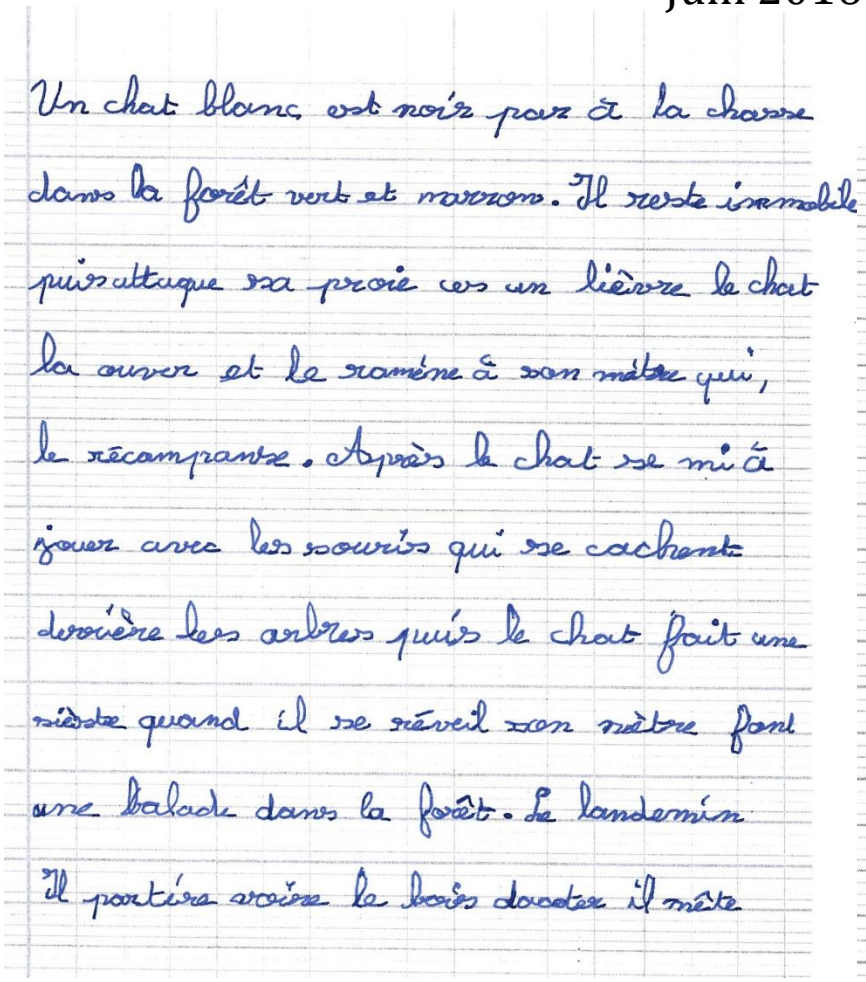
Le loup se ~~promem~~ promene dans les boit camp un chat ~~sor~~ sor din buson et lui di quefaitu je me promene pourquoi on peus se promene ensemble set une bonne idet nes pas pourcoi pas suivre la riveire il nia pas de chaseur sest sur sinan il metir desu et je cour baucoup maime que je me cache dans un buison et il me chere et tire par tout et je cour a nouvau.

il fau faire atantion il sont nonbre et mete des pillege de partot et on fait trais atantion il met des corde de par tout.

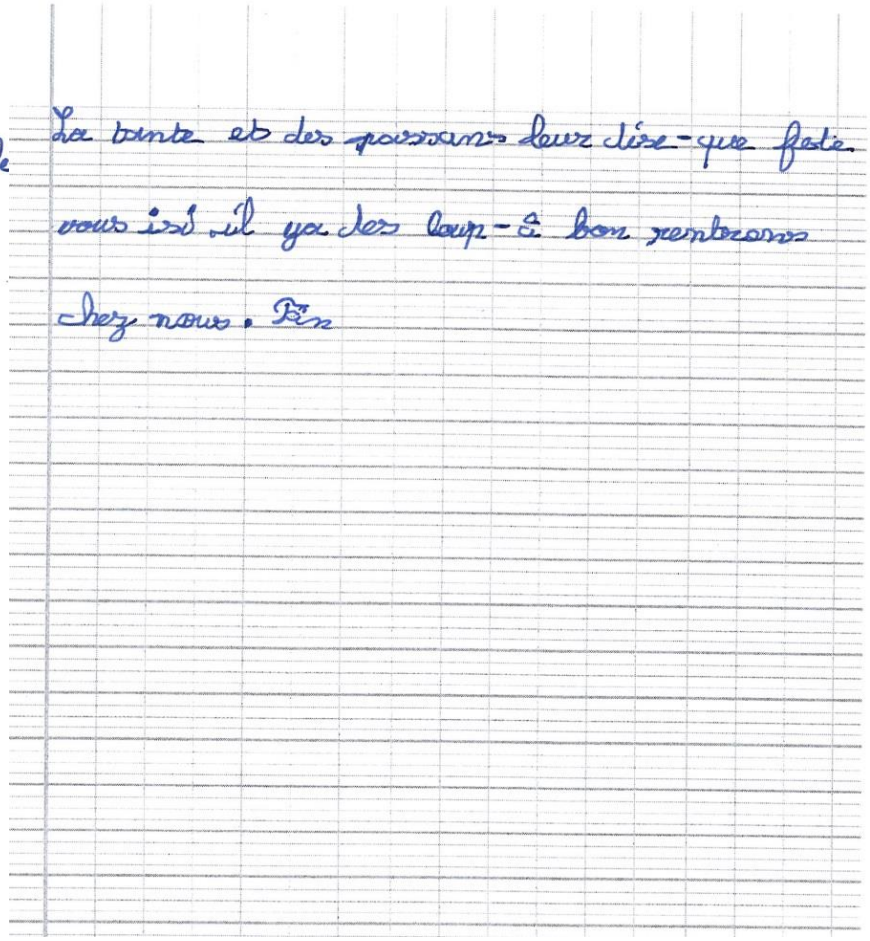
Données recueillies (CE2)

(ex. élève 1558)

Juin 2016 (fin CE2)



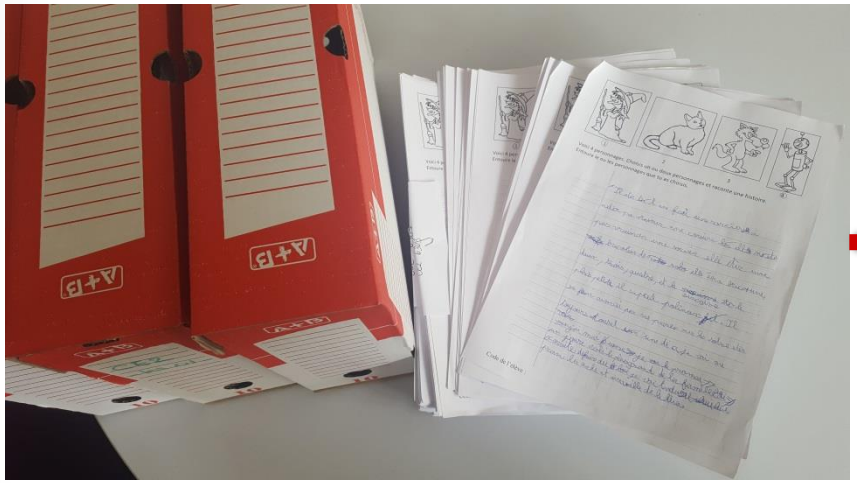
Un chat blanc est noir part à la chasse dans la forêt vert et marron. Il reste immobile puis attaque sa proie ces un lièvre le chat la auver et le ramène à son mètre qui, le récampanse. Après le chat se mi à jouer avec les souris qui se cachent derrière les arbres puis le chat fait une siècte quand il se réveil son mètre font une balade dans la forêt. Le landemin Il partira croire le bois docter il mète

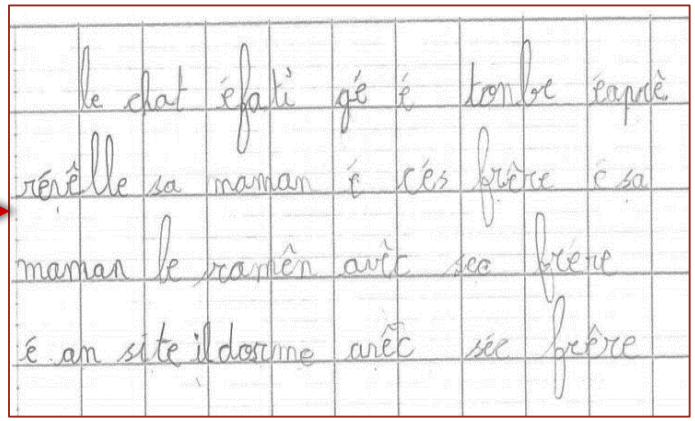


La tante et des passants leur dise-que faite vous ici il ya des loup-à bon rentrons chez nous. Fin

Numérisation du corpus

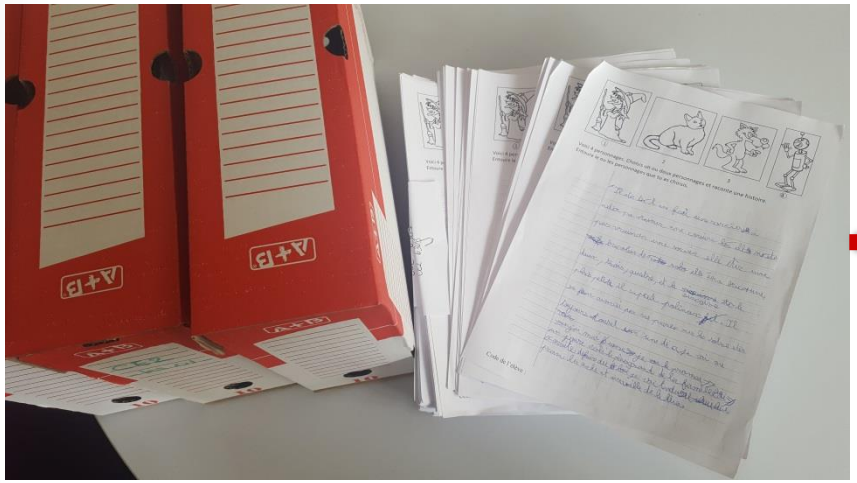
- **Scan du corpus**

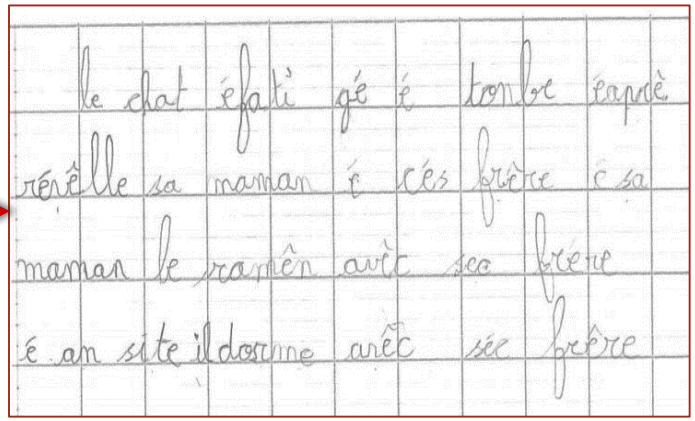




Numérisation du corpus

- **Scan du corpus**





- **Transcription**

le chat éfati gé é tonbe éaprê / révêlle sa maman é cés frêre é sa / maman le ramên avêc sec frére / é an site il dorme avêc séc frêre

lien

I – Recueil et numérisation

...nalyse

UNIVERSITÉ Grenoble Alpes

Scoledit (version CP)

Accueil Visualisation Commentaires A propos

Login : ClaireWolfarth | Déconnexion

Visualisation

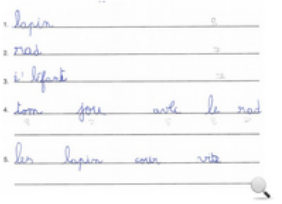
1 2 3 4 5 6 7 8 9 10 11 12

Recherche par :

Forme | N° d'élève | ok

Il y a actuellement à l'étude : 1169 élèves.

Elève 47

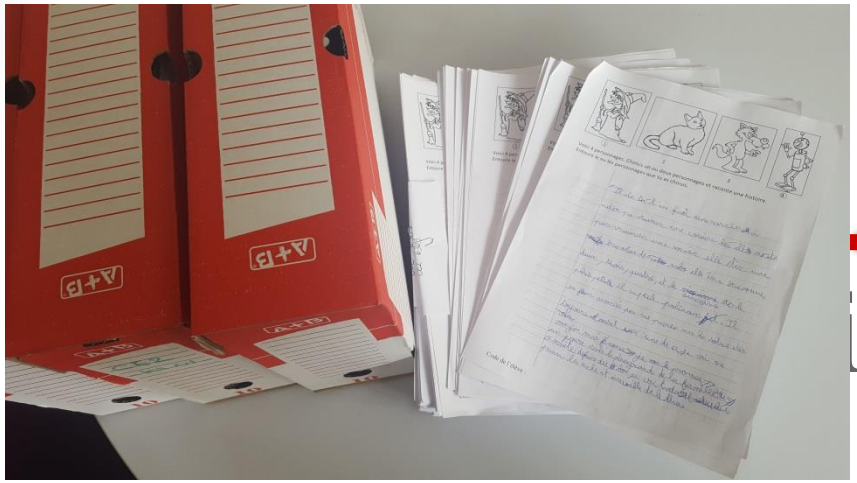
SCAN DICTÉE SEPTEMBRE	DICTÉE SEPTEMBRE	DICTÉE JUIN	SCAN DICTÉE JUIN
	LP RAPH LPIH LLULHLÏH	lapin rad é léfant tom jou avec le rad les lapin cour vite	

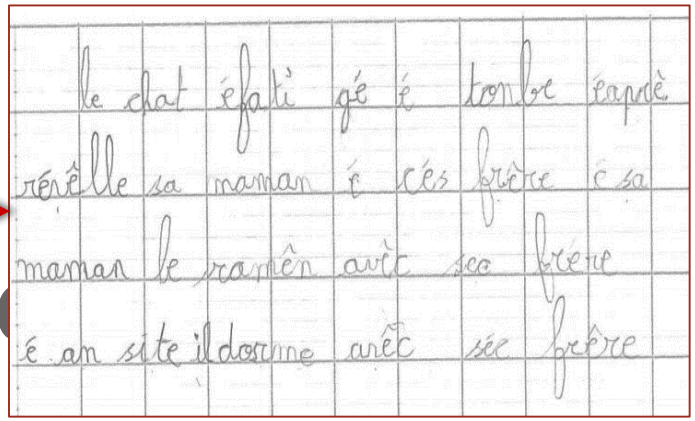
PRODUCTION	SCAN PRODUCTION
le chat et tonbe dant les zescalié et il pleur parse que il ses fe tre tre male et sa maman le pretg et il le lech.	

Elève 49

SCAN DICTÉE SEPTEMBRE	DICTÉE SEPTEMBRE	DICTÉE JUIN	SCAN DICTÉE JUIN
	W LAP U RA	lapin rat éléphant Tom jou avec le rat les lapin cors vite	

- **Scan du corpus**



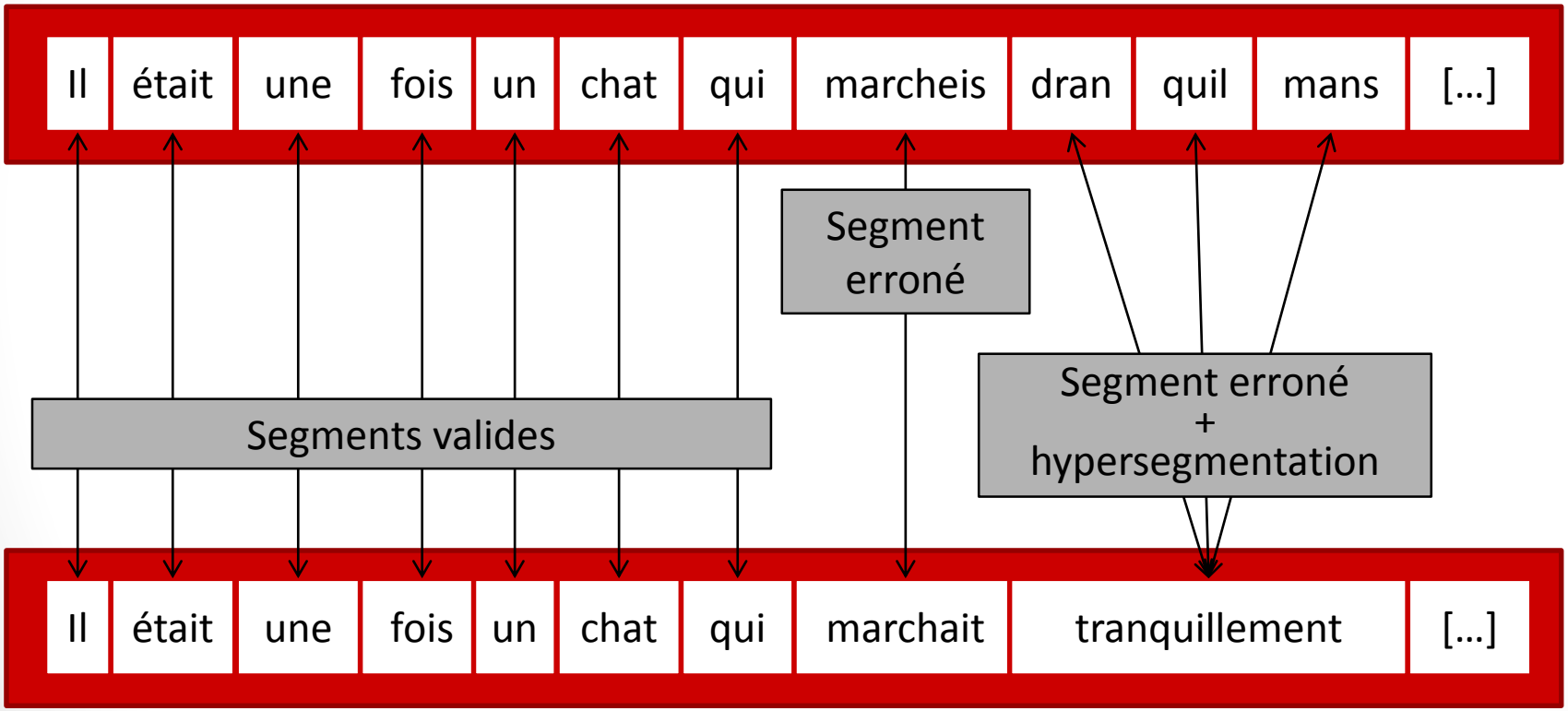


- **Transcription**

le chat éfati gé é tonbe éaprê / révêlle sa maman é cés frêre é sa / maman le ramên avêc sec frére / é an site il dorme avêc séc frêre

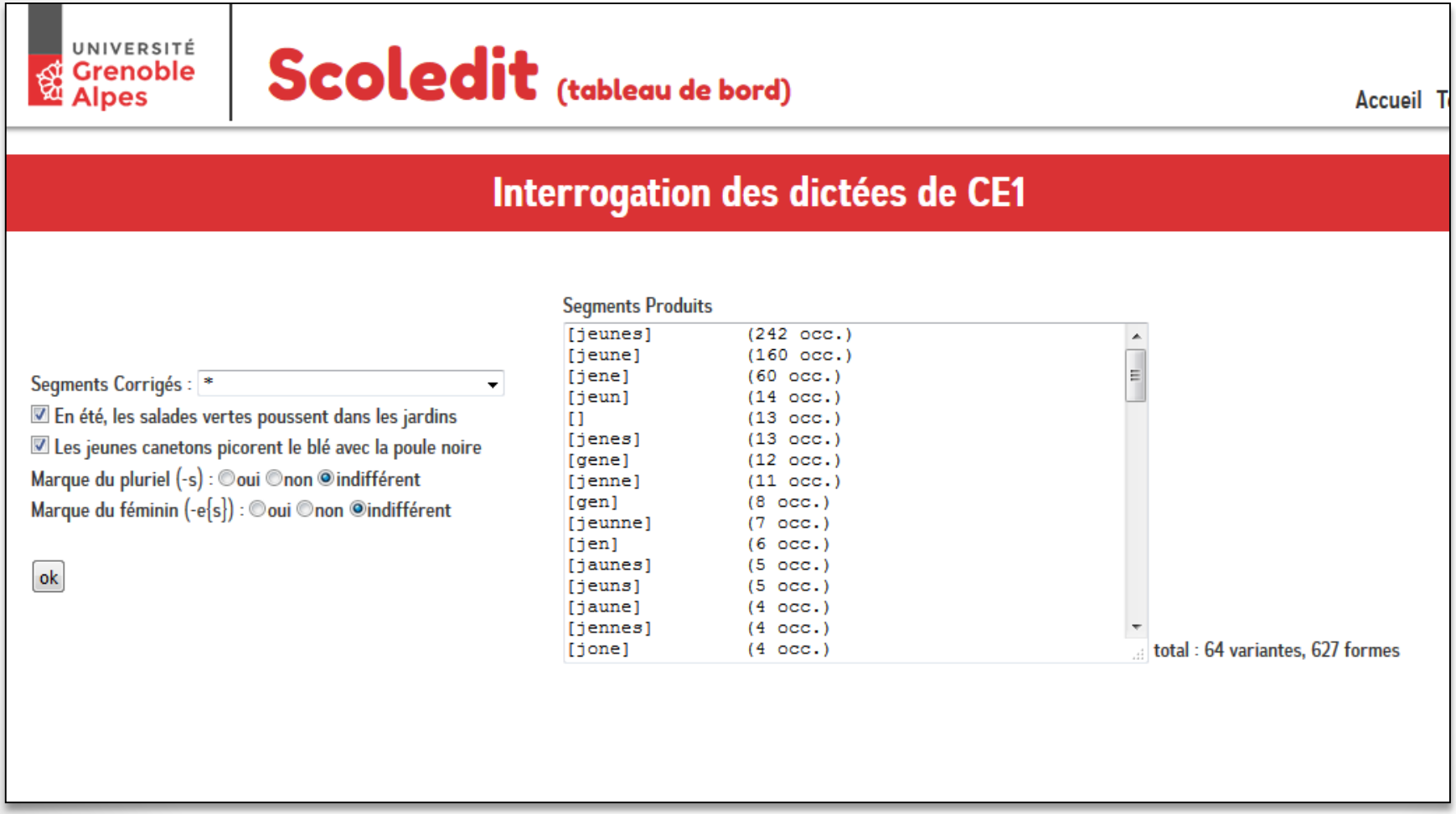
- **Analyse**

Analyse par comparaison

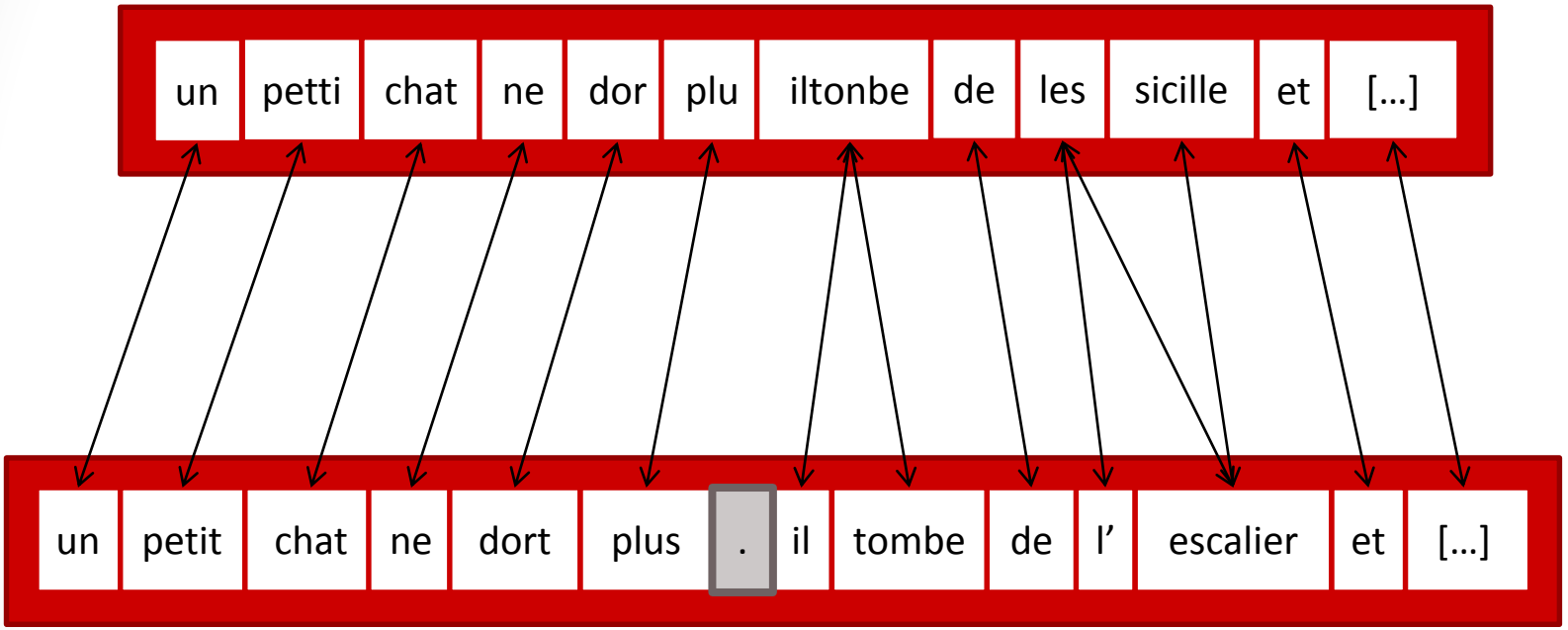


(élève 77 - production de CP)

lien

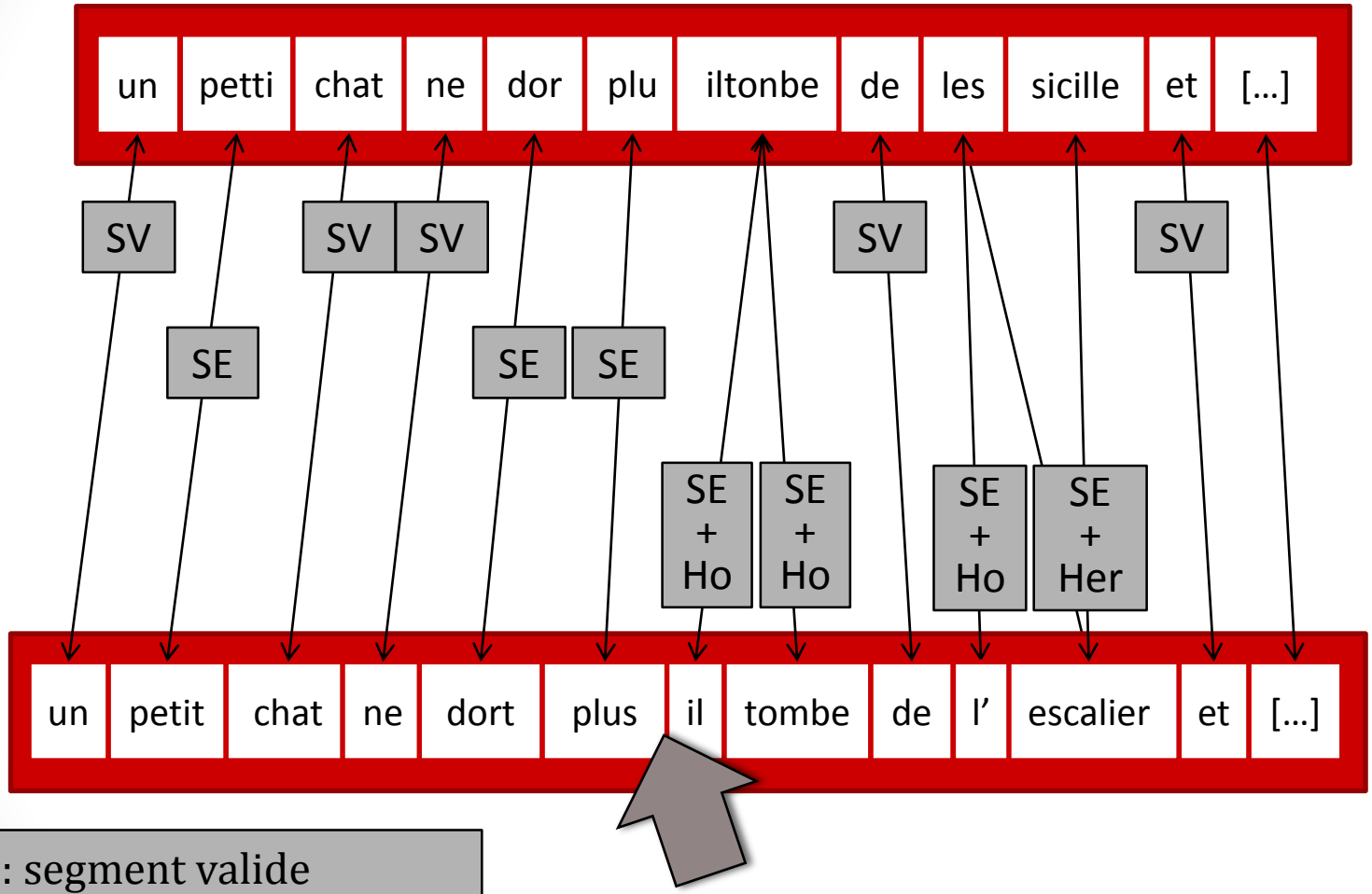

UNIVERSITÉ Grenoble Alpes
Scoledit (tableau de bord)
Accueil
Interrogation des dictées de CE1
Segments Produits
[jeunes] (242 occ.)
[jeune] (160 occ.)
[jene] (60 occ.)
[jeun] (14 occ.)
[] (13 occ.)
[jenes] (13 occ.)
[gene] (12 occ.)
[jenne] (11 occ.)
[gen] (8 occ.)
[jeunne] (7 occ.)
[jen] (6 occ.)
[jaunes] (5 occ.)
[jeuns] (5 occ.)
[jaune] (4 occ.)
[jennes] (4 occ.)
[jone] (4 occ.)
total : 64 variantes, 627 formes
Segments Corrigés : *
En été, les salades vertes poussent dans les jardins
Les jeunes canetons picorent le blé avec la poule noire
Marque du pluriel (-s) : oui non indifférent
Marque du féminin (-e{s}) : oui non indifférent
ok

Analyse par comparaison



(élève 93 - production de CP)

Analyse par comparaison



SV : segment valide
SE : segment erroné
Ho : hyposegmentation
Her: hypersegmentation

(élève 93 - production de CP)

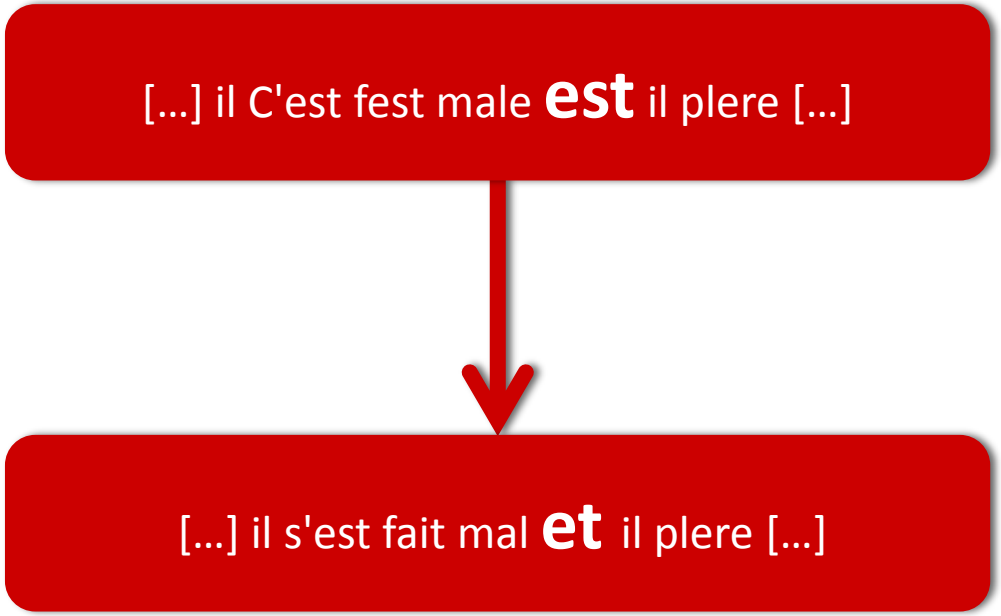
Le cas de la ponctuation

[…] Iltonbe de les sicille **et** il C'est fest male […]

[…] Il tombe de l'escalier **et** il s'est fait mal […]

(élève 93 - production de CP)

Le cas de la ponctuation



(élève 93 - production de CP)

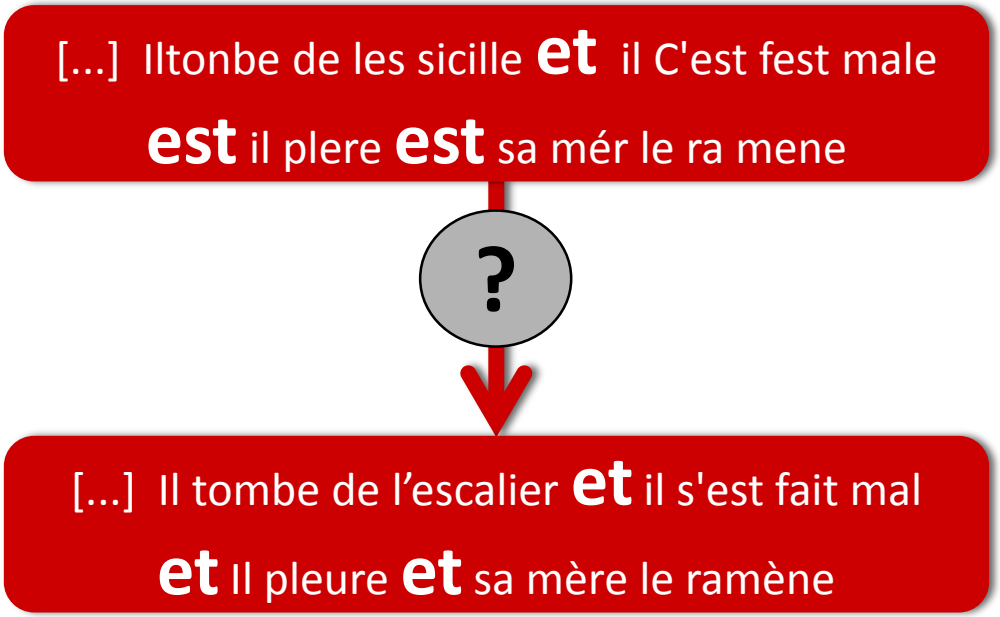
Le cas de la ponctuation

[…] il plere **est** sa mér le ra mene

[…] Il pleure **et** sa mère le ramène

(élève 93 - production de CP)

Le cas de la ponctuation



(élève 93 - production de CP)

Le cas de la ponctuation

[...] Iltonbe de les sicille **et** il C'est fest male **est** il plere **est** sa mér le ra mene

?

- **Proposition 1 : Alterner avec des *points***

Exemple 1 : Il tombe de l'escalier. Il s'est fait mal. Il pleure. Sa mère le ramène .
Exemple 2 : Il tombe de l'escalier. Il s'est fait mal. Il pleure et sa mère le ramène.
Exemple 3 : Il tombe de l'escalier et il s'est fait mal. Il pleure et sa mère le ramène .

➢ Choix différents selon les transcripteurs

➢ Cohérence avec la production de l'enfant :

un petti chat ne dor plu iltonbe/ de les sicille et il C'est/ fest male est il plere/est sa mér le ra mene

Le cas de la ponctuation

- **Proposition 1 : Alterner avec des *points***
- **Proposition 2 : Alterner avec des *virgules***

➢ Choix différents selon les transcripteurs

➢ Cohérence avec la production de l'enfant :
un petti chat ne dor plu iltonbe/ de les sicille et il C'est/ fest male est il plere/est sa mér le ra mene

➢ La virgule n'apparait que plus tard (Simon, 1973)

Ponctuation et connecteurs au CP (70 productions)

Ponctuation interpropositionnelle			**217**
Ponctuation			**49**
point	45	point d'exclamation	4
retour à la ligne	16	virgule	0
Connecteurs			**152**
coordination			**145**
et	112	ensuite	1
après	10	alors	1
mais	8	maintenant	1
puis	6	car	1
tout à coup / tout d'un coup	3	où	1
		soudain	1
subordination			**7**
parce que	4	pendant que	3
Aucune marque			**69**
Ponctuation intrapropositionnelle			**11**
Connecteurs			**10**
Énumération			10
et	10		
Aucune marque			**12**

Le cas de la ponctuation

- **Proposition 1 : Alterner avec des *points***
- **Proposition 2 : Alterner avec des *virgules***
- **Proposition 3 : Balises marquant la ponctuation**

Un petit chat ne dort plus **<ponctuation/>** il tombe de l'escalier […]

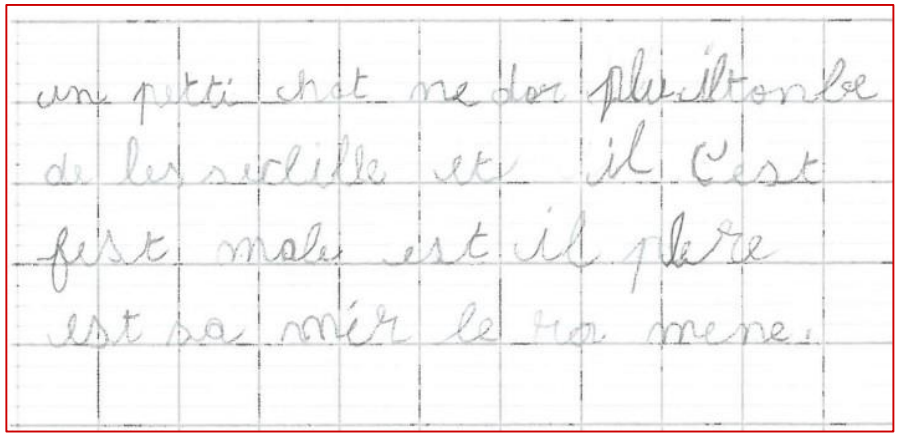
- Choix indépendant du transcripteur

- Comment gère-t-on les ponctuations non obligatoires ?
- Faut-il normaliser les connecteurs « et » ?
- Combien de balises différentes ?

Le cas de la ponctuation

- Combien de balises différentes ?
  - Interpropositionnelles
    - Avec ellipse du sujet (*sa maman se réveilla d'un coup et descend la marche*)
    - Sans ellipse du sujet (*Il tombe de l'escalier et il s'est fait mal*)
  - Intrapropositionnelles
    - Apposition (*le chat, il marche*)
    - Énumération (*avec* [sa mère,] *ses sœurs et ses frères*)
    - Insertion / Incise (*Le chat va faire une aventure, sa première aventure*)

Le cas de la ponctuation : conclusion



un	petti	chat	ne	dor	plu	iltonbe	de	les	sicille	et	[…]

un	petit	chat	ne	dort	plus	<ponctuation/>	il	tombe	de	l'	escalier	?	[…]

(élève 93 - production de CP)

III – Perspectives

Perspectives

- Poursuite du travail d'analyse
  - Réflexion sur la « normalisation »
  - Élaboration d'outils de traitement automatiques

Perspectives

- Poursuite du travail d'analyse
  - Réflexion sur la « normalisation »
  - Élaboration d'outils de traitement automatiques

- Enrichissement du site en ligne
  - Ajout des autres années de recueil
  - Requêtes et fonctionnalités

lien

Recherche par :

Forme | Afficher tous les élèves

2525

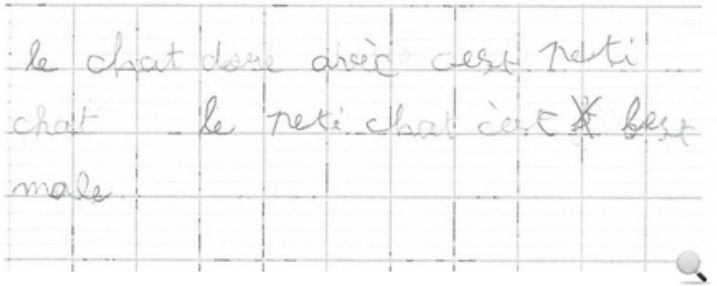
ok

Résultat de votre recherche :

Il y a actuellement à l'étude : 1169 élèves.

Elève 2525

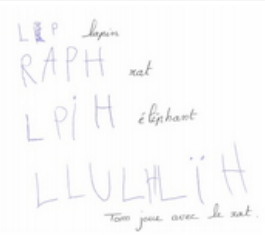
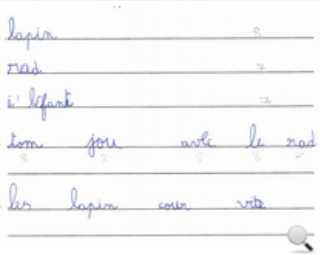
SCAN DICTÉE SEPTEMBRE	DICTÉE SEPTEMBRE	DICTÉE JUIN	SCAN DICTÉE JUIN
	A LA È TUA	lapun rae èllèfen tome gou avèc le rae lest lapun qour vite	

PRODUCTION	SCAN PRODUCTION
le chat dore avèc cest peti chat le peti chat cest fest male	

586 élèves correspondent à votre recherche (*chat * dans la Production) :0

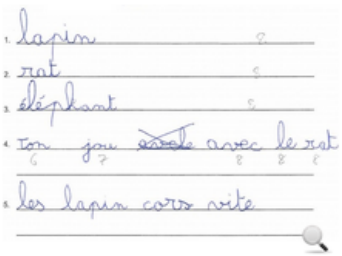
Il y a actuellement à l'étude : 1169 élèves.

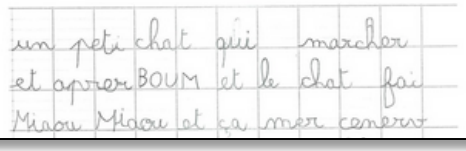
Elève 47

SCAN DICTÉE SEPTEMBRE	DICTÉE SEPTEMBRE	DICTÉE JUIN	SCAN DICTÉE JUIN
	LP RAPH LPIH LLULHLÏH	lapin rad é léfant tom jou avec le rad les lapin cour vite	

PRODUCTION	SCAN PRODUCTION
le chatet tonbe dant les zescalié et il pleur parse que il ses fe tre tre male et sa maman le pretg et il le lech.	

Elève 40

SCAN DICTÉE SEPTEMBRE	DICTÉE SEPTEMBRE	DICTÉE JUIN	SCAN DICTÉE JUIN
	W LAP U RA éLFON TomRA	lapin rat éléphant Tom jou avec le rat les lapin cors vite	

PRODUCTION	SCAN PRODUCTION
un peti chatqui marcher et aprer BOUM et le chatfai Miaou Miaou et ça mer ceherv et ça mer la sover et ail rantr chés elle avec ce peti chat.	

Perspectives

- Poursuite du travail d'analyse
  - Réflexion sur la « normalisation »
  - Élaboration d'outils de traitement automatiques

- Enrichissement du site en ligne
  - Ajout des autres années de recueil
  - Requêtes et fonctionnalités

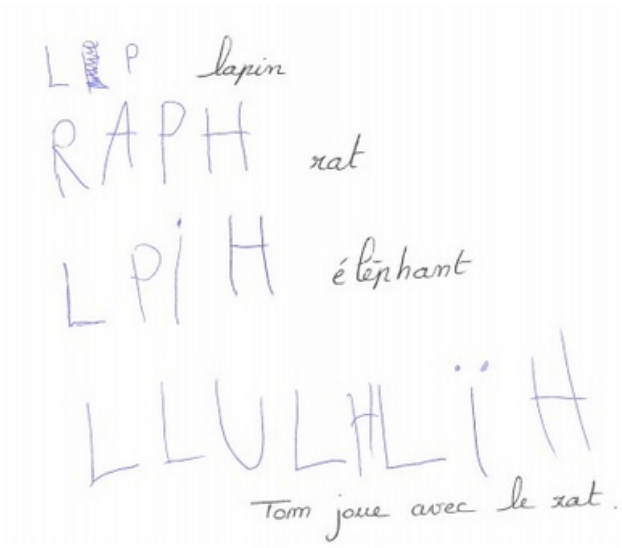
Perspectives

- Poursuite du travail d'analyse
  - Réflexion sur la « normalisation »
  - Élaboration d'outils de traitement automatiques
- Enrichissement du site en ligne
  - Ajout des autres années de recueil
  - Requêtes et fonctionnalités
- Dialogue avec les enseignants et les utilisateurs
  - Amélioration des données
  - Pertinence des outils de traitement

Laissez un commentaire sur cette application
N'hésitez pas à nous laisser un commentaire sur l'application pour nous aider à l'améliorer.
Laissez votre commentaire...
Enregistrer
Annuler
Aucun commentaire

lien

Transcription de la dictée de septembre pour l'élève 47



MARKHA
LP
RAPH
LPIH
LLULHLÏH

Commentaires sur la transcription

Enregistrer

Annuler

lien

Bibliographie

- Cappeau, P. & Roubaud M.-N. (2005). Enseigner les outils de la langue avec les productions d'élèves, Paris : Bordas.
- Catach, N. (1994). La ponctuation:(histoire et système) (Vol. 2818). Presses Universitaires de France-PUF.
- David, J., & Doquet, C. (2016). Les écrits d'élèves: un corpus de référence pour le français contemporain. In SHS Web of Conferences (Vol. 27, p. 11001). EDP Sciences.David, J., & Vaudrey-Luigy, S. (2014). Enseigner la ponctuation. Le français aujourd'hui, 187.
- Elalouf, M. L. (2005). Ecrire entre 10 et 14 ans: un corpus, des analyses, des repères pour la formation. SCEREN-CRDP de l'Académie de Versailles.
- Elalouf, M. L. (2011). Constitution de corpus scolaires et universitaires: vers un changement d'échelle?. Pratiques. Linguistique, littérature, didactique, (149-150), 56-70.
- Garcia-Debanc, C., & Bonnemaison, K. (2014). La gestion de la cohésion textuelle par des élèves de 11-12 ans: réussites et difficultés. In SHS Web of Conferences (Vol. 8, pp. 961-976). EDP Sciences.
- Gunnarsson-Largy, C., & Auriac-Slusarczyk, E. (2013). Présentation du corpus Grenouille. Ecriture et réécriture chez les élèves. Un seul corpus, divers genres discursifs et méthodologies d'analyse, (10), 7-14.
- Kraif, O., & Ponton, C. (2007). Du bruit, du silence et des ambigüités: que faire du TAL pour l'apprentissage des langues, Actes de TALN, Toulouse. url: http://www. u-grenoble3. fr/ponton/perso/docs/TALN07. pdf.
- Corpus lancaster : http://www.lancaster.ac.uk/fass/projects/lever/index.htm
- Mauroux, F., & Garcia-Debanc, C. (2013). Analyse d'épreuves pour évaluer les compétences de scripteur des élèves à l'entrée de l'école élémentaire. Repères. Recherches en didactique du français langue maternelle, (47), 149-170.
- Popin, J., & Thomasset, C. A. (1998). La ponctuation. Nathan.
- Simon, J. (1973). La langue écrite de l'enfant. Presses universitaires de France.
- Smith, N., & McEnery, T. (1998). Issues in Transcribing a Corpus of Children's Hanwritten Projects, Literary and linguistic computing, 13(4): 217-225.

Merci de votre attention

LIDILEM

UNIVERSITÉ Grenoble Alpes

Mise à disposition d'un corpus d'apprenants : quelle méthodologie pour quels enjeux ?

Claire Wolfarth, Catherine Brissaud, Claude Ponton

Lidilem

Université Grenoble Alpes