

Évaluer le lexique des enfants de 2 à 4 ans en milieu scolaire : quelles méthodes ? quels outils ?

Samantha Ruvoletto* et Marie-Laurence Knittel

Université de Lorraine (Nancy) et Laboratoire CNRS ATILF

*samantha.ruvoletto@univ-lorraine.fr

Résumé. Notre travail se concentre sur la présentation des outils disponibles pour comprendre, décrire et évaluer le vocabulaire des enfants âgés de 2 à 4 ans, ainsi que sur l'introduction du projet Lexique à l'Ecole Maternelle (Lex.E.M) et du corpus Jeunes Enfants en Milieu Scolaire (JEMS) qui en découle. Cette présentation est suivie d'une analyse préliminaire, à la fois quantitative et qualitative, d'une partie de ce corpus, permettant ainsi de donner un portrait général du vocabulaire utilisé par les enfants à l'école maternelle dans le cadre des interactions enseignant-groupe d'enfants. Ces analyses nous permettent de poser des questions sur les méthodes les plus efficaces pour récolter des données en milieu scolaire qui puissent permettre d'évaluer efficacement ce lexique spécifique.

Introduction

Notre travail est dédié à l'étude des outils disponibles pour évaluer le lexique des enfants de 2-4 ans, ainsi qu'à la présentation du projet Lexique à l'Ecole Maternelle (Lex.E.M) et du corpus Jeunes Enfants en Milieu Scolaire (JEMS), élaborés dans ce cadre. Dans un premier temps, nous présentons les outils et les ressources employés pour l'évaluation du lexique des enfants en détaillant leurs points positifs et leurs limites.

Nous présentons ensuite nos objectifs et l'intérêt du projet Lex.E.M. et de la création d'un nouveau corpus en milieu scolaire, le corpus JEMS. Cette présentation est suivie d'une analyse préliminaire (quantitative et qualitative) d'une partie de ce corpus, qui permet de décrire le lexique employé d'une manière générale par les enfants à l'école maternelle. Enfin, nous nous interrogeons sur la taille et la variété du lexique utilisé par les enfants en contexte scolaire, et aux meilleurs moyens de l'évaluer et de favoriser son développement.

Au-delà du développement d'outils de remédiation, le corpus JEMS nous permettra d'entreprendre des analyses qualitatives et quantitatives portant sur cette composante essentielle du langage, en examinant notamment le lexique de l'école en comparaison avec le lexique employé à la maison et dans la sphère familiale, ou encore avec le lexique de l'adulte en contextes scolaire et familial.

Outils et ressources employés pour l'évaluation du lexique des enfants

Pour évaluer le lexique employé par les enfants, plusieurs méthodes peuvent être utilisées. Nous les présentons dans cette section.

1.1 Le questionnaire

Pour décrire le lexique des enfants en milieu familial, les chercheurs peuvent questionner les parents en utilisant des questionnaires parentaux. Ce type de questionnaire est rempli par le/les parents, qui coche(nt) les items produits naturellement par l'enfant en milieu familial. Le DLPP (Développement du Langage et Production en Français) en constitue un exemple. Ce questionnaire en trois parties, élaboré par Bassano, Labrell et Bonnet (2020) vise à recueillir des données lexicales, syntaxiques et pragmatiques émanant des enfants de 1 à 4 ans.

Concernant la partie lexicale, ce questionnaire, inspiré du *MacArthur-Bates Communicative Development Inventories* (MCDI) (Fenson et al. 1993), se présente sous la forme de listes de mots relevant de divers

champs lexicaux (personnes, animaux, véhicules, etc.), susceptibles d'être employés par les enfants. Ces listes ont été constituées à partir des productions des enfants recensées dans les bases PLPNat (Bassano 2005, 2007) et CHILDES (MacWhinney 2000).

Comme d'autres types de questionnaire, le DLFP permet d'obtenir rapidement une idée générale du lexique de l'enfant à différents âges, sur les plans quantitatif et qualitatif. Cependant, ce lexique est construit *via* le regard de parents, qui peuvent parfois manquer d'objectivité. De plus, il est limité au lexique utilisé par les parents, et ne prend pas en compte l'entourage proche possiblement présent au domicile (baby-sitter, grands-parents, etc.). Il est en outre évident que de tels questionnaires ne fournissent aucune information concernant le lexique employé en milieu scolaire.

1.2 Les corpus

Les productions des enfants peuvent être évaluées à partir de corpus. La création d'un corpus et sa transcription demandent du temps et de l'investissement de la part du groupe de recherche impliqué. À la différence du questionnaire, un corpus permet d'avoir des informations objectives concernant le lexique de l'enfant enregistré, grâce à la mise en place d'enregistrements écologiques au domicile familial, dans le cadre d'une situation naturelle et spontanée.

On peut citer en exemple le corpus CoLaJE (Communication Langagière chez le Jeune Enfant, Morgenstern & Parrisé 2012), disponible sur ORTOLANG, ainsi que sur CHILDES, sur le nom de *corpus Paris*. Il comporte des suivis longitudinaux de dialogues parents-enfants de 7 enfants francophones natifs, de la naissance jusqu'à l'âge de 6 ans. Les enregistrements transcrits et interrogeables peuvent donc être utilisés comme ressource comparative pour examiner les productions lexicales des jeunes enfants, sous les aspects qualitatifs et quantitatifs.

D'autres corpus francophones sont disponibles sur la plateforme CHILDES et consultables en ligne. La plupart de ces corpus se basent sur des enregistrements écologiques au domicile de l'enfant, ne permettant pas de généraliser ses productions au milieu scolaire. De plus, ces corpus ayant été réalisés préalablement à 2012, ils ne garantissent pas une description du lexique de l'enfant à l'époque actuelle.

Deux corpus d'enregistrements en milieu scolaire sont mentionnés sur la plateforme CHILDES.

D'une part du corpus de Palasis (Palasis, 2010) regroupe les productions longitudinales de 18 à 20 enfants de 2,5 ans à 5,11 ans enregistrés en milieu scolaire pendant 3 ans. Il se compose d'enregistrements de dialogues en petits groupes entre enfants et évaluateur. Ce corpus est donc semi-écologique, étant donné l'intervention de l'évaluateur, dont l'objectif était de faire raconter aux enfants leurs diverses activités en milieu scolaire, de les faire participer à des jeux ou raconter des histoires à partir de livres destinés à ce public. Toutefois, ce corpus et les activités proposées ayant pour objectif d'étudier le développement et l'emploi des pronoms et des syntagmes nominaux chez les enfants, certaines productions de l'enfant risquent d'avoir été impulsées ou favorisées (voir à ce propos Bornstein, Tamis-LeMonda & Haynes 1999). Il ne permet pas donc d'évaluer le lexique spontané de l'enfant à l'école.

D'autre part, le corpus DylNet (Nardy et al. 2016-2020) comporte des enregistrements longitudinaux d'environ 200 enfants et intervenants d'une école maternelle pendant 3 ans. Ces enregistrements ont été réalisés pendant une semaine chaque mois, grâce à des capteurs, qui ont enregistré en parallèle les proximités entre individus. Les objectifs du projet DylNet étaient d'étudier les relations entre l'apprentissage du langage oral et la socialisation enfantine par la mise en relation de trois facteurs : interactions sociales, performances linguistiques et usages sociolinguistiques des enfants. Le corpus, lorsqu'il sera accessible, constituera certainement une ressource précieuse pour l'étude du lexique employé par les enfants à l'école maternelle.

Mentionnons enfin le corpus CLEA, recueilli en 2014 (voir Péroz 2016), qui regroupe des transcriptions écrites d'enregistrements de 12 "séances de langage" dans une classe de Moyens-Grands. Il s'agissait pour les élèves de reformuler une histoire racontée par un enseignant, et de donner leur avis sur certains de ses

aspects. Cependant, les productions des élèves ne sont pas totalement spontanées, puisqu'ils ont choisi de répéter pour se préparer à l'enregistrement.

1.3 Les bases de données

La base de données lexicales Manulex (Lété et al. 2004) peut également être employée comme outil de référence pour évaluer le lexique des enfants. Manulex comporte 23900 lemmes (correspondant à 49000 formes fléchies), extraits de manuels scolaires destinés à l'école primaire (CP à CM2). Il s'agit donc d'un lexique écrit, destiné à un public plus âgé, et potentiellement éloigné des productions réelles des enfants. Cette base n'est donc pas adaptée au public de l'école maternelle. Quant à la base Lexique.org (New et al., 2001), elle a été constituée à partir de livres et de films qui ne sont pas spécifiquement destinés aux enfants. Les textes et les dialogues adultes ne sont certainement pas représentatifs du lexique de l'enfant ou nécessaire à l'enfant.

1.4 Les tests

Enfin, plusieurs batteries de tests permettant l'évaluation des compétences lexicales des enfants sont fréquemment utilisées par les orthophonistes pour établir des bilans orthophoniques. Nous pouvons citer EVALO 2-6 (Coquet, Roustit & Ferrand 2009) ou EXALANG 5-8 (Croteau et al. 2010). Ces tests reposent principalement sur deux types d'épreuves : la désignation d'images, au cours de laquelle l'enfant désigne l'image correspondant à l'item prononcé par l'évaluateur, qui permet d'évaluer la compréhension des items lexicaux, et la dénomination, au cours de laquelle l'enfant doit nommer l'objet ou la notion abstraite qui lui sont présentés sous forme d'image. Cependant, ces tests sont contestables, notamment en raison du choix des items lexicaux utilisés, qui s'avère problématique (Da Silva-Genest et al. 2020).

1.5 Bilan

Les divers outils, ressources et méthodes de recueil et/ou d'évaluation lexicale qui viennent d'être présentés mettent en évidence l'absence de dispositif permettant l'évaluation du lexique des enfants à l'école maternelle, en particulier en Petite et Moyenne Sections, soit 2-4 ans. Deux raisons principales peuvent être évoquées conjointement pour rendre compte de l'inadaptation des moyens existants : d'une part, les outils destinés au milieu scolaire sont faits pour des enfants à l'école primaire, et d'autre part, les corpus disponibles pouvant servir de référence pour les enfants de 2 à 4 ans sont des enregistrements soit réalisés en milieu familial (CoLaJe), soit faisant intervenir l'évaluateur (Palasis 2010), ou bien ne sont pas encore disponibles (DyINet). Quant au DLPF, il est spécifiquement destiné aux parents, et s'inscrit donc dans le milieu familial. De ce fait, une base de données propre au lexique de l'enfant de 2-4 ans apparaît comme une ressource nouvelle, répondant à un besoin.

2 Le projet Lex.E.M. et le corpus JEMS

Le projet LEXique à l'Ecole Maternelle (Lex.E.M) s'est donné pour objectif l'élaboration d'une nouvelle base de données, fondée sur les productions réelles des enfants de 2-4 ans, et des interactions adultes/enfants dans un contexte scolaire. Cette base permettra ainsi de combler le manque décrit ci-dessus (Section 2.5). Au-delà de ce premier objectif, elle permettra également aux chercheurs d'étudier le lexique des jeunes enfants, et d'en élaborer une cartographie reposant sur les mêmes critères que ceux qui sont employés pour décrire celui des adultes (classes de mots, champs lexicaux, relations lexicales, etc.)

Nous nous proposons dans ce cadre de créer un nouveau corpus, le corpus Jeunes Enfants en Milieu Scolaire (JEMS), élaboré à partir d'enregistrements inédits en milieu scolaire, et transcrits. Notre projet s'inscrit dans le cadre d'une collaboration avec des écoles maternelles, y compris appartenant au Réseau d'éducation Prioritaire (REP) de Lorraine, notamment des écoles de Nancy et de la proche banlieue, et en milieu rural, dans les Vosges et concerne les enfants de 2 à 4 ans (classes de Petits et Petits-Moyens).

2.1 Origine et objectifs du projet

Le projet émane de la demande d'enseignants d'écoles maternelles situées en REP, dont le public est hétérogène et comporte de nombreux enfants allophones, avec ou sans connaissance du français. Dans les classes, les enseignants ont repéré chez les enfants des lacunes importantes au niveau lexical, y compris concernant le vocabulaire basique utilisé à l'école. Ces lacunes se situent à la fois en compréhension et en production. Elles sont d'autant plus flagrantes que les enfants sont accueillis dès 2 ans et demi.

Or, le lexique, son étude et le développement du vocabulaire font l'objet de recommandations officielles dès le cycle 1 de l'École Maternelle, et l'enseignement du vocabulaire donne lieu à un guide et des fiches destinés aux enseignants². Le travail sur le lexique doit en effet permettre l'enrichissement du vocabulaire des élèves par l'acquisition de nouveaux mots, tant sur leurs aspects de forme, de sens et d'usage, dans le but d'améliorer les capacités de compréhension et de favoriser l'apprentissage de la lecture et la compréhension de l'écrit. Dans ce contexte, des lacunes lexicales peuvent être lourdes de conséquences, y compris à l'école primaire et au-delà, pour tout type de public, mais encore plus pour celui des zones REP et REP+, particulièrement fragile.

Notre projet à moyen et long terme se fixe pour objectif de développer des outils pour la remédiation, spécifiquement destinés à un public de 2 à 4 ans. Ces outils incluent notamment des activités visant à développer le lexique des enfants en difficulté linguistique, bilingues ou francophones natifs, en favorisant l'intégration et l'échange linguistique.

Toutefois, pour que les outils développés soient adaptés aux enfants en milieu scolaire, il est nécessaire de connaître le lexique qui y est employé, et comment il l'est. Cette connaissance porte sur plusieurs de ses aspects, notamment :

- la taille et la variété du lexique utilisé,
- les champs lexicaux employés,
- les catégories lexico-grammaticales prédominantes,
- la répartition adulte / enfant des tours de parole.

Notre objectif est donc de cerner le lexique employé par les enfants, à la fois dans ses aspects quantitatifs et qualitatifs.

De ce fait, préalablement à toute autre activité, nous avons entrepris de cartographier le lexique employé en situation scolaire par les enseignants et les enfants de 2 à 4 ans tout-venant. Pour ce faire, nous avons procédé à l'enregistrement d'activités en milieu scolaire, regroupés dans le corpus JEMS, dont nous décrivons la constitution ci-dessous.

2.2 Le corpus JEMS

Le corpus JEMS³ est constitué de captations audio et vidéo écologiques réalisées en 2020 dans trois écoles maternelles : l'École Jean Jaurès (Nancy, 54), l'école de Hadol et l'école Jean Rostand de Bruyères, situées dans les Vosges (88). Étant donné le caractère bruyant de l'environnement, les captations ont été réalisées au moyen de perches télescopiques, qui ont permis de pointer vers la source-cible, et donc de maximiser le rendu audio. Les enregistrements ont été réalisés par des stagiaires-étudiants de L3 Sciences du Langage, qui ne pouvaient pas intervenir pendant les enregistrements.

Pour respecter le RGPD (Règlement Général sur la Protection des Données), des consentements écrits ont été diffusés aux parents pour signature préalablement à l'enregistrement. Les enregistrements ont été transcrits au moyen du logiciel PHON (Hedlund & Rose 2020). Les productions des enfants ont fait l'objet d'une transcription phonétique (lignes *IPA Actual* & *IPA Target* dans le logiciel), les productions adultes ont été transcrites orthographiquement.

Plusieurs types d'activités ont été filmées dans chaque école. Toutes incluent un groupe d'enfant en présence d'un ou deux adultes (enseignant et/ou ATSEM). Plusieurs sessions d'enregistrement ont été réalisées, en

Petite et/ou Moyenne Section, notamment en cas de classe à double niveau. Notre corpus comporte au total 234,4 minutes d'enregistrement transcrits, soit 6381 records dans le logiciel PHON, où un record correspond à un tour de parole par un locuteur adulte ou enfant.

Le Tableau 1 présente une vision globale des enregistrements transcrits et des activités réalisées lors de ces enregistrements.

Tableau 1. Enregistrements transcrits

École ⁴	Sessions d'enregistrement		Activité(s)	Classe(s)	Nombre de records	Durée de l'enregistrement (minutes, secondes)
Bru	01_10	18	Activité : lexique de la cuisine	MS	265	10
	01_13	17	Activité sur le langage	PS	181	8
	02_10	27	Activité sur les prénoms	PS	154	7
	03_09	2	Activité sur les prénoms	PS	120	6
	01_10	16	Activités diverses	MS-PS	419	13,5
		19		MS-PS	314	11,2
	01_13	14		MS	464	14,1
		21		PS	464	14,1
	02_10	26		MS	390	10,6
		28		PS	501	15
		30		MS-PS	283	9,6
	02_03_20	2	Activité : mathématiques placer des éléments dans un ordre numérique sur un support	PS	320	9,5
		3		PS	394	21
	07_20_20	1	Activité : coloriage du clown	MS-PS	445	25
		3	Activité : autour des prénoms	MS-PS	286	9
		2		MS-PS	380	11,5
JJ	02_0320	1	Activité en petit groupe : le train des animaux	PS	417	14,5
	01_03_20	6	Activité : le jeu des couleurs	PS	515	17,8
JJ	01_03_20	5	Activité : puzzle	PS	69	5

3 Premières analyses

Nous avons effectué nos premières analyses à partir du corpus recueilli à Bruyères (Vosges). C'est en effet l'école dans laquelle le plus grand nombre d'enregistrements a été effectué. Elle offre donc une quantité de données assez importante pour réaliser un nombre significatif d'analyses quantitatives et qualitatives.

3.1 Données chiffrées

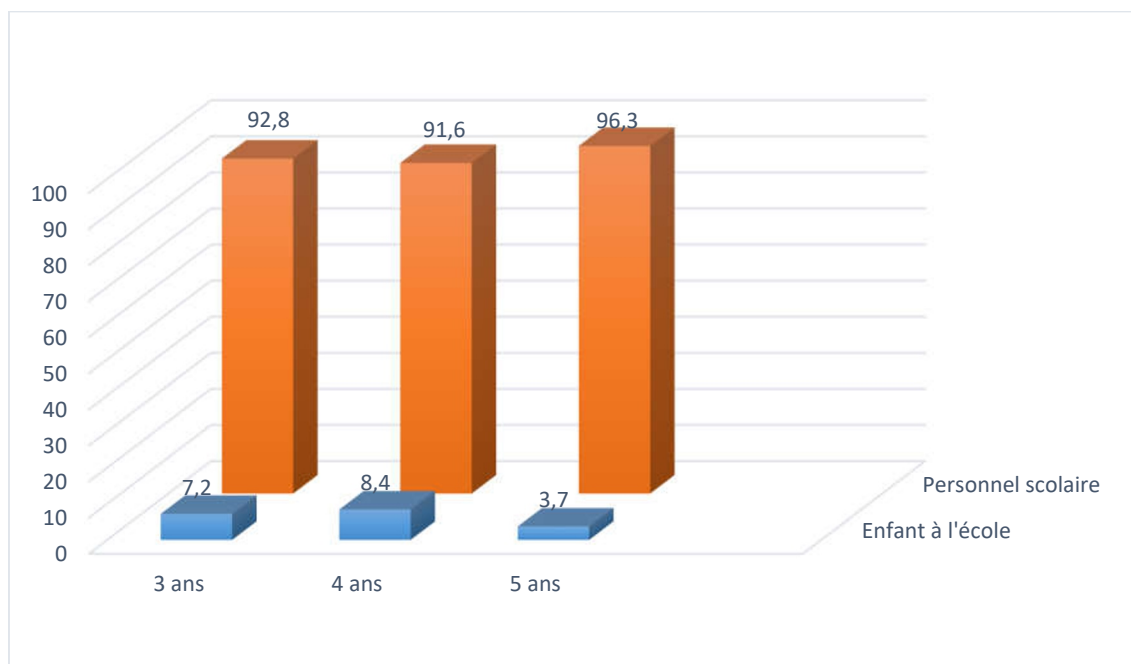
En regroupant les différentes sessions d'enregistrement, nous avons recueilli les productions de dix-huit locuteurs, seize enfants et deux adultes (l'enseignante et l'ATSEM).

Le Tableau 2 présente les données concernant les locuteurs impliqués, classés par âge croissant et anonymisés. Les locuteurs adultes apparaissent en fin de tableau.

Tableau 2. Synthèse des données concernant les enfants et les adultes enregistrés à Bruyères

Locuteurs	Age (AA;MM;JJ)	Nombre d'enregistrements
KZ	03;03,22	2
NM	03;04,09	5
CL	03;05,13	4
TM	03;05,00	4
KZ	03;09,01	3
NL	03;09.17	3
NT	03;09.17	3
TH	03;11,20	5
RB	04;02.27	4
GL	04;03,26	7
MY	04;03,22	3
MS	04;06,26	5
JY	04;06,26	6
LV	04;11,05	3
CS	05;00,24	6
MA	05;01,15	4
VG	Adulte (enseignante)	11
SA	Adulte (ATSEM)	3

Le Graphique 1 présente les pourcentages de mots produits par les enfants selon leur âge au moment de l'enregistrement, en comparaison avec le personnel scolaire (enseignante/ATSEM).

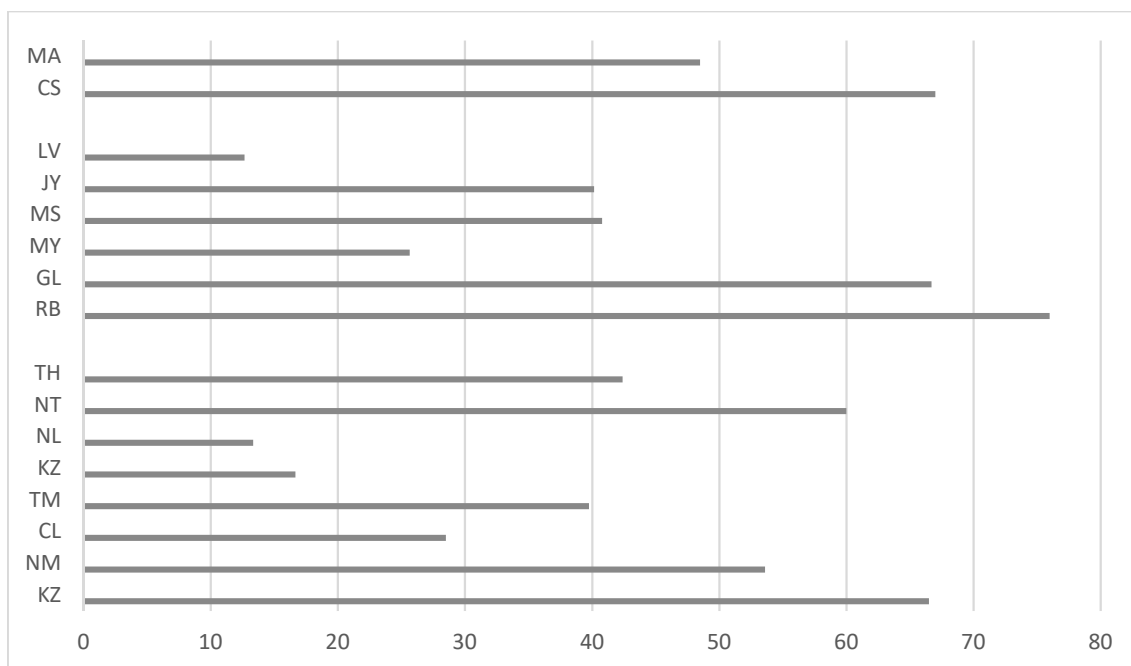


Graphique 1. Pourcentages de mots produits par le personnel scolaire et les enfants

Le graphique 1 permet de constater que les pourcentages de mots produits par les adultes se monte à environ 90 % de la production totale enregistrée. Cette répartition est en accord avec la tendance observée par Kidd & Rowland (2021) au Royaume-Uni, qui ont noté que les productions des enseignants représentent la majorité du temps de parole, même si le taux relevé est inférieur au nôtre (63% de productions des enseignants vs. 37% de productions des enfants). De plus, le taux de productions de l'adulte ne change pas en fonction de l'âge de l'enfant, à la différence de ce qu'on observe en analysant les corpus parents-enfants et les tendances du langage adressé à l'enfant en milieu familial, où les temps de parole des parents diminuent en relation à l'augmentation de l'âge de l'enfant (Raneri et al. 2020). En outre, les caractéristiques spécifiques du langage « en classe » fait émerger l'idée d'un « parler enseignant » qui diffère, sur certains aspects, du français dans le discours adressé aux jeunes enfants d'âge préscolaire. Ceci confirme la conclusion que Buson et Nardy (2020) tirent de l'analyse contrastive des échanges enseignants-enfants et enseignants-autres adultes dans des classes de Petite Section d'École Maternelle.

Le taux élevé de temps de parole de l'enseignant peut provenir du rôle spécifique du personnel scolaire. Les adultes en milieu scolaire dirigeant les activités collectives, il est normal qu'ils prennent plus souvent la parole face aux enfants de tous âges confondus, notamment pour les guider ou les encourager à s'exprimer. Cependant, les « espaces de parole » créés ne sont pas remplis par des productions des enfants, au contraire. Ils restent très à l'écart avec un taux de productions qui ne dépasse pas 6-7%. Le taux de production de chaque enfant pris individuellement est donc très faible. Des observations similaires ont été faites par plusieurs chercheurs (voir par exemple Péroz 2010 ; Canut, Espinosa et Vertalier 2013).

En ce qui concerne les productions des enfants, le Graphique 2 présente le nombre de mots moyen produit par enfant par enregistrement, selon les groupes d'âges au moment de l'enregistrement. Les enfants de 5 ans environ apparaissent haut du graphique, les enfants de 4 ans environ au milieu et ceux de 3 ans environ en bas.



Graphique 2. Nombre de mots moyen produit par enfant dans chaque enregistrement

Le Graphique 2 ci-dessus nous permet de faire un certain nombre d'observations.

En premier lieu, on remarque chez les enfants d'importantes variations interindividuelles. Ainsi, on peut observer, à l'intérieur des groupes les plus nombreux (3 ans et 4 ans), des décalages importants entre enfants. Chez les enfants de 3 ans, KZ a produit plus de 65 mots en moyenne par enregistrement, par contraste avec NL, dont les productions sont inférieures à 15 mots. Chez les enfants de 4 ans, les productions de LV sont inférieures à 15 mots (et donc comparables à celles de NL), alors que celles de RB — le plus loquace tous groupes d'âge confondus — se montent à plus de 75 mots en moyenne.

Le graphique 2 montre donc que, dans une même tranche d'âge, certains enfants parlent beaucoup plus que d'autres. A cet âge toutefois, de telles variations sont normales. Selon Le Normand, Parisse & Cohen (2008), elles sont imputables à des facteurs socioculturels, ce qui va dans le sens des observations formulées au début de la Section 3.1. On observe en outre que les productions moyennes des deux enfants de 5 ans sont proches ou supérieures à 50 mots, et excèdent les moyennes des autres groupes d'âge. Considérant que les enfants les plus âgés disposent d'un lexique plus vaste, on s'attend effectivement à ce qu'ils soient à même de produire plus de mots que les enfants plus jeunes.

3.2 Analyse qualitative

Dans cette section, nous présentons d'abord quelques observations résultant du dépouillement du lexique utilisé au cours d'une activité filmée à l'École de Bruyères, que nous comparons avec quelques données relevant de l'ensemble des enregistrements réalisés dans cette même école.

3.2.1 Analyse d'activité

Dans cette section, nous présentons quelques observations résultant du dépouillement du lexique utilisé au cours de l'activité "Lexique de la cuisine" du Tableau 1. Cet extrait est présenté dans la Figure 1.

VG : Maintenant c'est l'inverse je vous montre une image

VG : vous regardez on ne dit rien on regarde d'abord
VG : on utilise les yeux
VG : on utilise ses yeux
VG : LV qu'est-ce que tu en penses ?
MS : une casserole
LV : une casserole
VG : une casserole. Maëlys aussi ?
MS : une casserole
VG : c'est une casserole ?
VG : et toi MA qu'est-ce que tu en penses ?
MA : casserole
VG : et bah moi je ne suis pas d'accord
VG : ce n'est pas une casserole
VG : LV va voir me chercher la casserole dans la dinette
VG : on va voir si ça ressemble
MA : c'est pour cuire les œufs
VG : c'est pour cuire oui
MA : les œufs
VG : les œufs alors est ce qu'on fait cuire les œufs dans la casserole ?
Célestin : non
Célestin : dans la passoire
VG : ah alors qu'est-ce qu'elle nous ramène LV ? est-ce qu'elle nous ramène la casserole ?
VG : est-ce que c'est ça la casserole ?
CS : non
MA : c'est pour cuire les œufs
VG : ça c'est pour faire cuire les œufs. Va voir chercher l'œuf
VG : là-bas qu'on fait cuire
VG : là. Comment ça s'appelle ça pour faire cuire les œufs ?
JY : casserole

Figure 1. Extrait transcrit de l'activité "Lexique de la cuisine"

À partir de cet extrait, nous avons répertorié les classes de mots lexicaux employés (noms, verbes et adjectifs) et leur nombre d'occurrences, répartis selon le locuteur (adulte ou enfant). Ces données sont présentées dans le Tableau 3.

Tableau 3. Lexique employé dans l'extrait de l'activité « Lexique de la cuisine » et répartition

		Nombre d'occurrences	Productions adulte (VG)	Productions enfants
Noms	inverse	1	1	—
	image	1	1	—
	yeux	2	2	—
	casserole	12	6	6
	dinette	1	1	—
	œuf(s)	8	4	4
	passoire	1	—	1
Total Noms	7	26	15	11
	utiliser	2	2	—
	montrer	1	1	—
	regarder	2	2	—
	dire	1	1	—
	penser	2	2	—
	chercher	2	2	—
	ressembler	1	1	—
	ramener	2	2	—
	s'appeler	1	1	—
	faire	4	4	—
	cuire	7	5	2
	aller	3	3	—
	voir	3	3	—
	être	12	10	2
Total Verbes	14	43	39	4
Total Noms et Verbes	21	69	54	15

Nous pouvons formuler quelques observations à partir du Tableau 3.

Il est à noter en premier lieu que notre extrait ne comporte aucun adjectif. En ce qui concerne les noms, on remarque que seulement 4 d'entre eux sur 7 appartiennent au lexique spécifique à l'activité : *dinette*, *casserole*, *œuf(s)*, *passoire*. Seuls *casserole* et *œuf(s)* sont employés à part égale par l'adulte et les enfants,

alors que *passoire* est produit exclusivement par un enfant, et *dinette* par l'adulte. Les noms qui n'appartiennent pas au lexique de la cuisine (*inverse*, *image*, *yeux*) sont employés par l'adulte dans les instructions données aux enfants ; il en est de même pour *dinette*.

Considérons à présent les verbes. En premier lieu, on note que le nombre de verbes différents employés (14) excède le nombre de noms (7), ce qui est inattendu. Parmi ces verbes, on observe une prédominance de *être*, le seul verbe à apparaître plus de 10 fois. Si on observe les productions dans la Figure 2, on constate en effet que ce verbe apparaît fréquemment dans des présentatifs (*c'est* (4 fois) / *ce n'est pas* (1 fois)), ou des interrogatives (*est-ce que* (3 fois), *qu'est-ce que* (3 fois)). Le seul autre emploi de *être* est *je suis pas* (*d'accord*). L'importance de ce verbe dans nos données est donc surtout liée à des emplois dans des structures figées. On remarque en outre que ce verbe est majoritairement produit par l'adulte (12 occurrences sur 14). Deux autres verbes sont également utilisés de manière similaire : *faire*, employé comme un semi-auxiliaire causatif (*faire cuire*), et non pas comme un verbe plein, et *voir* dans 2 occurrences sur 3, en tant que synonyme de *donc* (*va voir pour va donc*). Le nombre de verbes pleins se réduit donc à 12. Parmi ces verbes, un seul verbe relève du lexique spécifique de la cuisine. Il s'agit de *cuire*, qui est le verbe le plus fréquent dans l'extrait, et le seul à être produit par un enfant, MA (2 fois). On remarque d'ailleurs que c'est également MA qui a produit la première occurrence de ce verbe.

Comme dans le cas des noms, le lexique verbal est donc principalement fourni par l'adulte, et ne correspond pas exclusivement à celui de la thématique de l'activité —ici la cuisine—, mais plutôt aux instructions et commentaires émanant de l'adulte qui guide l'activité. Lorsqu'il concerne l'activité elle-même, le lexique employé par les enfants se rapproche de celui de l'adulte. Par exemple, *œuf* et *casserole* sont produits à égalité par les enfants et l'adulte, et *cuire*, le seul verbe du champ lexical de l'activité est employé 2 fois sur 7 par un enfant.

Pour conclure, les données étudiées indiquent que le lexique employé par les enfants est plus restreint, à la fois qualitativement (5 mots sur 21) et quantitativement (21,7% du total), que celui de l'adulte (14 mots sur 21 et 78,3% du total). Cependant, cet état de fait s'avère finalement compréhensible, du fait même du rôle de l'adulte, qui, en guidant l'activité et en incitant les enfants à s'exprimer, est amené à parler plus, et en utilisant un lexique plus varié.

Toutefois, nos observations doivent être relativisées. En effet, l'activité présentée ici n'est que l'une de celles qui sont menées à l'école. Il serait intéressant de comparer l'étendue et la variété du lexique utilisé entre cette activité et d'autres, centrées sur le langage (telles que la lecture de livres par l'enseignant, suivie par la restitution du contenu par les élèves, comme dans le corpus CLEA), ou non (jeux libres entre enfants, par exemple).

3.2.2 Synthèse des productions globales

Observons à présent les verbes utilisés dans l'ensemble de notre corpus. Les données présentées dans le Tableau 4 correspondent aux pourcentages d'utilisation (nombre d'occurrences pour chaque item produit sur nombre total de mots produits par locuteurs enfants et adultes).

Tableau 4. Pourcentages de productions verbales chez l'adultes et chez l'enfants

	% productions	
	Enfants	Adultes
être	2,23	2,03
avoir	1,19	0,84
aller	0,49	0,26

On observe que les verbes les plus produits chez l'adulte et chez l'enfants sont *être* (dans sa forme *c'est*), *avoir* (dans la forme *a*) et *aller* (dans les formes *allez* et *vais*). Ceci s'explique aisément, dans la mesure où *être* apparaît dans le présentatif *c'est*, dont l'emploi est également notable dans l'extrait présenté en 3.2.1. Il

en est de même pour *aller*. Concernant *avoir*, son emploi résulte vraisemblablement de son utilisation en tant qu'auxiliaire. Les verbes lexicaux n'apparaissent pas parmi les items verbaux les plus employés.

Pour les noms, nous n'avons pas repéré d'items ayant un nombre d'occurrences plus important que les autres. Ceci est cohérent avec les données du Tableau 3, dans lequel nous avons remarqué que 4 noms sur les 7 employés sont en relation avec la thématique que l'activité. Ainsi, les productions et le nombre d'occurrences dépendent beaucoup de l'activité enregistrée et de l'exercice proposé par l'adulte. Si chaque activité demande un lexique nominal spécifique, il est attendu qu'aucune constante n'émerge dans les productions de noms, mis à part peut-être les noms employés par l'adulte dans les consignes (cf. *image*, *yeux*, voir 3.2.1.).

Conclusion et perspectives

Notre travail, qui repose sur des enregistrements écologiques en milieu scolaire, ne nous a fourni que peu de données sur le lexique utilisé par les enfants. Comme nous l'avons vu, la majorité du temps de parole et du lexique émane des adultes. Nous n'avons donc que peu de données émanant des enfants en tant que groupe, et encore moins lorsqu'ils sont pris individuellement. Il est probable que la taille de notre corpus entraîne ce manque de données et ces difficultés de généralisation. L'analyse d'un corpus plus grand, comme celui récolté dans le cadre du projet ANR DyLNet (Nardy et al. 2016) dans plusieurs écoles maternelles pourra potentiellement enrichir cette réflexion. Il est possible que l'activité enregistrée et étudiée soit en fait peu propice à la production. Son but étant seulement la dénomination d'objets, elle peut sembler artificielle et décontextualisée aux enfants. Par ailleurs, les activités proposées par l'enseignant peuvent pousser ou limiter certaines productions. Hindman et al. (2021) montrent que dans la cadre de la lecture d'un livre en milieu scolaire, certains enseignants posent plutôt des questions fermées, qui réduisent la production des enfants et/ou la limitent, alors que d'autres préfèrent créer un espace d'échange et impulser des commentaires, ce qui favorise une production plus riche et plus variée.

Si tel est le cas, le dépouillement des enregistrements d'autres activités, comme la lecture de livres par l'enseignant, suivie de la restitution par les enfants de l'histoire racontée, nous permettra sans doute d'enrichir nos données. De plus, une telle activité, qui tend à être aussi pratiquée en milieu familial, pourra sembler plus familière, et donc plus naturelle, à l'enfant.

Le faible nombre de données émanant des enfants nous amène également à nous interroger sur l'intérêt de la méthode que nous avons employée, celle des enregistrements écologiques. Pour favoriser le recueil du lexique produit par les enfants, il faut favoriser sa production. Ainsi, Palasis (2010) a mené des activités en petits groupes, comme des séances au cours desquelles les enfants racontent des histoires, ou leurs activités à l'école et en dehors, ou encore participent collectivement à des jeux. Nous envisageons d'adopter à notre tour de telles méthodes.

Une autre question soulevée est d'ordre plus qualitatif. L'objectif de notre projet était le recueil du lexique jugé nécessaire aux enfants par les enseignants d'école maternelle de REP. Or, l'extrait d'enregistrement présenté plus haut montre que, si l'on exclut les consignes émanant de l'enseignant, le lexique est spécifique à l'activité dans laquelle les enfants sont engagés. Il est donc pertinent d'interroger les enseignants, *via* des questionnaires par exemple, sur le lexique qu'ils jugent le plus utile aux enfants sur le modèle de Bassano (2020) présenté dans la Section 1.2. Un questionnaire à destination des enseignants est en cours d'élaboration.

Dans le même ordre d'idées, on peut soulever la question de la variation lexicale induite par les activités réalisées en milieu scolaire, et par celles qui ont lieu à la maison (soins corporels, repas, coucher, etc). Il est probable que le lexique de l'école et celui de la maison sont différents. Dans ce cas, quel est le plus pertinent pour l'enfant ? S'agit-il du lexique spécifiquement scolaire — s'il existe —, du lexique de la maison, ou des deux ? Pour les enfants ayant peu d'interactions en milieu familial, on peut supposer que le lexique à développer en priorité est celui qui est commun à l'école et à la maison. L'un de nos objectifs à court terme est la comparaison des items du DLFP avec ceux que le questionnaire mentionné ci-dessus nous fournira.

Vis-à-vis de notre projet, le milieu familial présente un autre intérêt, celui de mettre généralement en relation un ou deux adulte(s) et un enfant. De ce fait, il est possible que les productions de l'enfant soient plus nombreuses, puisque ce contexte met en relation moins de participants qu'une classe. On peut s'attendre de la part de l'enfant à des productions à la fois plus nombreuses et plus variées, dans un contexte naturel. Un premier dépouillement du lexique employé dans les enregistrements réalisés dans le cadre de CoLaJe et une comparaison avec notre corpus JEMS nous permettra de vérifier cette hypothèse.

Références bibliographiques

- Bassano, D. (2005). Production naturelle précoce et acquisition du langage : l'exemple du développement des noms. *LIDIL 31: Corpus oraux et diversité des approches* : 61-84.
- Bassano, D. (2007). Emergence et développement du langage : enjeux et apports des nouvelles approches fonctionnalistes. *L'acquisition du langage et ses troubles*, E. Demont & MN. Metz-Lutz (eds.) :13-46. Marseille : SOLAL Editeurs.
- Bassano, D. Labrell, F. & Bonnet, P. (2020). *Évaluer les débuts du langage avec le DLFP. Lexique, grammaire et pragmatique*. <https://halshs.archivesouvertes.fr/halshs-02436994>
- Bornstein M.H., Tamis-LeMonda C.S., Haynes, O.M. (1999). First words in the second year: Continuity, stability, and models of concurrent and predictive correspondence in vocabulary and verbal responsiveness across age and context. *Infant Behavior & Development* 22 : 65-85.
- Buson, L & Nardy, A. (2020). Le(s) français scolaire(s) à l'école maternelle. Pratiques langagières d'enseignant.e.s et d'élèves. *Le Français Aujourd'hui*, 2020, Les pratiques langagières « ordinaires » des élèves, 208 : 93-103.
- Canut, E., Espinosa N., Vertalier M. (2013). Corpus et prise de conscience des processus interactionnels d'apprentissage du langage pour repenser les pratiques enseignantes en maternelle. *Linx* 68-69 : 68-93.
- Coquet, F., Roustit, J. & Ferrand, P. (2009). *EVALO 2-6*. Isbergues : Ortho Édition.
- Croteau, B., Helloin, M.-C. & Thibault, M.-P. (2010). *EXALANG 5-8 : Batterie informatisée d'examen du langage oral et écrit*. Mont-Saint-Aignan : Orthomotus.
- Da Silva-Genest, C. Liégeois, L., Masson, C. Benzitoun, C. & Le Mené, M. (2020). Rabot, sécateur, téléphérique : interroger la pertinence des choix lexicaux dans les outils d'évaluation du langage en orthophonie. *Lidil* 62 :1-19.
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego : Singular Publishing Group.
- Hedlund, G. & Rose, Y. (2020). Phon 3.1 [Computer Software]. Consultable à l'adresse <https://phon.ca>
- Hindman, A.H., Farrow, J., Anderson, K., Wasik, B.A., & Snyder, P.A. (2021). *Understanding Child-Directed Speech Around Book Reading in Toddler Classrooms: Evidence From Early Head Start Programs*. *Frontiers in Psychology*, 12.
- Kidd, L., & Rowland, C. F. (2021). The effect of language-focused professional development on the knowledge and behaviour of preschool practitioners. *Journal of Early Childhood Literacy*, 21-1, 27-59
- Le Normand M-T, Parisse C, Cohen H. (2008). Lexical diversity and productivity in French preschoolers: Developmental, gender, and sociocultural factors. *Clinical Linguistics & Phonetics* 22 : 47-58.
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods* 36 (1) : 156-166.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition*. Mahwah, NJ : Lawrence Erlbaum Associates.
- Morgenstern, A. & Parisse, C. (2012). The Paris Corpus, *Journal of French Language Studies* 22-1, First language acquisition of French grammar : 7-12.

- Nardy, A., Fleury, E., Chevrot, J.-P., Karsai, M., Buson, L., Bianco, M., Rousset, I., Dugua, C., Liégeois, L., Barbu, S., Crespelle, C., Busson, A. Léo, Y., Bouchet, H. (2016-2020). *DyLNet – Language Dynamics, Linguistic Learning, and Sociability at Preschool: Benefits of Wireless Proximity Sensors in Collecting Big Data* (ANR-16-CE28-0013-0).
- New, B., Pallier, C., Ferrand L. & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'Année Psychologique*, 10 : 447-462.
- Palasis, K. (2010). Introducing New French Child Data: Thoughts on their Gathering and Coding. *Corpus 9 La syntaxe de corpus* : 33-51.
- Péroz, P. (2010). Apprentissage du langage oral à l'école maternelle. Pour une pédagogie de l'écoute. Nancy, Scéren/CRDP Lorraine.
- Péroz, P. (2016). Apprentissage du langage oral à l'école maternelle. Mise en œuvre d'un modèle polylogal d'apprentissage du langage oral en moyenne et grande section de maternelle en 2010-2013. *Pratiques*, 169-170. <https://doi.org/10.4000/pratiques.3100>
- Raneri, D., Von Holzen, K., Newman, R., & Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language* 47 :1263-1275.

¹ Une version transcrite du Corpus CLEA est accessible en ligne dans *Pratiques* : <https://journals.openedition.org/pratiques/3919>

² https://eduscol.education.fr/83/j-enseigne-au-cycle-1?menu_id=68

³ D'autres écoles maternelles sont également incluses dans le projet, mais les enregistrements réalisés n'ont pas encore été transcrits.

⁴ Abréviations : Bru : École de Bruyères ; Ha : École de Hadol ; JJ : École Jean Jaurès