

Annotation de résumés oraux d'élèves de primaire pour l'analyse automatique des capacités de compréhension de la lecture



Auteurs : Étienne Labbé² Brice Brossette³ Nathalie Camelin⁴ Tiphaine Caudrelier⁴ Eddy Cavalli⁴ Isabelle Ferrané²
Barbara Lutz¹ Véronique Moriceau² Thomas Pellegrini² Julien Pinquier² Cantin Prat⁴ Lucile Gelin^{1, 2}



(1) Lalilo by Renaissance Learning, France
(3) Laboratoire d'Etude des Mécanismes Cognitifs (EMC), Université Lumière Lyon 2, Lyon, France

(2) IRIT, Université Paul Sabatier, CNRS, Toulouse, France
(4) Laboratoire d'Informatique d'Avignon (LIA), Avignon, France

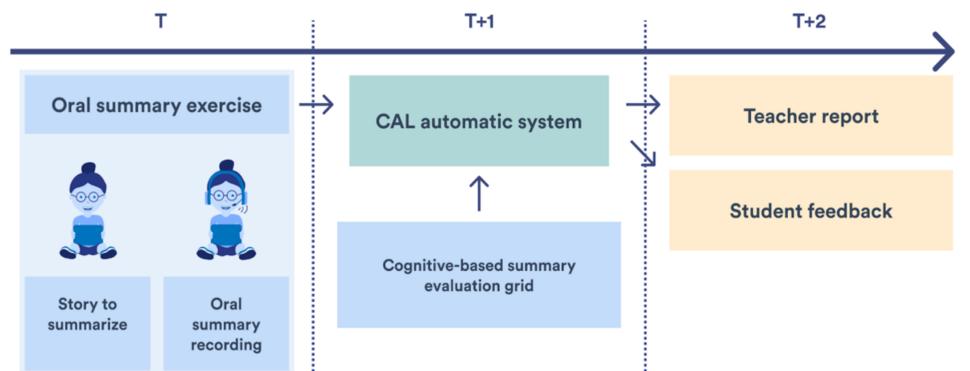


Contexte : le projet CHICA-AI (2024-2028)

• **Objectif** : Construire une activité assistée par ordinateur pour l'entraînement des compétences de compréhension de la lecture des élèves de primaire

• **Défis** :

- Adaptation de technologies
 - Parole d'enfant
 - Parole spontanée
 - Utilisation en classe
- Développer une expérience psychocognitive et pédagogique
 - Évaluation d'un résumé : Qu'est-ce qu'un "bon résumé" ?
 - Remédiation : Quels retours aux élèves ?



→ **Besoin de données annotées !**

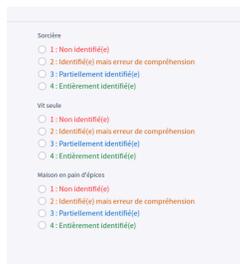
Grille d'évaluation

- Création de 5 textes à résumer pour des CM1-CM2, conçus par des experts pédagogiques
- Collecte des résumés audios à travers la plateforme Lalilo
- Attente pédagogiques des enseignant(es)
 - Grille d'évaluation des résumés [1, 2]
 - Liste de critères prédéfinis pour chaque histoire

- Critères par catégorie :
 - Compréhension : personnages, caractéristiques, lieux, idées et chronologie
 - Mise en forme : connecteurs logiques, temporels...
 - Expression orale : intonation, vitesse de parole...
- Score global dans chaque catégorie

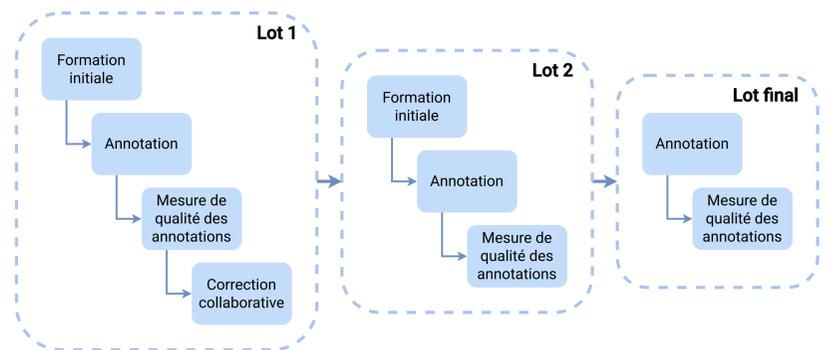
Tâche d'annotation

- Écarter les audios non annotables
 - Trop bruyé, hors exercice...
- Corriger la transcription automatique
 - Ce qui est prononcé
- Évaluer les critères de compréhension
 - Score chaque critère
 - Surlignage des mots concernés
- Évaluer la mise en forme
- Évaluer l'expression orale



Capture d'écran de l'interface (scores de quelques critères de compréhension)

Protocole d'annotation



Résultats

- Lot final : 2085 annotations, dont 1020 annotables et 875 fichiers audio
- Métriques d'accord :
 - **ICC3** : Coefficient de Corrélation Intraclasse [3] (acceptable si >0.6)
 - **Levenshtein** : Ratio des insertions, délétions et substitutions selon la taille des séquences (acceptable si >0.8)
- Observations :
 - Transcription valide malgré certains cas ambigus
 - Scores de compréhension se sont améliorés après le lot 1
 - Score de mise en forme et expression plutôt bas → reste difficile à évaluer
 - Accords suffisant la suite du projet, car l'annotation est complexe

Champ d'annotation	Métrique	Lot 1	Lot 2	Lot final
Transcription	Levenshtein	0,9102	0,8559	0,8781
Compréhension globale	ICC3	0,6510	0,9089	0,8141
Personnages principaux	ICC3	0,6775	0,9282	0,9306
Idées principales	ICC3	0,7303	0,8092	0,8859
Lieux	ICC3	0,9666	0,8922	0,8836
Personnages secondaires	ICC3	0,7816	0,9666	0,8633
Chronologie des idées	Levenshtein	0,7776	0,8135	0,8551
Mise en forme globale	ICC3	0,3173	0,7517	0,5092
Expression globale	ICC3	0,6945	0,6708	0,5594

À venir

- Étude approfondie des données :
 - Quels critères sont pris en compte dans les scores annotés ?
 - Peut-on retrouver toutes les annotations automatiquement ?
- Adaptation de systèmes RAP et CLP
- Comment évaluer l'efficacité sur le niveau des élèves ?

Références

- [1] CASAZZA M. E. (1993). Using a model of direct instruction to teach summary writing in a college reading class. *Journal of Reading*, 37(3), 202–208.
- [2] CÈBE S., GOIGOUX R. & THOMAZET S. (2004). Enseigner la compréhension. Principes didactiques, exemples de tâches et d'activités. In *Lire écrire, un plaisir retrouvé*. MEN-DESCO.
- [3] KOCH G. (2006). Intra-class Correlation Coefficient.