



HAL
open science

“ Le chat et tonbe dant les zescalié ”, Faut-il s’inquiéter ?

Claire Wolfarth, Louise-Amélie Cougnon, Catherine Brissaud, Claude Ponton

► To cite this version:

Claire Wolfarth, Louise-Amélie Cougnon, Catherine Brissaud, Claude Ponton. “ Le chat et tonbe dant les zescalié ”, Faut-il s’inquiéter?. Posters Day 2018 – ILC, Mar 2018, Louvain-La-Neuve, France. hal-01940682

HAL Id: hal-01940682

<https://hal.univ-grenoble-alpes.fr/hal-01940682>

Submitted on 30 Nov 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 chat et tonbe dant les zescalie » FAUT-IL S'INQUIÉTER ?

PROJET SCOLEDIT

Objectifs

- **Corpus** : constituer et diffuser un vaste corpus longitudinal d'écrits d'apprenants de 7 à 11 ans (école primaire)
- **TAL** : concevoir et élaborer, à l'aide du traitement automatique des langues, des méthodes et un outillage servant d'aide à la constitution et à l'exploitation de corpus scolaires
- **Description linguistique** : décrire finement les performances réelles des élèves dans le domaine de la production d'écrit et analyser les traces visibles de l'évolution diachronique au cours de l'apprentissage de l'écriture

01 Corpus

Le projet SCOLEDIT vise la constitution d'un corpus numérique longitudinal de textes scolaires produits par des élèves de 7 à 11 ans et rencontrés à plusieurs reprises lors de leur scolarité élémentaire. La consigne donnée est la même pour l'ensemble des élèves d'une classe d'âge.

1^{ère} année

*un petit chat me des plouillonke
 di les sorcille est il cest
 fire mais est il phre
 est sa miter le tra mene.*

2^e année

*Il s'agit de ce fait un petit chat qui
 s'appelle mimi il arrive dans un
 maison avec son fratrie un jour il
 rentre au magasin puis il ne savait
 pas comment faire son son fratrie mais
 il se se fit par accident des @poules
 avec chez lui en toute innocence alors
 devant le jardin mais le chien l'arrêt
 n'arriver pas de l'entree, moi des
 question fait le lui, le fait rentrer au
 marcher, ou demander a la @poule?
 et le vrai comment fait pour de ses saas!*

*dit le chat a son fratrie. Pour qui ne
 pas le feu faire qu'on lui demande
 son fratrie. Et les a lui faire pas
 est mimi tu va qui demander au
 grand maître de chat qu'on dit tu
 Non j' ne va pas, je ne va pas
 je kope pas. Ah! ah! ah! ah! ah! ah!
 qui kope les ses un chat qui mimi
 @poules n'arrive aller tu dit que tu n'arrive
 pour lui dire que non dit le chat l'on
 mimi tu comprend l'histoire mimi
 mimi alors tu n'arrive pas*

3^e année

*Il était une fois, dans une maison un petit
 chat qui aime jouer. Il voulait voir à
 qu'il ressemble la maison des voisins.
 Il se demande si les crocotte son
 malleur que chez lui. Il dit:
 Salut les amis!!! cri le chat
 Salut tu fait quoi? dit l'un
 ben je va voir si le crocotte son
 malleur chez les voisins! dit le chat
 alors, et Et il continue sa route.*

*Il arrivait devant la maison des voisins.
 Malle paillet par la fenêtre il entrat.
 Mais ou sont elle ils passai? Selon son
 le chat j' est une devant lui sont venir
 des crocotte. En revenant à la maison,
 il avait mangé l'élément de crocotte
 qu'on l'appella crocotte.*

4^e année

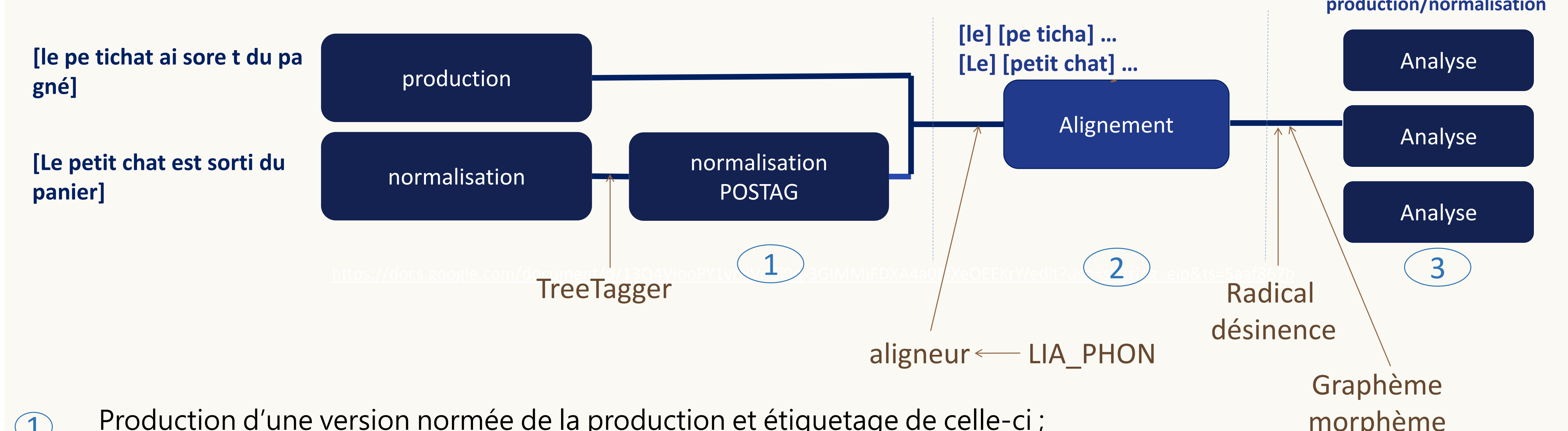
*le dit
 Il était une fois un chatte qui avait une
 ses deux chatte son son malleur chez
 lui. Elle ne faisait pas de malleur
 quelle crocotte elle dit la même chose
 elle arrive dans une maison et
 la elle traverse mimi et elle très content
 de les voir elle demande à mimi
 pourquoi est il parti elle ne va pas
 par mimi lui dit qu'il était parti à
 la recherche de son fratrie et il
 rentre à la maison.*

5^e année

Recueil en mai-juin 2018

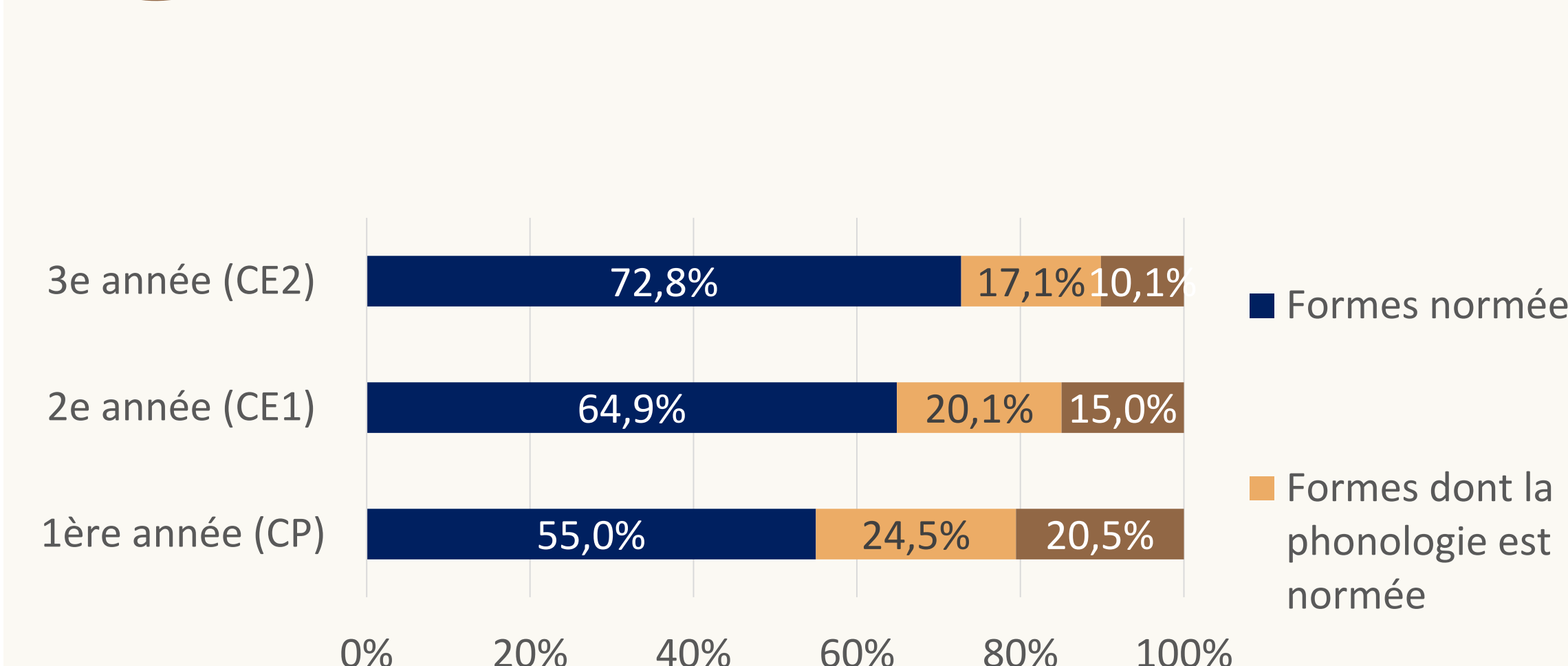
02 Traitement du corpus

Méthodologie : Aligner forme produite et forme normée



- 1 Production d'une version normée de la production et étiquetage de celle-ci;
- 2 Alignement forme à forme de la transcription et de la normalisation de la production :
 - Usage d'indices phonologiques et archiphonologiques ;
 - Calcul des distances d'édition (Levenshtein, 1966) ;
 - Post-traitements des cas de sur-segmentation et sous-segmentation ;
- 3 Traitements postérieurs (alignement graphème par graphème par exemple) et extraction de données linguistiques.

03 Statistiques générales



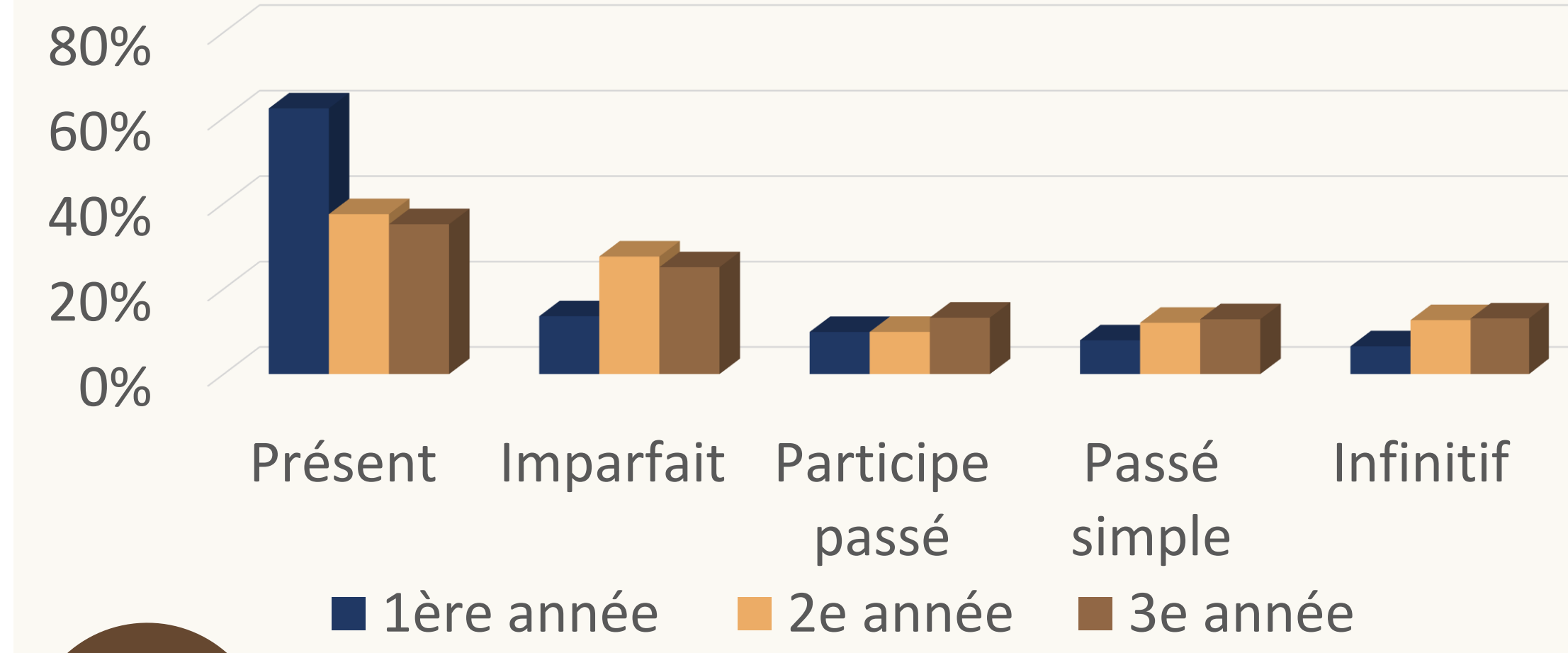
Variations et réussites orthographiques (974 productions)

2 ^e année		3 ^e année		4 ^e année	
Rang	Forme	Rang	Forme	Rang	Forme
8	chat	8	chat	7	chat
9	sorcière	9	sorcière	10	sorcière
11	loup	14	loup	16	loup
18	robot	21	robot	27	dit
25	fois	25	dit	39	robot
31	dit	37	jour	47	fois
35	jour	38	fois	49	jour
37	forêt	47	forêt	57	maison
46	maison	53	maison	61	fin
54	petit	58	petit	71	coup

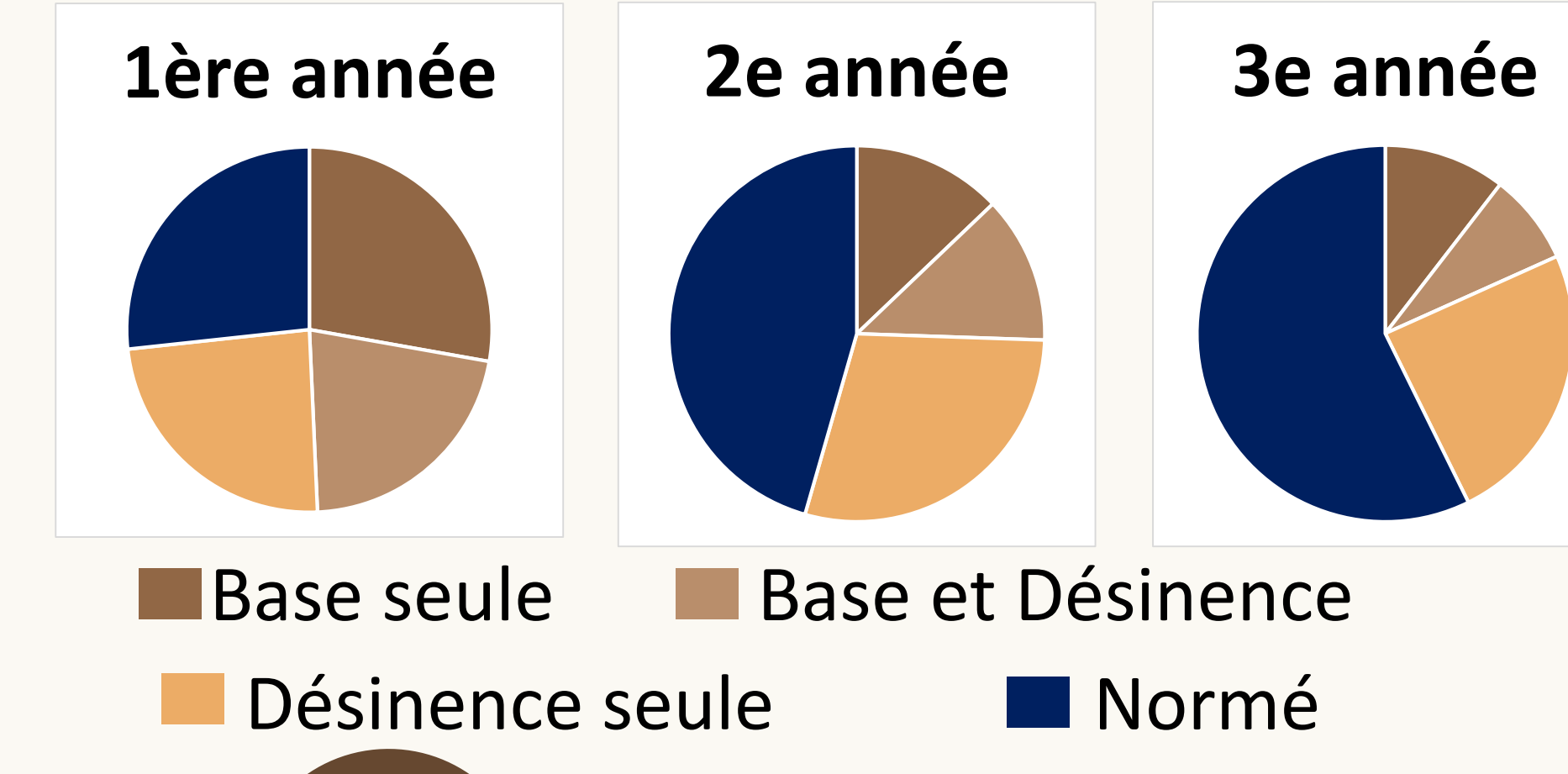
Unités lexicales les plus fréquentes (1104 productions)

04 Étude des verbes

Répartition des temps et modes verbaux selon le niveau de scolarité (903 productions)



Répartition des erreurs sur les bases et les désinences des formes verbales (903 productions)



Conclusion

Ce projet applicatif répond à un besoin de société, tant pour les chercheurs (description linguistique) que pour les enseignants (outil pédagogique).
 Après 5 ans, la collecte est désormais presque aboutie, tandis que la phase d'analyse des données n'en est qu'à ses débuts. De nombreuses analyses linguistiques sont encore possibles et des traitements plus qualitatifs permettront de compléter les résultats obtenus grâce au traitement automatique des langues.

Perspectives

Recherche en cours : analyse de l'évolution des compétences des élèves sur base de digrammes et trigrammes (cf. prospectus de présentation)

Pour découvrir le corpus dans son intégralité :
<http://otus.u-grenoble3.fr/scoledit/>

	CP	CE1	CE2	CM1	CM2	Total
Nombre de productions	977	876	1190	1102	1100 (prévision)	5244
Nombre de mots	16 000	67 000	153 500	180 000	/	416 300
Longueur moyenne (en mots par production)	24,7	76,4	129,5	163,1	/	102,5