



HAL
open science

Recueil non supervisé de données audio massives à l'école maternelle : de l'enregistrement au développement d'un logiciel d'enrichissement et d'analyse

Aurélie Nardy, Isabelle Rousset, Rachel Gaubil

► **To cite this version:**

Aurélie Nardy, Isabelle Rousset, Rachel Gaubil. Recueil non supervisé de données audio massives à l'école maternelle : de l'enregistrement au développement d'un logiciel d'enrichissement et d'analyse. 2025. <hal-05105934v1>

HAL Id: hal-05105934

<https://hal.science/hal-05105934v1>

Preprint submitted on 10 Jun 2025 (v1), last revised 12 Dec 2025 (v2)

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.



HAL Authorization

Recueil non supervisé de données audio massives à l'école maternelle :

de l'enregistrement au développement d'un logiciel d'enrichissement et d'analyse

NARDY, Aurélie (1), MC, <https://orcid.org/0000-0001-6593-5307>. Mail : aurelie.nardy@univ-grenoble-alpes.fr.

(1) Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France

ROUSSET, Isabelle (1), IR, <https://orcid.org/0000-0003-1020-4819>. Mail : isabelle.rousset@univ-grenoble-alpes.fr.

(1) Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France

GAUBIL, Rachel (1,2), IE, <https://cv.hal.science/rachel-gaubil>. Mail : rachel.gaubil@univ-grenoble-alpes.fr.

(1) Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France

(2) Univ. Grenoble Alpes, CNRS, Litt&Arts, 38000 Grenoble, France

- Domaine scientifique principal

Développement du langage

- Domaines et disciplines scientifiques autres ou plus spécifiques

programmation informatique, logiciel

- Mots clés en français

compétences langagières, enfant, maternelle, adulte

- Mots clés en anglais

linguistic skills, child, preschool, adult

- Résumé en français

Afin d'examiner les relations entre socialisation enfantine et développement du langage oral en maternelle, l'équipe multidisciplinaire réunissant des linguistes, des éthologistes et des chercheurs en science des réseaux du projet ANR DyLNet a mis en place un recueil de données massives dans une école maternelle. Pendant 2 ans et demi, une semaine par mois, grâce à un dispositif portatif, ils ont enregistré les proximités entre individus (enfants et adultes de l'équipe pédagogique) ainsi que leurs paroles. Environ 35 000 heures d'enregistrements audio ont été collectées dont un peu plus de 800 ont été annotées et transcrites par des linguistes.

Dans cet article, nous abordons les enjeux liés à l'enregistrement et au traitement des données audio massives. Dans un premier temps, nous exposons les démarches et précautions éthiques mises en place puis notre méthodologie de recueil ainsi que les pré et post-traitements automatiques des enregistrements que nous avons effectués. Nous rendons compte ensuite des procédures d'annotation et de transcription des enregistrements et présentons le logiciel d'enrichissement, d'extraction et de comptage de formes langagières

que nous avons développé. Enfin, nous décrivons le corpus de 705 transcriptions issues des enregistrements de 96 enfants et 11 adultes que nous mettons à disposition de la communauté scientifique sur Ortolang. Ce corpus viendra étayer les recherches sur le développement langagier chez l'enfant entre 3 et 6 ans et permettra une description de leur environnement langagier à l'école maternelle.

- Résumé en anglais

To examine the relationships between early childhood socialization and the development of oral language in preschool, the multidisciplinary team of the ANR DyLNet project – composed of linguists, ethologists, and network science researchers – implemented a large-scale data collection in a preschool setting. Over a period of two and a half years, one week per month, they used a portable device to record both the proximity between individuals (children and educational staff) and their speech. Approximately 35,000 hours of audio recordings were collected, of which just over 800 hours were annotated and transcribed by linguists.

In this article, we address the challenges related to the recording and processing of massive audio data. First, we outline the ethical considerations and precautions that were taken, followed by our data collection methodology and the automatic pre- and post-processing procedures applied to the recordings. We then detail the annotation and transcription procedures and introduce the software we developed for enriching, extracting, and counting linguistic forms. Finally, we describe the corpus of 705 transcriptions from recordings involving 96 children and 11 adults, which we are making available to the scientific community via Ortolang. This corpus will support research on language development in children aged 3 to 6 and will allow for a detailed description of their linguistic environment in preschool.

Informations courtes sur le corpus	
Titre	Corpus DyLNet.
Auteur-e-s	NARDY, Aurélie – Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France (https://orcid.org/0000-0001-6593-5307), responsable scientifique, auteure. ROUSSET, Isabelle – Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France (https://orcid.org/0000-0003-1020-4819), chercheure, auteure. GAUBIL, Rachel – Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France ; Univ. Grenoble Alpes, CNRS, Litt&Arts, 38000 Grenoble, France (https://cv.hal.science/rachel-gaubil), développeuse, auteure. BOUCHET, Hélène – Univ. Grenoble Alpes, LIDILEM, F-38000 Grenoble, France (https://orcid.org/0000-0002-0317-4909), chercheure, auteure.
Date de publication	2025-05-13
Version	Version 2
Lien	https://hdl.handle.net/11403/corpus-dylnet
Identifiant unique	https://hdl.handle.net/11403/corpus-dylnet
Disponibilité	Dépôt avec accès sur demande motivée à la responsable du projet.
Licence	CC BY-NC-ND 3.0
Formats et	Fichiers textes et images pour un total de 763,2 Mo : <ul style="list-style-type: none"> • 1245 fichiers .eaf

métriques	<ul style="list-style-type: none"> • 5 fichiers .eag • 3 fichiers images (.jpg, .eps, .pdf) • 1 fichier .csv • 3 fichiers .pdf <p>Organisation en collections d'items :</p> <ul style="list-style-type: none"> • Corpus_complet <ul style="list-style-type: none"> • adulte <ul style="list-style-type: none"> • <i>fichiers.eaf</i> • enfant <ul style="list-style-type: none"> • <i>fichiers.eaf</i> • Corpus_longitudinal <ul style="list-style-type: none"> • Cohorte_MS-GS <ul style="list-style-type: none"> • <i>fichiers.eaf</i> • Cohorte_PS-MS <ul style="list-style-type: none"> • <i>fichiers.eaf</i> • Images <ul style="list-style-type: none"> • <i>logo_dylnet.eps</i> • <i>logo_dylnet.jpg</i> • <i>logo_dylnet.pdf</i> • Informations_sociodemographiques <ul style="list-style-type: none"> • <i>dico_variabies_sociodemo_ortolang.pdf</i> • <i>sociodemo_ortolang.csv</i> • Procedures_annotations_transcription <ul style="list-style-type: none"> • Requetes_ELAN <ul style="list-style-type: none"> • <i>fichiers .eag</i> • <i>MANUEL_DyLNet_ortolang.pdf</i> • <i>Description_Corpus_DyLNet_ortolang .pdf</i> <p>Convention de nommage des fichiers .eaf : les informations suivantes sont jointes avec le symbole '-' :</p> <ul style="list-style-type: none"> • identifiant du locuteur (ex. '0039') • période de recueil des enregistrements (ex. 'P1') • numéro de fichier transcrit (ex. '01') <p>Nom du fichier d'exemple : <i>0039-P1-01.eaf</i></p>
Langue	français
Responsable	Nardy, Aurélie, MC, Lidilem (Univ. Grenoble Alpes). aurelie.nardy@univ-grenoble-alpes.fr
Publications liées	https://hal.science/search/index/?q=dylnet&rows=30
Référence bibliographique	Nardy, A., Rousset, I., Gaubil, R., Bouchet, H. (2025). <i>Corpus DyLNet</i> [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). https://hdl.handle.net/11403/corpus-dylnet .

1. INTRODUCTION

1.1. CONTEXTE SCIENTIFIQUE

Comme le pointent depuis plusieurs années les rapports de l'OCDE, la France est le pays où l'origine sociale pèse le plus sur les inégalités scolaires et ce, dès le début de la scolarisation. Parce que l'école maternelle est la première étape de la scolarisation, il est

nécessaire de comprendre comment les enfants de tous les groupes sociaux s'y intègrent et s'y adaptent. Le langage oral joue un rôle central dans ce contact précoce avec le monde scolaire car il y revêt un double rôle : il est un outil au service de la socialisation (l'enfant se socialise par le langage) et un résultat de la socialisation (l'enfant acquiert le langage en se socialisant). Les enfants s'intègrent et s'adaptent à l'école en communiquant. Réciproquement, une socialisation scolaire réussie multiplie les opportunités de communication avec les pairs et les adultes en charge des enfants, favorise les apprentissages et renforce les compétences linguistiques. Il peut donc s'établir un cercle vertueux – ou au contraire une spirale d'échec – entre sociabilité enfantine, communication orale, apprentissages scolaires. Afin de cerner les relations entre socialisation enfantine et développement du langage oral en maternelle, un recueil de données sans précédent a été mené dans le cadre du projet ANR DyLNet (Language Dynamics, Linguistic Learning, and Sociability at Preschool: Benefits of Wireless Proximity Sensors in Collecting Big Data, ANR-16-CE28-0013, 2016-2021). Ce projet, porté par le Lidilem (Univ. Grenoble Alpes, UGA), et mené en partenariat avec l'équipe DANTE de l'Inria (ENS de Lyon), a réuni une équipe multidisciplinaire composée de sociolinguistes, psycholinguistes, éthologistes et chercheurs en sciences des réseaux.

1.2. RECUEIL DE DONNEES

La mise en œuvre de ce projet a consisté à suivre, pendant 2 ans et demi, les enfants et intervenants pédagogiques d'une école maternelle socialement mixte d'une zone urbaine du département de l'Isère (≈ 200 individus par an). Chaque mois, pendant une semaine, ils ont porté des petits boîtiers contenant deux types de dispositifs de collecte : (1) des capteurs RFID¹ enregistrant les proximités entre individus (via des mesures de RSSI² prises toutes les 5 secondes) ; (2) des micros enregistrant leurs paroles. À différents moments du suivi longitudinal, des tests de langage ont été administrés aux enfants et grâce à un questionnaire adressé aux familles, nous avons collecté pour chacun d'eux des informations sociodémographiques. Enfin, lors de chaque semaine de collecte des données, une chercheuse de l'équipe était présente dans l'enceinte scolaire afin de mener des observations ethnographiques sur les routines scolaires.

Au final, nous avons enregistré 655 958 381 signaux RSSI correspondant à 1 705 126 interactions sociales et disposons de $\approx 35\,000$ heures d'enregistrements audio dont un peu plus de 800 heures ont été annotées et transcrites par des linguistes. Une partie des données (10 mois de données sur les interactions sociales, des résultats aux tests de langage ainsi que des informations sociodémographiques sur les participants) a été mise à disposition de la communauté scientifique internationale via une publication dans la revue *Scientific Data* (Dai et al., 2022). Les transcriptions des enregistrements audio ont quant à elles été mises à disposition de la communauté scientifique à l'échelle nationale sur la plateforme Ortolang (<https://hdl.handle.net/11403/corpus-dylnet>).

Dans cet article, nous aborderons plus spécifiquement les enjeux liés à l'enregistrement et au traitement des données audio massives. Après avoir présenté les démarches et précautions éthiques mises en œuvre, nous exposerons notre méthodologie de recueil et de traitements des enregistrements audio puis nous expliciterons les procédures d'annotation et de transcription adoptées et présenterons le logiciel d'enrichissement, d'extraction et de comptage de formes langagières que nous avons développé. Enfin, nous décrirons le corpus de transcriptions mis à disposition sur Ortolang et illustrerons son intérêt pour la communauté scientifique et son potentiel de réutilisation.

¹ RFID : Radio Frequency Identification.

² RSSI : Received Signal Strength Indicator.

2. METHODE

2.1. ÉTHIQUE, DONNEES PERSONNELLES ET QUALITE

La mise en place d'un tel projet a nécessité d'être très attentifs à l'éthique de la recherche. Dans ce cadre, différentes mesures ont été prises de façon à protéger et à respecter le bien-être et les droits des personnes.

2.1.1. Éthique de la recherche et confidentialité

Une convention de partenariat entre l'Université et l'Éducation Nationale a été signée par la présidence de l'Université Grenoble Alpes, la Direction Académique des Services de l'Éducation Nationale de l'Isère (par délégation du recteur de l'Académie de Grenoble), l'inspection de circonscription, la direction de l'école maternelle et la chercheuse responsable du projet DyLNet.

Le protocole de l'étude a été validé par le Comité Opérationnel d'Évaluation des Risques Légaux et Éthiques – COERLE – de l'INRIA (avis favorable n° 2017-014) et a fait l'objet d'une déclaration auprès de la Commission Nationale de l'Informatique et des Libertés – CNIL – (avis favorable n° 2017_0980683). En outre, une autre déclaration auprès de la CNIL a été effectuée pour la constitution d'une base de données de recherche (avis favorable n° 2019_0980683).

Les participants ont été amenés à signer 2 formulaires de consentement : un formulaire de consentement de participation à l'étude et un formulaire de consentement concernant la conservation des données après l'étude et leur accessibilité pour d'autres études.

Les données récoltées au cours du projet DyLNet ont été traitées avec la plus grande confidentialité :

- Toutes les personnes impliquées dans le recueil et le traitement des données ont signé une clause de confidentialité et ont reçu des consignes déontologiques strictes. Ces personnes sont soumises au secret professionnel (selon les conditions définies par les articles 226-13 et 226-14 du code pénal).
- Toutes les données ont été traitées de manière anonyme : les participants ont été identifiés par des numéros d'anonymat.
- L'ensemble des données (40 To) a dans un premier temps été stocké sur un serveur sécurisé de l'INRIA puis a été transféré sur un serveur sécurisé de l'UAR GRICAD Grenoble Alpes Recherche.
- S'agissant des transcriptions effectuées à partir des enregistrements : les consignes de transcription protègent la vie privée des participants et de leur famille (tout élément contenant un nom de lieu ou de personne, une adresse, des informations susceptibles de révéler l'identité des enfants et des familles, ou d'autres informations privées n'ont pas été transcrites).

2.1.2. Qualité et conformité du matériel utilisé

Les boîtiers utilisés pour enregistrer les proximités entre individus et les paroles échangées au sein de l'école ont été développés spécialement pour le projet DyLNet par l'entreprise SEQUANTA. Ils ne sont pas susceptibles de blesser les enfants qui les portent et ne contiennent aucun élément détachable qui puisse être avalé.

Ils émettent de faibles signaux pour communiquer sans fil à une fréquence de 2,44 GHz, proche de celles utilisées par les systèmes Bluetooth et Wi-Fi. Chaque boîtier émet pendant 384 microsecondes (environ 0,4 millièmes de seconde) toutes les 5 secondes. Ainsi, le boîtier d'un participant émet au total 6,6 secondes durant une semaine d'école (sur 24h de temps passé à l'école). Sur une année scolaire, un boîtier émet au total 66 secondes. Ces boîtiers sont conformes aux normes en vigueur :

- Respect de la norme AFNOR « EN 71.1 : Sécurité des jouets » sur les propriétés mécaniques et physiques des objets utilisés par les jeunes enfants ;
- Respect de la norme « ETSI EN 300 328 V1.8.1 » relative aux radiofréquences (*Electromagnetic compatibility and Radio Spectrum Matters*) ;
- Conformité aux prescriptions de la norme EN 50566 (visant à démontrer la conformité des champs radiofréquence produits par les dispositifs de communication sans fil tenus à la main ou portés près du corps), suivant les restrictions de base de la recommandation du Conseil Européen 1999/519/CE, selon la norme de mesure EN 62209-2.

Enfin, nous avons fait réaliser une mesure de DAS (Débit d'Absorption Spécifique) par le laboratoire d'essai EMITECH Angers (accrédité par le Comité Français d'Accréditation). Le DAS est l'indicateur d'exposition de référence pour les systèmes de communication sans fil utilisant des fréquences radio. Il est exprimé en Watt par kilogramme (W/kg) et ne doit pas dépasser 2 W/kg (moyenné dans 10 g) selon la réglementation en vigueur en 2017. Les mesures ont été faites dans la situation où l'appareil est situé au contact du corps. Dans les conditions réelles d'utilisation (un signal émis toutes les 5 secondes), le DAS des boîtiers du projet DyLNet est de l'ordre de 0.00017 W/kg.

2.2. ADHESION AU DISPOSITIF

Avant même le dépôt du projet en vue de la demande de financement, nous avons pris contact avec l'Inspectrice de l'Éducation Nationale de la DSDEN de l'Isère afin de lui présenter notre projet et solliciter son soutien pour engager une école socialement mixte dans notre projet. Le financement du projet acquis, elle nous a mis en contact avec une école maternelle répondant aux caractéristiques sollicitées puis nous avons rencontré l'équipe pédagogique afin de leur présenter les objectifs de notre recherche, le dispositif de collecte des données déployé et répondre à leurs questions. Nous avons présenté le projet à chaque réunion de rentrée aux parents et avons également mis en place des permanences à destination des parents en fin de journée au sein de l'école pour répondre à leurs questions. Nous avons aussi mis en place un site web dédié au projet (hébergé par l'UGA et fermé en 2023) avec une page destinée aux familles qui explicitait le projet, rappelait nos créneaux de permanences et informait des semaines de collecte des données. Enfin, nous nous sommes adjoints les services d'un membre du Haut Conseil de la Santé Publique, spécialisée en biophysique dans le domaine des risques électromagnétiques, pour de l'accompagnement et du conseil sur la question des ondes électromagnétiques. Son intervention au sein du projet s'est principalement déroulée en amont du recueil de données et son accompagnement s'est effectué à la fois auprès des chercheurs du projet et des membres de l'entreprise en charge de la conception et de la fabrication des capteurs, notamment au travers d'une formation sur les aspects médicaux et sociétaux liés aux ondes électromagnétiques. Si le besoin s'était manifesté, elle aurait également participé aux réunions à destination des parents afin de répondre à leurs interrogations sur ces questions.

Au final, sur les 2 ans et demi de la phase de collecte de données, le taux de participation à l'étude a été de 81% (283/351) chez les enfants et de 97% (62/64) chez les adultes (enseignants et Atsem³).

2.3. RECUEIL ET TRAITEMENT DE DONNEES AUDIO MASSIVES

2.3.1. Dispositif d'enregistrement

Dans le cadre de cette recherche, nous avons mis en place un dispositif d'enregistrement portatif spécifiquement conçu pour documenter les usages langagiers des participants au cours de l'ensemble de leurs déplacements au sein de l'établissement scolaire. Ce dispositif a été pensé de manière à répondre simultanément aux contraintes de notre terrain d'enquête

³ Agent Territorial Spécialisé des Écoles Maternelles.

et à nos exigences méthodologiques. Les participants, en majorité de jeunes enfants âgés de 2 ans et demi à 6 ans et demi, ont été enregistrés de manière continue une semaine par mois pendant 2 ans et demi, durant leurs activités scolaires quotidiennes (en classe, dans la cour de récréation, lors des séances d'éducation physique, etc.). De fait, le dispositif d'enregistrement devait être facile à porter, sans risque pour les usagers, tout en étant simple à installer. Par ailleurs, il devait permettre un enregistrement d'une durée relativement importante (24 heures par semaine), garantir une qualité acoustique satisfaisante et intégrer un système d'horodatage.

En collaboration avec une entreprise implantée dans la région Auvergne-Rhône-Alpes, nous avons conçu des boîtiers mesurant 58,15 x 50 x 15 mm (hauteur x largeur x profondeur). Ces boîtiers, fixés au niveau du col à l'aide d'une pince bretelle et permettant d'enregistrer les proximités interindividuelles toutes les 5 secondes, intègrent également deux micros sur leur face avant (cf. Figure 1).



Figure 1 – Boîtier d'enregistrement des proximités et de la parole (source : SEQUANTA)

Les signaux audio ont été enregistrés au format .wav, selon un encodage stéréo 16-bit signed PCM à une fréquence d'échantillonnage de 32 kHz, et stockés sur une carte SD d'une capacité de 32 Go. Le système comprend également un module d'horodatage.

2.3.2. Pré et post-traitements automatiques des signaux

Le nombre élevé de participants, couplé à la volumétrie importante des données collectées (environ 35 000 heures d'enregistrement), soulève deux défis majeurs en matière de traitement des fichiers audio. Le premier concerne l'identification de la voix du porteur du micro dans un environnement caractérisé par la présence simultanée de multiples locuteurs et par un bruit de fond significatif. Cette tâche s'avère particulièrement complexe dans le cas des voix des jeunes enfants dont les fréquences fondamentales moyennes des deux sexes ne sont pas différentes (Busby et Plant, 1995; Potamianos et al., 1997; Weinberg et Bennett, 1971). Le second défi réside dans l'optimisation du processus de transcription, fastidieux et chronophage afin de le rendre plus efficient tout en garantissant une qualité suffisante des données pour répondre aux objectifs de l'étude. Pour ce faire, différents dispositifs de pré- et post-traitement automatique ont été appliqués aux enregistrements recueillis.

Une fois les enregistrements extraits des capteurs (soit, pour chaque participant, deux fichiers d'environ trois heures correspondant à l'enregistrement continu de la matinée et de l'après-midi), un premier traitement est effectué pour segmenter les enregistrements selon des tranches horaires. Cette opération, qui préserve l'horodatage d'origine, permet de générer des fichiers d'une durée maximale d'une heure. Par exemple, une matinée peut donner lieu à quatre fichiers distincts : 8h36-9h, 9h-10h, 10h-11h, 11h-11h22. Pour chacun de ces fichiers, plusieurs versions sont générées : la version stéréo d'origine ainsi que des versions post-traitées.

La version originale stéréo est conservée sous deux formats : .wav, utilisé comme support pour la transcription et .flac, format compressé sans perte, destiné au stockage⁴. À partir de

⁴ Une évaluation menée sur un sous-ensemble de nos fichiers audio a mis en évidence une réduction de l'espace de stockage d'environ 65 % comparativement aux mêmes enregistrements au format

la version stéréo originale, un post-traitement automatique des signaux fondé sur un filtrage spatial (beamforming) est appliqué. Il se fonde sur le décalage physique entre les deux micros intégrés à chaque boîtier.

La Figure 2 illustre le fonctionnement général de ce processus de traitement automatique des signaux.

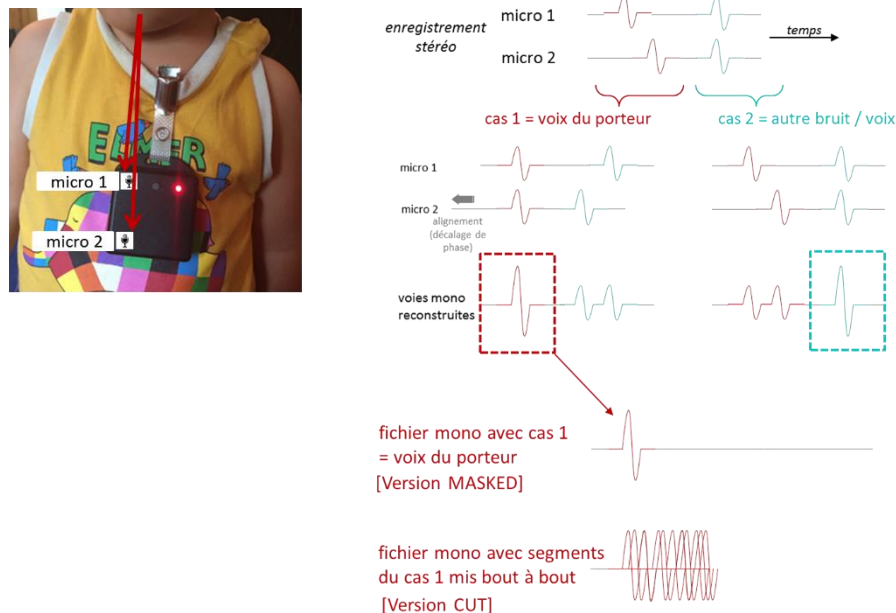


Figure 2 – Illustration du post-traitement automatique des signaux par filtrage spatial (source : Lidilem, Univ. Grenoble Alpes)

Comme l'illustre la photographie présentée sur la Figure 2, lorsque le porteur du boîtier parle, sa voix parvient en premier au micro le plus proche de sa bouche (micro 1) puis, quelques millisecondes plus tard, au second micro (micro 2). Les sons ambiants – qu'il s'agisse des voix d'autres individus ou de bruits environnementaux – atteignent quant à eux généralement les deux micros de manière quasi simultanée ou avec un décalage qui diffère de celui induit par la voix du porteur. Le post-traitement exploite cette propriété physique : il consiste à réaligner les signaux identiques enregistrés par les deux micros en tenant compte du décalage connu, lié à la distance fixe qui les sépare. Une fois ce réajustement effectué, les signaux des deux micros sont sommés afin de générer une voie mono qui permet ensuite de produire 2 versions post-traitées des fichiers audio originaux :

- une version MASKED dans laquelle seuls les segments détectés comme correspondant à la parole du porteur sont conservés. Les autres segments, correspondant aux bruits environnants ou à d'autres voix que celle du porteur, sont remplacés par du silence. Cette version permet de préserver l'intégralité du déroulement temporel de l'enregistrement ;
- une version CUT, constituée uniquement des segments reconnus comme émanant du porteur, qui sont concaténés.

Ce processus de post-traitement automatique, bien qu'efficace dans l'ensemble, présente certaines limites inhérentes à sa dépendance à la position du locuteur par rapport aux micros. Ainsi, certaines prises de parole du porteur peuvent ne pas être détectées, tandis que certains bruits parasites, comme le frottement du boîtier, peuvent, à l'inverse, être

.wav. Un autre avantage du format .flac est de permettre une conversion simple et rapide vers le format .wav.

interprétés à tort comme des segments vocaux émanant du porteur. Malgré ces limites, ce traitement a été particulièrement utile pour deux raisons principales.

Premièrement, il a facilité la sélection des fichiers audio à transcrire. En effet, le rapport entre la durée du fichier CUT (ne contenant que les prises de parole attribuées au porteur) et celle du fichier stéréo original a fourni une estimation approximative du temps de parole effectif du porteur sur l'ensemble de l'enregistrement.

Deuxièmement, ce traitement a soutenu la phase de transcription en aidant à repérer les moments où le porteur s'exprime. Un script exécuté dans le logiciel PRAAT (Boersma et Weenink, 2013) a segmenté automatiquement le signal en intervalles de parole et de silence. Cette opération a permis de générer un fichier au format .TextGrid, contenant les annotations temporelles correspondantes. Ce fichier a ensuite été importé dans le logiciel de transcription alignée ELAN (Wittenburg et al., 2006) en parallèle avec le fichier audio stéréo original. Les segments temporels ainsi affichés guidaient le transcripateur dans l'identification des passages où le porteur du boîtier prenait la parole.

2.3.3. Procédure d'annotation et de transcription en 5 étapes

Conscients de la complexité inhérente à l'activité d'annotation et de transcription, ainsi que de l'exigence cognitive qu'elle impose (Baude et Dugua, 2011; Hriba et al., 2011), nous avons élaboré une procédure en cinq étapes successives répondant à un double objectif : améliorer l'efficacité du travail de transcription tout en garantissant la qualité des données produites. Cette procédure repose sur l'utilisation du logiciel ELAN, qui permet une transcription alignée sur le signal audio et l'annotation simultanée de différents types d'informations linguistiques ou contextuelles. Chaque étape de la procédure est conçue pour guider le transcripateur à travers plusieurs écoutes successives de l'enregistrement, en focalisant son attention sur des aspects spécifiques à chaque passage.

Étape 1 : Ajustement des bornes temporelles placées automatiquement et annotation d'éléments paralinguistiques (1ère écoute)

Lors de la première étape, le transcripateur ouvre simultanément, dans le logiciel ELAN, le fichier audio stéréo original (.wav) et le fichier .TextGrid généré précédemment. Cette configuration lui permet de visualiser, d'une part, les délimitations temporelles automatiquement placées à partir du post-traitement du signal, et d'autre part, la grille d'annotation préétablie mise à sa disposition. Sa tâche, à ce stade, consiste à écouter les segments identifiés comme appartenant au locuteur porteur, à ajuster manuellement les bornes temporelles si nécessaire, et à annoter un certain nombre d'éléments paralinguistiques contextuels présentés dans des menus déroulants.

Trois types d'informations contextuelles ont été annotées dans les transcriptions :

- Activité en cours (lieu dans lequel la production orale a été enregistrée) : en classe, en récréation, hors classe (trajet dans les couloirs, toilettes, bibliothèque, etc.), en sport, indéterminée
- Interlocuteur(s) (interlocuteur(s) du porteur du micro) : un ou plusieurs enfant(s), enseignant de la classe, Atsem de la classe, autre adulte que celui de la classe ou adulte indéterminé, classe entière, parole auto-adressée, indéterminé
- Situation langagière (type de prise de parole) : récitation (chanson, comptine, etc.), lecture à voix haute et transmission de consignes (codées uniquement quand le porteur du micro est un adulte)

Les deux premiers types d'informations ('Activité en cours' et 'Interlocuteur') ont été annotées pour chaque prise de parole du locuteur. La situation langagière a quant à elle été annotée uniquement les fois où cela s'avérait pertinent.

Cette première phase vise à limiter la charge cognitive du transcripateur en focalisant son attention sur un ensemble restreint d'éléments tout en lui permettant de se familiariser

progressivement avec l'enregistrement avant d'aborder la phase de transcription proprement dite.

Étape 2 : Transcription des paroles du porteur (2ème écoute)

Lors de cette seconde écoute, le transcripteur passe dans un mode de visualisation sous ELAN qui lui permet d'écouter et de transcrire les segments délimités au cours de l'étape précédente en suivant les conventions établies pour le projet.

Étape 3 : Vérification et complétion de certains codages (3ème écoute ciblée)

La troisième étape correspond à une écoute ciblée de certains passages transcrits, visant à vérifier ou compléter le codage de plusieurs variables sociolinguistiques du français. Cette phase repose sur l'exploitation de requêtes préprogrammées, exécutables directement depuis ELAN, qui permettent d'isoler automatiquement les occurrences pertinentes pour chacune des variables ciblées. Le transcripteur est ainsi orienté vers les éléments à examiner de manière approfondie. Les variables sont les suivantes :

- la production ou non du 'u' de *tu* devant voyelle ou *h* muet ;
- la production ou non du 'l' dans 'il(s)' et la production complète ou non de 'il(s)';
- la réalisation ou l'effacement du schwa dans les 9 monosyllabiques les plus fréquents : *je, ce, de, le, me, ne, que, se, te* ;
- la production ou non de /@/ et /λ/ en position post-consonantique finale.

Étape 4 : Pré-anonymisation des signaux

Conformément aux engagements éthiques pris auprès des participants et des instances éthiques, tous les éléments susceptibles d'identifier des individus ou des lieux ont été exclus de la transcription. Ces éléments ont été systématiquement remplacés par des chaînes de caractères en majuscules, telles que *PRENOM* ou *LIEU* lorsqu'un prénom ou le nom d'un lieu était mentionné.

À cette étape, le transcripteur exécute une requête dans ELAN qui recherche toutes les suites de caractères en majuscules dans la transcription puis, pour chaque information à anonymiser, il crée un segment sur le signal audio dans la ligne 'Anonymisation' de la grille d'annotation. Cette étape avait été prévue afin de pouvoir ensuite bruitez les enregistrements audio. Les signaux audio n'ont toutefois pas pu être anonymisés car il arrive très souvent que lorsque le porteur parle on entende en même temps un adulte ou un autre enfant produire un prénom.

Étape 5 : Relecture orthographique et formelle de la transcription

La cinquième et dernière étape de la procédure a consisté à exporter la transcription réalisée dans ELAN vers un éditeur de texte afin de procéder à une relecture orthographique et formelle et de corriger les éventuelles coquilles et erreurs orthographiques dans le fichier de transcription.

Enfin, afin de communiquer efficacement avec l'ensemble de l'équipe des transcripteurs et répondre à leurs questions, nous avons utilisé la plateforme de communication collaborative SLACK qui permettait de leur répondre en temps réel et de conserver une mémoire commune des réponses apportées.

Afin que les transcripteurs prennent le temps de s'approprier le manuel de transcription d'une cinquantaine de pages ainsi que les différentes procédures mentionnées lors de chacune des 5 étapes, nous leur avons payé 7 heures de formation correspondant à 4 temps distincts :

- Avant la formation : envoi du manuel de transcription pour lecture et de la version d'ELAN à installer

- 1er temps de formation en présentiel (2h30) : présentation générale du projet, prise en main d'ELAN, présentation de la plateforme SLACK et de son utilisation, signature d'un engagement de confidentialité
- Entre les 2 temps de formation en présentiel : transcription d'un enregistrement de 15 minutes effectué auprès d'un enfant en notant les éventuelles questions qui se sont posées puis envoi à la responsable du projet
- 2ème temps de formation en présentiel (2h30) : retours sur les transcriptions reçues, échanges sur la mise en œuvre des conventions de transcription

2.3.4. Développement d'un logiciel d'enrichissement, d'extraction et de comptage de formes langagières

Dans le but d'exploiter automatiquement le corpus transcrit, nous avons développé le logiciel EAF² (Nardy et al., 2020) qui permet 3 grands types de traitement que nous détaillons dans les sections suivantes.

A/ Normalisation et nettoyage des transcriptions

Nous avons une équipe d'une vingtaine de transcripateurs, principalement étudiants en licence de Sciences du Langage et en deuxième cycle d'école d'orthophonie qui ont dû s'approprier un nombre relativement important de conventions de transcription. Afin de pallier d'éventuelles erreurs de codage et d'orthographe, nous avons développé un premier script que nous avons appelé 'normalisation'. Ce script a été établi à partir de l'analyse « manuelle/humaine » de la liste des mots avec une fréquence d'occurrence supérieure à 3 dans un sous-corpus de $\approx 508\ 000$ mots. Cette analyse a mené à une liste de corrections afin d'uniformiser l'orthographe et les codages et résoudre les éventuelles erreurs d'espace et de ponctuation. Par exemple, pour les mots qui apparaissaient sous plusieurs formes comme *t-shirt* et *tee-shirt*, nous avons choisi une orthographe comme étant celle qui apparaîtrait dans les transcriptions. Puis, sur cette version 'normalisée' de la transcription, nous avons fait tourner un script de 'nettoyage' qui consistait quant à lui à retirer des codages que nous avons fait mettre aux transcripateurs afin d'avoir une version de la transcription la plus proche d'un écrit « standard » dans le but d'effectuer des traitements automatiques ultérieurs. Par exemple, ce script supprimait les parenthèses qui avaient été mises pour signaler l'absence de production d'un phonème comme dans *t(u) as vu* (version transcrite pour indiquer la prononciation *t'as vu*).

B/ Enrichissement des transcriptions

Un autre type de traitement que permet le logiciel est d'enrichir les transcriptions. Le premier script 'tagging' effectue un étiquetage morphosyntaxique via TreeTagger⁵ avec modèle de langue PERCEO adapté pour l'oral (mot|catégorie|lemme). Deux autres scripts viennent enrichir les transcriptions en annotant automatiquement la présence et l'absence de 'ne' dans les contextes de négation. Les règles de détection ont été établies en limitant les contextes de négation à des contextes repérables générant le moins de bruit possible. Le premier recherche tous les contextes de négation impliquant les mots² de négation suivants : 'pas', 'jamais', 'personne', 'rien', 'aucun', 'aucune', 'aucuns', 'aucunes'. Pour chaque contexte détecté, un codage est inséré dans la transcription pour indiquer la présence ([NN1]) ou l'absence ([NN0]) du 'ne' de négation. Le second script quant à lui recherche uniquement les contextes de négation avec le mot² 'pas' et insère les mêmes codages que le précédent. Ce dernier script a fait l'objet d'une évaluation qui révèle un taux d'accord de 97% avec le codeur expert (2105 contextes de négations codés par le script et 2063 codées par l'expert) (Buson et al., 2022). Chacun de ces scripts de codage de contextes de négation génère une nouvelle ligne dans le fichier de transcription.

⁵ <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

Après passage dans le logiciel EAF², notre fichier de transcription initial (celui rendu par le transcritteur qui contient 5 lignes : la ligne de transcription, les 3 lignes d'informations contextuelles et la ligne 'Anonymisation') comprend 10 lignes grâce aux procédures de normalisation, de nettoyage et d'enrichissement mises en œuvre (cf. Figure 3).

*0232 [461]	t(u) as pas d(e) bague.
Activité en cours [461]	Récréation
Interlocuteur [461]	un ou plusieurs enfant(s)
Situation langagière [24]	
Anonymisation [47]	
*0232_normalise [461]	t(u) as pas d(e) bague.
*0232_nettoyé [454]	tu as pas de bague.
*0232_tagging [454]	tu PRO c s tu as VER:pres avoir pas ADV pas de PRP de bague NOM bague . SEN
*0232_Négation [454]	tu as pas[NN0] de bague.
*0232_NégationPAS [454]	tu as pas[NN0] de bague.

Figure 3 – Vue d'un fichier de transcription (source : Lidilem, Univ. Grenoble Alpes)

C/ Extraction et comptage de formes langagières en vue de l'analyse quantitative des données

Le logiciel EAF² permet enfin 2 grands types d'extraction pour l'analyse quantitative des données. La première, que nous avons appelée 'sortie brute' présente, sous forme de tableau au format .csv, une ligne par occurrence de la forme recherchée avec en colonne l'identifiant (id) du locuteur, le tour de parole et l'énoncé dans lequel se trouve la forme recherchée, ainsi que les informations sur l'activité en cours, le ou les interlocuteurs et la situation langagière. Le second type d'extraction, nommée 'sortie organisée' se présente également sous forme de tableau au format .csv et est assimilable à un tableau de données sur lequel un traitement quantitatif des données est possible. On y trouve en ligne, l'id du locuteur, puis en colonne son temps de parole, le total nombre de mots produits, puis dans les colonnes suivantes le nombre d'occurrences de chaque forme en fonction des 140 combinaisons possibles d'informations contextuelles (5 catégories pour 'Activité en cours', 7 catégories pour 'Interlocuteurs' et 4 catégories pour 'Situations langagière').

À ces données peuvent être ajoutées les informations sociodémographiques dont nous disposons sur les participants. La jointure des fichiers est réalisée par un script R qui tient compte de l'id du locuteur mais également de la période de recueil de données afin de calculer automatiquement l'âge en mois des participants. Dans le cadre des données mises à disposition sur Ortolang, le fichier d'informations sociodémographiques donne pour chaque individu son âge en mois à chaque période de collecte des données.

3. DESCRIPTION DU CORPUS

Nous mettons à disposition 705 fichiers de transcriptions au format .eaf (Wittenburg et al., 2006). Les enregistrements audio ont été effectués lors de 3 périodes nommées P1, P2 et P3 qui correspondent à différents moments de l'année scolaire (P1 : début année scolaire 1, P2 : fin année scolaire 1, P3 : milieu année scolaire 2). Ils concernent les 96 enfants et 11 adultes pour lesquels nous avons un accord pour la conservation et le partage des données pour d'autres études. Ces transcriptions, produites à partir de la version 5.3 d'ELAN, sont compatibles avec les versions ultérieures du logiciel.

Le nombre de fichiers transcrits par individu est variable. Initialement, nous avons prévu de transcrire 2 fichiers audio par individu pour chaque période, ce qui correspond au nombre moyen de fichiers effectivement transcrits par individu. Toutefois ce critère s'est avéré insuffisant pour certains enfants – souvent les plus jeunes – qui parlaient très peu. Pour ces

derniers, nous avons visé la transcription d'au moins 1000 mots à partir de davantage de fichiers audio en allant jusqu'à transcrire 7 fichiers audio pour un même enfant. Dans certains cas, nous ne sommes pas parvenus à atteindre ce seuil. Cela concerne 10 enfants à P1, 5 enfants à P2 et 6 enfants à P3. Pour les adultes, plus prolixes, ce cas de figure ne s'est pas présenté.

Les transcriptions sont présentées dans deux dossiers distincts :

- 'Corpus complet' qui contient 2 dossiers : 'enfant' et 'adulte' ;
- 'Corpus longitudinal' qui contient uniquement les transcriptions des enfants présents aux 3 temps longitudinaux et pour lesquels 2 cohortes ont été distinguées :
 - 'Cohorte_PS⁶-MS' (contient les transcriptions de 37 enfants enregistrés de la PS à la MS)
 - 'Cohorte_MS-GS' (contient les transcriptions de 32 enfants enregistrés de MS à GS)

En outre, nous mettons à disposition un fichier de métadonnées au format .csv avec le point-virgule comme séparateur de champ. Ce fichier contient, pour chaque participant, son id ainsi que 23 colonnes contenant différentes informations (âge en mois aux différentes périodes, sexe, diplôme des parents, mode de garde avant l'école maternelle, etc.). Le détail sur les variables est présenté dans le dictionnaire des variables sur Ortolang.

3.1. DESCRIPTION DU CORPUS COMPLET

Le corpus complet contient environ 86 heures de prises de parole transcrites pour un total de plus de 746 000 mots. Il contient les transcriptions de 96 enfants, 4 enseignants et 7 Atsem. Le Tableau 1 indique le nombre total de fichiers en distinguant enfants et adultes et donne les temps de paroles transcrites ainsi que le nombre de *tokens* et de *types* dans chacun de ces 2 sous-corpus.

	Nombre de fichiers	Temps de parole	Nombre total de mots (<i>tokens</i>)	Nombre total de mots différents (<i>types</i>)
Enfants (96)	660	4 423 minutes (≈ 74 heures)	593 882	91 060
Adultes (4 enseignants, 7 Atsem)	45 (19 fichiers enseignants et 26 fichiers Atsem)	744 minutes (≈ 12 heures)	152 540	14 912

Tableau 1 – Présentation du corpus complet

Le corpus complet 'enfant' compte 45 filles et 51 garçons. Parmi eux, 60 comprennent et/ou parlent uniquement le français et 36 comprennent ou parlent une ou plusieurs autres langues que le français.

Vingt-quatre enfants ont une mère ayant un niveau bac ou inférieur, 22 ont une mère avec un niveau équivalent à bac + 2/3 et 41 enfants dont la mère a un niveau équivalent à bac + 4/5 ou plus. Pour 9 enfants, l'information ne nous a pas été communiquée.

3.2. DESCRIPTION DU CORPUS LONGITUDINAL

Le corpus longitudinal contient environ 60 heures de prises de paroles qui correspondent à un total d'un peu plus de 487 000 mots (cf. Tableau 2).

⁶ PS : Petite Section de Maternelle, MS : Moyenne Section de Maternelle, GS : Grande Section de Maternelle.

	Nombre de fichiers	Temps de parole	Nombre total de mots (<i>tokens</i>)	Nombre total de mots différents (<i>types</i>)
Cohorte PS-MS (37)	328	2 049 minutes (≈ 34 heures)	263 325	38 313
Cohorte MS-GS (32)	212	1 573 minutes (≈ 26 heures)	224 041	35 849

Tableau 2 – Présentation du corpus longitudinal

Il réunit les transcriptions de 69 enfants enregistrés lors de 3 temps longitudinaux répartis en 2 cohortes – cohorte PS-MS et cohorte MS-GS – pour lesquelles nous donnons quelques informations sociodémographiques dans le Tableau 3 ci-dessous.

		Cohorte PS-MS (37)	Cohorte MS-GS (32)
sexe	filles	17	16
	garçons	20	16
comprend ou parle une ou plusieurs autres langues que le français	comprend ou parle uniquement le français	22	21
	comprend ou parle une ou plusieurs autres langues que le français	15	11
niveau de diplôme de la mère	niveau bac ou inférieur	8	9
	niveau équivalent à bac + 2/3	10	10
	niveau équivalent à bac + 4/5 ou plus	17	9
	données manquantes	2	4

Tableau 3 – Corpus longitudinal : informations sociodémographiques

4. INTERET SCIENTIFIQUE ET POTENTIEL DE REUTILISATION

L'utilisation première des transcriptions issues du projet DyLNet mises à disposition sur Ortolang concerne principalement le domaine de la linguistique à des fins de description et d'analyse du langage enfantin et de son développement durant la période de scolarisation en maternelle. Elles permettent aussi une description des usages et discours des adultes en charge des enfants. Des travaux sur les usages langagiers enfantins et adultes en maternelle ont déjà été publiés au sein de l'équipe.

Nous avons abordé la question de la variation stylistique en contexte scolaire à travers l'analyse des usages de la négation bipartite variable (le fait de pouvoir supprimer le « ne » dans l'usage oral) chez 13 adultes (enseignants et Atsem) et 61 élèves âgés de 3 à 6 ans. À partir d'un corpus de parole correspondant à 640 000 mots transcrits, nous montrons que les enseignants manifestent des ajustements stylistiques en produisant davantage les deux particules de négation lorsqu'ils s'adressent à leurs élèves plutôt qu'à d'autres adultes. Quant aux enfants, leur usage des deux particules reste très marginal jusqu'à l'âge de 6 ans et ne manifeste pas d'ajustement stylistique (Buson et al., 2023).

Un autre travail s'est intéressé à l'acquisition des interrogatives entre 3 et 6 ans à partir du corpus DyLNet et du corpus PARM qui contient de la parole enfantine suscitée à partir d'albums lors d'ateliers menés en classe de maternelle (Buson et al., 2024; Buson et al., à paraître). Cette étude montre que bien que plusieurs variantes d'interrogatives soient présentes dans les conversations spontanées des enfants de maternelle, d'autres variantes, plus rares, sont complètement absentes (corpus DyLNet). Ces formes d'interrogatives, absentes des conversations spontanées ne sont toutefois pas complètement absentes des ré-

pertoires langagiers enfantins puisqu'on en trouve dans le corpus PARM amorcées grâce aux albums (Buson et al., à paraître).

Une autre recherche, menée cette fois-ci sur le discours des enseignants, s'est attachée à la manifestation des stéréotypes de genre dans les interactions enseignant-élèves. Différents indicateurs ont été examinés tels que la quantité et la nature du discours adressé aux élèves en fonction de leur sexe et une analyse plus fine du contenu pragmatique a été menée (Peuzin et al., 2022; Peuzin et al., 2023).

D'autres exploitations, via des collaborations avec des informaticiens et des chercheurs en science des réseaux, auront également lieu dans le cadre de la chaire *Socialization and Language at School: An Opportunity and a Challenge for AI in the Exploitation of Massive Data* (SoLaS) financée par MIAI Cluster de 2025 à 2029 (ANR-23-IACL-0006). Les recherches menées au sein de cette chaire visent 3 objectifs principaux constituant chacun un work package : 1/ la visualisation dynamique du réseau d'interactions sociales à l'école maternelle ; 2/ le développement d'un outil de transcription automatique de la parole spontanée des enfants, incluant des outils pour évaluer les niveaux de développement langagier ; 3/ la modélisation du réseau dynamique d'interactions scolaires et de ses influences réciproques avec l'usage du langage. Les données extraites des transcriptions viendront alimenter chacun des work packages pour 1/ visualiser l'évolution des usages langagiers au sein des réseaux d'interactions, 2/ servir de corpus de référence pour entraîner et évaluer l'outil de transcription, 3/ permettre l'exploration des effets des liens sociaux sur le développement langagier.

Le corpus de transcriptions mis à disposition sur Ortolang présente un accès restreint en conformité avec les engagements pris auprès des participants ayant signé le formulaire de consentement concernant la conservation des données après l'étude et leur accessibilité pour d'autres études : « Seuls des chercheurs affiliés à des institutions de recherche reconnues pourront utiliser ces données. Ils obtiendront un mot de passe permettant l'accès à la base après avoir décliné leur identité et explicité leur projet de recherche à A. Nardy, responsable du projet DyLNet. Seuls les projets à visée strictement scientifique pourront être acceptés. Les exploitations à visée commerciale ou partisane seront refusées. Les données seront exploitées uniquement dans un cadre de recherche scientifique, celle-ci étant soumise à la réglementation sur les données personnelles. »

L'accès au corpus se fait ainsi en cliquant sur « contacter le producteur » sur la page dédiée au corpus sur Ortolang.

5. REMERCIEMENTS

Nous remercions l'ANR (Agence Nationale de la Recherche), INRIA Rhône-Alpes et l'Université Grenoble Alpes pour le soutien financé apporté au projet DyLNet (ANR-16-CE28-0013).

Nous remercions l'ensemble des chercheurs impliqués dans le projet DyLNet. Hélène Bouchet, Jean-Pierre Chevrot, Laurence Buson (Lidilem, Univ. Grenoble Alpes), Stéphanie Barbu (EthoS, Université de Rennes), Céline Dugua (LLL, Université d'Orléans), Maryse Bianco (LaRAC, Univ. Grenoble Alpes), Loïc Liégeois (LLF, Université de Paris & CLILLAC-ARP, Université Paris Cité), Éric Fleury, Márton Karsai, Christophe Crespelle, Anthony Busson, Sicheng Dai (équipe DANTE, INRIA Grenoble Rhône-Alpes, ENS de Lyon, UCB Lyon I).

La mise en place de ce projet a également nécessité l'engagement d'autres collaborateurs que nous remercions pour leur participation :

Les collègues :

Mathieu Loiseau (Lidilem, Univ. Grenoble Alpes) : pour toutes les discussions autour de l'acquisition des données audio ; Sylvain Hatier et Arnaud Bey (Litt&Arts, Univ. Grenoble

Alpes) : pour leur participation au développement du logiciel EAF² ; Solange Rossato (LIG, Univ. Grenoble Alpes) : pour les discussions autour de l'acquisition des données audio et son script PRAAT de repérage des silences/paroles ; Christophe Braillon (INRIA Rhône-Alpes) : pour son support technique lors des phases de récupération des données ; Jean-François Cuniberto (INRIA Rhône-Alpes) : pour la conception et fabrication de présentoirs ad hoc pour le rechargement des boîtiers dans les salles de classes ; Jean-Marc Joseph (INRIA Rhône-Alpes) : pour la mise en place du serveur dylserv.

L'équipe des transcripateurs, étudiants, stagiaires :

Sophie Loiseau, Emmanuelle Roman, Lisa Paillussière, Bérénice Bonson, Sarah Viratelle, Beatriz Villa, Mélanie Mascré-Pothier, Myriam Salze, Louis Roché, Marie Peuzin, Constance Manent, Antonine Massat, Soline Felice, Charlotte Chauvet, Kim Nalin, Célia Bron, Charlotte Froissard, Tiphaine Mout, Claire Wolfarth, Lou Maume, Lili Fourré, Rose Moreau Raguene, Chenxi Zhang, Léonie-Rose Zecchi--Morales

L'équipe des développeurs :

Mélanie Veloso, Julien Fagot, Youssef Hamrouni, Bin Yang, Kaiyan Li

6. CONTRIBUTION

Aurélié Nardy : Conceptualisation, préparation des données, analyse formelle, acquisition de financements, investigation, méthodologie, administration du projet, ressources, supervision, validation, présentation visuelle, rédaction – ébauche du manuscrit initial, rédaction – relectures et correction

Isabelle Rousset : Conceptualisation, préparation des données, analyse formelle, investigation, méthodologie, ressources, logiciel(s), supervision, validation, présentation visuelle, rédaction – ébauche du manuscrit initial, rédaction – relectures et correction

Rachel Gaubil : dépôt en ligne des données, logiciel(s), rédaction – relectures et correction

7. DECLARATION D'INTERETS CONFLICTUELS

Nous ne déclarons pas d'intérêts conflictuels.

8. REFERENCES BIBLIOGRAPHIQUES SCIENTIFIQUES

Baude, O. et Dugua, C. (2011). (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? *Corpus*, 99-118. <https://doi.org/https://doi.org/10.4000/corpus.2036>

Boersma, P et Weenink, D. (2025). *Praat: doing phonetics by computer* [Computer program]. Version 6.4.34. <http://www.praat.org/>

Busby, P. A. et Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America*, 97(4), 2603-2607. <https://doi.org/https://doi.org/10.1121/1.412975>

Buson, L., Chevrot, J.-P., Nardy, A. et Rousset, I. (à paraître). Acquisition of syntactic variation: Production of interrogatives in French preschool children. Dans L. Rosseel et E. Zenner (dir.), *Variation in Language Acquisition: Unity in diversity*. John Benjamins.

Buson, L., Nardy, A., Rousset, I. et Zhang, C. (2022). Pratiques stylistiques en contexte scolaire : usages contrastés du 'ne' de négation chez les adultes et les enfants d'une école maternelle française. *Congrès Mondial de Linguistique Française. SHS Web of Conferences* 138(Article 06018). <https://doi.org/https://doi.org/10.1051/shsconf/202213806018>

Buson, L., Nardy, A., Rousset, I. et Zhang, C. (2023). Style in a school community—"Ne" deletion in French preschool. *Journal of Sociolinguistics*, 27, 384-401. <https://doi.org/310.1111/josl.12611>

- Buson, L., Rousset, I. et Rossato, S. (2024). « S'exprimer dans un langage oral syntaxiquement correct et précis ». *Syntaxe et variation en maternelle. Pratiques* [en ligne], 201-202. [https://doi.org/ https://doi.org/10.4000/122u3](https://doi.org/https://doi.org/10.4000/122u3)
- Buson, L., Rousset, I. et Rossato, S. (à paraître). *Méthode Parm - PARler Raconter en Maternelle*. UGA Éditions.
- Dai, S., Bouchet, H., Karsai, M., Chevrot, J.-P., Fleury, E. et Nardy, A. (2022). Longitudinal data collection to follow social network and language development dynamics at preschool. *Scientific Data*, 9(Article 777). <https://doi.org/10.1038/s41597-41022-01756-x>
- Hriba, L., Baude, O. et Dugua, C. (2011, 27-28 mai 2011). *Transcrire : la norme, la variation et le linguiste*. Colloque du CerLiCO : Transcrire, écrire, formaliser 2, Orléans, France.
- Nardy, A., Rousset, I., Bey, A., Fagot, J., Hatier, S., Hamrouni, Y. et Gaubil, R. (2020). EAF² – Logiciel d'enrichissement et d'analyse de fichiers de transcriptions faites sous ELAN. <https://hal.science/hal-03604589>
- Nardy, A., Rousset, I., Gaubil, R., Bouchet, H. (2025). *Corpus DyLNet* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). <https://hdl.handle.net/11403/corpus-dylnet>.
- Peuzin, M., Nardy, A. et Chevrot, J.-P. (2022). Discours adressé à l'enfant par les enseignants en maternelle : des différences filles garçons ? *Congrès Mondial de Linguistique Française. SHS Web of Conferences*. 138(Article 01002). <https://doi.org/10.1051/shsconf/202213801002>
- Peuzin, M., Nardy, A. et Chevrot, J.-P. (2023). Évaluer à l'école maternelle : des pratiques sexuées ? *Genre Éducation Formation (GEF)* [en ligne], 7. <http://journals.openedition.org/gef/1248>
- Potamianos, A., Narayanan, S. et Lee, S. (1997). Automatic speech recognition for children. *Proceedings of the Eurospeech 1997*, 2371-2374.
- Weinberg, B. et Bennett, S. (1971). Speaker sex recognition of 5- and 6-year-old children's voices. *Journal of the Acoustical Society of America*, 50(4), 1210-1213. <https://doi.org/https://doi.org/10.1121/1.1912757>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. et Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. Dans *Proceedings of the Fifth International conference on Language Resources and Evaluation* (p. 1556-1559).