

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDE DU DÉVELOPPEMENT DU NOM DANS LES PRODUCTIONS D'ENFANTS ÂGÉS DE DEUX ANS
À CINQ ANS

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE DE LINGUISTIQUE

PAR

MANON BROUSSE

AVRIL 2025

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude à toutes les personnes qui m'ont soutenue tout au long de ce parcours de recherche et d'écriture.

Tout d'abord, je voudrais remercier chaleureusement ma direction de recherche, Marine Le Mene Guigoures et Denis Foucambert, pour leurs expertises, leurs patiences et leurs précieux conseils tout au long de ce projet.

Mes plus sincères remerciements vont également à mes lecteurs, Caroline Masson et Grégoire Winterstein, pour leur temps, leur expertise et leurs suggestions constructives qui ont grandement enrichi ce travail.

À mon partenaire de vie, Paul, je suis infiniment reconnaissante pour ton soutien inconditionnel et ta compréhension tout au long de cette période exigeante.

Je souhaite également adresser mes remerciements les plus sincères à ma famille en France, en particulier mes parents et ma sœur, pour leur soutien indéfectible et leurs encouragements constants.

Enfin un grand merci à mes amis, qui ont été présents à mes côtés à chaque étape de ce projet. Vos encouragements, vos discussions stimulantes et votre soutien moral ont été d'une valeur inestimable.

À toutes ces personnes, je vous suis profondément reconnaissante pour votre contribution à la réalisation de ce mémoire. Vos encouragements et votre soutien ont été des sources d'inspiration et de motivation tout au long de ce parcours.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
LISTE DES FIGURES.....	vi
LISTE DES TABLEAUX	vii
RÉSUMÉ.....	viii
INTRODUCTION	1
CHAPITRE 1 Revue de littérature.....	3
1.1 Les différentes approches théoriques en acquisition	3
1.1.1 Les théories innéistes.....	3
1.1.1.1 La notion centrale de pauvreté du stimulus.....	4
1.1.1.2 Grammaire universelle et dispositif spécifique d'acquisition.....	4
1.1.2 Les approches basées sur l'usage	5
1.1.2.1 Les fondements de cette approche	6
1.1.2.2 L'acquisition du langage précoce : des « îlots verbaux » à la compétence syntaxique	7
1.2 Le développement lexical	8
1.2.1 Les rythmes d'acquisition	8
1.2.2 La structuration du premier lexique	10
1.2.3 La variation individuelle	12
1.2.4 Le rôle de l'adulte dans le processus d'acquisition	14
1.3 Le développement du nom	16
1.3.1 Les caractéristiques sémantiques du nom.....	16
1.3.2 La longueur du nom	18
1.3.3 La fréquence du nom	19
1.3.4 La valence émotionnelle	20
1.3.5 Le niveau de développement langagier.....	21
1.4 Question de recherche.....	22
CHAPITRE 2 MÉTHODE	23
2.1 Corpus et participants.....	23
2.1.1 La norme utilisée pour l'étude.....	28
2.2 Les variables et l'analyse du corpus	31
2.2.1 Les caractéristiques sémantiques du nom.....	34
2.2.2 La longueur du nom	36
2.2.3 La fréquence	37
2.2.4 La valence émotionnelle	37
2.2.5 Le nombre d'occurrences produites par les adultes	38
2.2.6 Le niveau de développement langagier.....	38

2.2.7	Le genre.....	42
2.2.8	Le modèle mixte généralisé de type « Poisson »	42
CHAPITRE 3 RÉSULTATS.....		44
3.1	Les analyses descriptives.....	44
3.1.1	Les caractéristiques sémantiques du nom.....	47
3.1.1.1	Le caractère animé/inanimé des référents associés aux noms	47
3.1.1.2	La concrétude	48
3.1.2	La longueur du nom	49
3.1.3	La valence émotionnelle	50
3.1.4	Le niveau de développement langagier	52
3.2	Les analyses du modèle sans interactions	53
3.2.1	L'effet de la variable <i>occurrences adultes</i>	54
3.2.2	L'effet du niveau de développement langagier (groupe MLU).....	55
3.2.3	L'effet du caractère animé/inanimé du référent associé au nom	56
3.2.4	L'effet des variables aléatoires	57
3.3	L'analyse du modèle avec les interactions.....	58
3.3.1	L'interaction entre le niveau de développement langagier (groupe MLU) et le caractère animé/inanimé des noms.....	59
3.3.2	L'interaction entre le niveau de développement langagier (groupe MLU) et la longueur des noms 61	
3.3.3	L'effet des variables aléatoires	64
CHAPITRE 4 DISCUSSION		65
4.1	Les variables qui influencent l'acquisition du nom chez l'enfant	65
4.1.1	L'effet de la variable <i>occurrences adultes</i> sur le nombre d'apparitions du nom chez l'enfant...66	
4.1.2	L'effet de la variable <i>niveau de développement langagier</i> sur le nombre d'apparitions du nom chez l'enfant.....	69
4.1.3	L'effet de la variable <i>caractère animé/inanimé du nom</i> — sur le nombre d'apparitions du nom chez l'enfant.....	71
4.2	Les variables non pertinentes dans le cadre de notre étude.....	72
4.3	Les variables qui influencent l'acquisition du nom chez l'enfant : interactions entre le niveau de développement langagier, le caractère animé/inanimé et la longueur du nom	74
4.3.1	L'effet de l'interaction entre le niveau de développement langagier et le caractère animé/inanimé du référent associé au nom.....	75
4.3.2	L'effet de l'interaction entre le niveau de développement langagier et la longueur du nom	76
4.4	Les variables aléatoires	77
4.5	Les limites de l'étude	79
CONCLUSION		81
ANNEXE A : CORPUS.....		83
ANNEXE B : RÉSULTATS COMPLETS DES ANALYSES SANS INTERACTIONS		89

ANNEXE C : RÉSULTATS COMPLETS DES ANALYSES AVEC INTERACTIONS92

RÉFÉRENCES95

LISTE DES FIGURES

Figure 3-1 Répartition des occurrences des noms en fonction du caractère animé ou inanimé des référents	47
Figure 3-2. Répartition des occurrences des noms en fonction de la concrétude	48
Figure 3-3. Répartition des occurrences des noms selon leur valence émotionnelle	51
Figure 3-4. Nombre d'apparitions en fonction des occurrences adultes.....	55
Figure 3-5. Nombre d'apparitions en fonction du niveau de développement langagier (groupe MLU)	56
Figure 3-6. Moyenne du nombre d'apparitions en fonction du caractère animé/inanimé.....	57
Figure 3-7. Moyenne du nombre d'apparitions en fonction du niveau de développement langagier (groupe MLU) et du caractère animé/inanimé	60
Figure 3-8. Nombre d'apparitions du nom en fonction de la longueur du nom.....	61
Figure 3-9. Moyenne du nombre d'apparitions en fonction du niveau de développement langagier (groupe MLU) et de la longueur du nom	62

LISTE DES TABLEAUX

Tableau 2.1 Corpus de CHILDES sélectionnés pour notre étude	24
Tableau 2.2 Âge des participants enfants	25
Tableau 2.3 Âge des enfants par séance	26
Tableau 2.4 Proportion de noms à fort accord et faible accord selon l'âge et le genre (Syssau et Monnier, 2009)	30
Tableau 2.5 Exemples d'attribution de valence pour six noms	33
Tableau 2.6 Répartition des noms animés et des noms inanimés	35
Tableau 2.7 Répartition des noms concrets et des noms abstraits	36
Tableau 2.8 Répartition des noms selon leur longueur	36
Tableau 2.9 Répartition des noms selon leur valence émotionnelle	38
Tableau 2.10 Production des noms selon le groupe MLU	39
Tableau 2.11 Répartition des séances des enfants selon leur groupe MLU	40
Tableau 3.1 Répartition du nombre d'occurrences et d'énoncés produits par enfant	45
Tableau 3.2 Répartition du nombre d'occurrences et d'énoncés produits par groupe MLU	46
Tableau 3.3 Répartition des occurrences selon leur longueur (en nombre de syllabes)	49
Tableau 3.4 Répartition des occurrences de noms selon la longueur moyenne des énoncés	52
Tableau 3.5 Distribution des variables dans le modèle sans interaction	53
Tableau 3.6 Distribution des variables aléatoires dans le modèle sans interaction	58
Tableau 3.7 Distribution des variables dans le modèle avec les interactions	59
Tableau 3.8 Relations entre le niveau de développement langagier (groupe MLU) et la longueur des noms	63
Tableau 3.9 Distribution des variables aléatoires dans le modèle avec les interactions	64

RÉSUMÉ

Ce mémoire a pour objectif d'évaluer l'influence de différentes caractéristiques des noms sur leur production (ou leur usage) chez des jeunes enfants âgés de deux à cinq ans. La présente étude analyse l'impact de divers facteurs sur la production des noms, tels que les caractéristiques sémantiques du nom (le caractère animé / inanimé et concret /abstrait du nom), la valence émotionnelle qui y est associée, sa longueur, sa fréquence, ainsi que le nombre d'occurrences du nom produites par les adultes interagissant avec les enfants. Elle prend également en compte le niveau de développement langagier et le genre des enfants. Cette étude se base sur dix suivis longitudinaux de cinq filles et cinq garçons, âgés de deux à cinq ans. Les noms étudiés dans notre travail sont présents à la fois dans les séances sélectionnées du corpus CHILDES, qui documente les productions langagières des enfants, et dans la norme ValEmo, qui renseigne sur leur valence émotionnelle. Cette sélection croisée permet d'analyser l'impact des caractéristiques émotionnelles des noms sur leur production ou leur usage chez les jeunes enfants. Les résultats montrent que certaines variables influencent significativement la production des noms par les enfants : le nombre d'occurrences du nom produites par les adultes, le niveau de développement langagier, le caractère animé/inanimé et la longueur du nom. Concernant la variable *niveau de développement langagier*, notre étude montre que les enfants ayant un développement plus avancé produisent davantage de noms, y compris des noms plus longs. Quant à la variable *caractère animé/inanimé* du nom, les noms animés prédominent dans le lexique du jeune enfant, mais une augmentation des noms inanimés est observée avec l'âge. Cependant, les résultats révèlent que les noms animés apparaissent plus fréquemment que les noms inanimés, même à un âge plus avancé. Enfin, pour la variable *longueur du nom*, les noms monosyllabiques dominent dans les premières étapes du développement, et les enfants produisent progressivement des noms plus longs en grandissant.

Mots clés : acquisition du langage, production des noms, caractéristiques sémantiques, longueur du nom, valence émotionnelle, fréquence, développement langagier, suivi longitudinal

INTRODUCTION

L'acquisition du langage chez l'enfant, en particulier l'acquisition des noms, constitue un domaine d'étude central en linguistique développementale. Les recherches scientifiques se penchent, depuis plusieurs décennies, sur la manière dont les enfants acquièrent leurs premières compétences langagières, et en particulier le lexique. Comprendre comment l'enfant passe de productions courtes et peu diversifiées à une maîtrise plus complexe du langage implique d'étudier l'influence — le plus souvent conjointe — d'une multitude de facteurs, tant cognitifs qu'environnementaux sur le développement langagier. Ce mémoire s'inscrit dans cette lignée de recherches où plusieurs facteurs sont étudiés pour comprendre leur influence sur le développement lexical des jeunes enfants. L'objectif principal de ce mémoire est d'examiner comment ces différentes variables influencent la production et l'usage des noms chez un groupe d'enfants âgés de deux à cinq ans.

Les théories cherchant à expliquer l'acquisition du langage peuvent être compilées en deux grandes approches : les théories innéistes et les théories non innéistes, le plus souvent aujourd'hui basées sur l'usage (usage-based). Selon l'approche innéiste, l'acquisition du langage repose sur des capacités cognitives innées, spécifiques au langage (Chomsky, 1996). En opposition, les théories basées sur l'usage avancent que l'acquisition du langage découle principalement de l'expérience et de l'interaction avec l'environnement social. Ces dernières approches mettent en avant le rôle de l'usage quotidien du langage dans la structuration des compétences langagières de l'enfant (Tomasello, 2003). C'est dans ce cadre théorique que ce mémoire propose d'examiner l'acquisition des noms, en tenant compte donc au moins en partie du rôle des productions adultes sur les formes produites par l'enfant.

Les noms (ou plus exactement les « candidats à noms »), souvent prédominants dans le premier lexique enfantin, jouent un rôle crucial dans le développement linguistique des enfants (Tardif, 1996). Cependant, il est difficile de les considérer comme les premières unités lexicales acquises, car la nature des premiers mots est bien plus diverse (Bloom, 2000). De plus, la prédominance des noms dans le lexique initial peut varier selon les styles d'acquisition et les langues, rendant difficile une catégorisation stricte selon les classes lexicales adultes (Gentner, 1982). L'objectif de ce mémoire est d'évaluer l'influence de certains facteurs sur la production (ou l'usage) des noms chez les jeunes enfants. La littérature sur ce sujet a montré que l'acquisition des noms est influencé par plusieurs facteurs, notamment des caractéristiques sémantiques, la fréquence d'utilisation ainsi que des dimensions spécifiques des interactions sociales,

telles que la quantité et la qualité du discours adressé à l'enfant et le contexte de jeu ou de routine dans lequel les noms sont présentés (Bates et al., 1994).

Ce travail s'appuie sur des données extraites de la base CHILDES (Child Language Data Exchange System), une base de données rassemblant des corpus (transcriptions et enregistrements associés) d'interactions le plus souvent spontanées entre des enfants et des adultes (MacWhinney, 2000). Dix suivis longitudinaux ont été extraits (cinq filles et cinq garçons) et le travail a porté sur les noms en commun entre les séances sélectionnées et ceux de la norme ValEmo. Pendant les séances retenues pour l'analyse, les enfants et les adultes sont engagés dans des interactions naturelles, permettant de recueillir un large éventail de productions langagières. Cette diversité de données permet d'observer comment les enfants utilisent et acquièrent les noms dans des contextes variés, tout en capturant la variabilité dans leurs productions langagières.

La première partie de ce mémoire est dédiée à la description du cadre théorique. Nous y aborderons, dans un premier temps, les différentes théories relatives à l'acquisition du langage. Les sections qui suivent aborderont tout d'abord les rythmes d'acquisition du lexique chez l'enfant. Ensuite, nous nous intéresserons à la structuration du premier lexique, en analysant les principes qui organisent les premières productions de l'enfant. La variation individuelle sera également explorée, en mettant en lumière les différences entre enfants au cours du processus d'acquisition. Nous discuterons par ailleurs du rôle de l'adulte dans ce processus. Enfin, nous nous concentrerons sur le développement du nom, en décrivant et expliquant les huit variables caractérisant le nom, pris en compte dans notre travail.

La deuxième partie décrit la méthode adoptée dans notre étude. Nous présentons le corpus et les participants, les activités réalisées par les enfants et la norme utilisée dans l'étude (ValEmo). Cette partie présente également les variables considérées pour l'étude et la méthode d'analyse du corpus.

Les résultats obtenus sont rapportés dans la troisième partie. Celle-ci est suivie par une discussion des résultats dans la quatrième partie. Nous tentons dans cette partie de faire des liens avec les théories et études abordées dans la première partie et les résultats de l'étude. Enfin, nous présentons les limites de l'étude.

CHAPITRE 1

REVUE DE LITTÉRATURE

La littérature sur l'acquisition du langage est vaste et présente une diversité de points de vue. Différents courants tentent d'expliquer les processus d'acquisition du langage chez l'enfant. Cependant, dans ce mémoire, nous nous limiterons aux approches majoritaires. Nous définirons les éléments principaux des théories innéistes et des théories non innéistes, notamment l'approche basée sur l'usage (usage-based).

Bien que ce mémoire n'ait pas pour vocation de proposer des réflexions théoriques approfondies, il nous semble nécessaire d'introduire ce chapitre de littérature par quelques repères généraux sur les théories de l'acquisition, qui conditionnent le traitement de l'acquisition du lexique dans la littérature. La section suivante (1.2) vise à présenter les principaux travaux portant sur l'acquisition des noms chez l'enfant, afin de situer notre recherche dans son contexte scientifique. Elle permettra également de justifier certains choix méthodologiques opérés dans le cadre de cette étude.

1.1 Les différentes approches théoriques en acquisition

1.1.1 Les théories innéistes

Les tenants de la théorie innéiste prônent l'idée selon laquelle l'acquisition du langage serait une capacité innée de l'être humain, reposant sur des capacités linguistiques universelles. L'être humain serait doté d'un module cognitif autonome dédié à l'acquisition du langage (Chomsky, 1996).

1.1.1.1 La notion centrale de pauvreté du stimulus

La pauvreté du stimulus est l'un des arguments récurrents et principaux de l'approche innéiste pour défendre la thèse générale du caractère inné du langage. Cet argument repose sur deux éléments : d'une part, l'idée que la connaissance linguistique acquise par l'enfant apprenant le langage est sous-déterminée par les données dont l'enfant dispose ; et d'autre part, l'idée d'une pauvreté des données linguistiques de son environnement (Chomsky, 1980). L'expérience linguistique de l'enfant serait trop limitée pour qu'il puisse apprendre de l'input de son entourage, et par ailleurs, les données linguistiques de l'environnement ne constitueraient pas de bons modèles et n'offriraient aucun accompagnement métalinguistique, ce qui justifierait que la grammaire de l'enfant soit nécessairement composée de structures innées (Blitman, 2015, p.83-113 ; Chomsky, 1959 ; 1965, 1980).

1.1.1.2 Grammaire universelle et dispositif spécifique d'acquisition

Outre l'argument de la pauvreté du stimulus, Chomsky met en avant dans ses travaux la notion de grammaire universelle. Pour lui, le système cognitif de l'enfant est muni d'un dispositif d'acquisition du langage, basé sur une grammaire universelle. La grammaire universelle se définit comme « le système des principes, des conditions et des règles qui sont des éléments ou des propriétés de toutes les langues humaines, pas simplement par accident, mais par nécessité — nécessité biologique et non logique » (Blitman, 2015, p.29-59). Autrement dit, la grammaire universelle est composée d'un ensemble de principes communs à toutes les langues humaines et d'une base innée permettant le développement de l'acquisition du langage chez l'enfant. Pour Chomsky, les structures linguistiques sont à considérer comme autonomes, non influencées par l'usage (Chomsky, 1991). Ainsi, il considère que le langage est une capacité innée reposant sur une grammaire universelle, elle-même caractérisée comme le « mécanisme d'apprentissage du langage ».

Dans ses travaux, Chomsky postule l'existence d'un dispositif d'acquisition du langage nommé le Language Acquisition Device (LAD). Il s'agit d'un ensemble de mécanismes permettant à l'enfant de découvrir les propriétés formelles de sa langue (Chomsky, 1965, 1971). Au moment de l'acquisition de sa langue première, l'enfant se comporte comme un linguiste face à une langue étrangère, c'est-à-dire qu'il extrait la grammaire de la langue cible tout en se basant sur les données linguistiques disponibles autour de lui. Au cours de l'apprentissage, l'enfant développe des règles grammaticales en fonction des données qu'il perçoit (McNeill, 1970). Autrement dit, il établit des règles concernant la structure de la langue, qu'il ajuste en fonction des informations qu'il rencontre, au sein des grammaires possibles définies par la grammaire universelle innée. Par la suite, l'enfant teste ces règles en les confrontant aux nouvelles données linguistiques, les affinant et les révisant progressivement jusqu'à atteindre une maîtrise adulte de sa langue. Ainsi, c'est la grammaire universelle et le LAD qui guident l'enfant dans l'apprentissage de sa langue première.

Bien que cette conception de l'acquisition puisse encore s'observer dans la littérature actuelle, elle tend à devenir minoritaire et à faire place à des approches dites basées sur l'usage, pour lesquelles d'autres facteurs, tels que l'interaction, l'entourage ou encore les variations individuelles sont à prendre en compte pour comprendre le développement langagier de l'enfant.

1.1.2 Les approches basées sur l'usage

Les linguistes adoptant une approche « usage-based » de l'acquisition soutiennent que les constructions grammaticales se développent à partir du sens et émergent à travers les usages que l'enfant fait du langage. Ainsi, la structure linguistique se construit progressivement à partir des interactions et des contextes dans lesquels le langage est utilisé.

1.1.2.1 Les fondements de cette approche

L'approche basée sur l'usage repose sur deux principes fondamentaux : le sens découle de l'usage et la structure émerge de l'usage. Le premier principe concerne la dimension fonctionnelle et sémantique de la communication linguistique, tandis que le second fait référence à la dimension structurelle et grammaticale. Tomasello (2003) a développé une théorie de l'acquisition basée sur l'usage. Selon lui, l'enfant commence le processus d'acquisition du langage avec deux ensembles de compétences cognitives essentielles : la lecture des intentions et la recherche de motifs.

La lecture des intentions, liée à la dimension fonctionnelle, permet à l'enfant de comprendre les objectifs ou intentions des locuteurs adultes lorsqu'ils utilisent des conventions linguistiques dans des contextes sociaux. Cette compétence inclut des capacités telles que l'attention conjointe, qui sont cruciales pour l'apprentissage des conventions linguistiques culturelles et sociales (Tomasello, 2009). La lecture des intentions est un élément clé de la théorie de l'acquisition sociale, qui postule que les enfants apprennent le langage en observant et en interprétant les intentions des autres (Bates et al., 1979).

La recherche de motifs, qui relève de la dimension grammaticale, permet à l'enfant de dépasser les énoncés individuels qu'il entend des adultes et d'identifier des régularités dans les structures linguistiques. A partir de ces régularités, l'enfant crée des schémas ou des constructions abstraites. Ce processus implique des compétences cognitives telles que la catégorisation, l'analogie et l'analyse distributionnelle (Bybee, 2006). La recherche de motifs repose sur l'idée que la langue est apprise par l'observation des fréquences et des contextes dans lesquels des formes linguistiques particulières sont utilisées (Lieven, 2006).

Ainsi, les linguistes qui adoptent l'approche usage-based rejettent l'idée d'une grammaire entièrement formelle, dépourvue de sens. Ils se concentrent plutôt sur la manière dont les constructions grammaticales, basées sur le sens, émergent à travers l'usage et les productions des enfants (Tomasello, 2003).

1.1.2.2 L'acquisition du langage précoce : des « îlots verbaux » à la compétence syntaxique

Dans l'approche usage-based de l'acquisition du langage, le développement précoce de la compétence grammaticale ne repose pas sur l'existence d'une grammaire abstraite et cohérente, mais plutôt sur l'extraction de constructions et de schémas linguistiques à partir de l'input. Les enfants commencent par acquérir des constructions simples, souvent associées à des verbes spécifiques, qu'ils apprennent dans des contextes significatifs. Ces constructions sont des « îlots verbaux », où chaque construction est semi-structurée et fondée sur une combinaison d'éléments lexicaux et d'expérience contextuelle (Tomasello, 2003).

Tomasello (2009) explique que, loin de concevoir l'acquisition grammaticale comme un processus de mise en relation de mots isolés, ce processus est basé sur la compréhension et la production d'énoncés complets. Les enfants, par la répétition, apprennent à reconnaître les relations syntaxiques et les rôles des constituants. Les premières constructions syntaxiques sont donc spécifiques aux verbes, et les enfants développent progressivement des structures plus complexes à mesure qu'ils rencontrent et utilisent un nombre croissant de verbes dans leurs interactions sociales (Lieven, 2006).

L'acquisition du lexique et de la syntaxe dans cette approche se fait de manière complémentaire : les enfants acquièrent d'abord des mots en contexte, leur attribuent une fonction communicative, puis ils réinvestissent ces mots dans des structures syntaxiques plus complexes (Bates et al., 1979). Cela montre que la compétence grammaticale émerge directement de l'usage et des constructions observées dans l'environnement social de l'enfant (Tomasello, 2009). Ainsi, l'acquisition précoce de la syntaxe repose sur l'interaction entre le lexique, les usages linguistiques et les modèles observés chez les adultes (Cameron-Faulkner et al., 2003).

1.2 Le développement lexical

C'est vers la fin de la première année que l'enfant commence à produire des éléments vocaux qui sont généralement assimilés à des mots de la langue de son environnement (Bassano, 2000). L'émergence du langage oral apparaît grâce à l'élaboration d'un stock de mots dans le système linguistique de l'enfant, c'est ce que l'on appelle le lexique. La deuxième et troisième année de vie de l'enfant sont particulièrement cruciales pour la constitution de ce lexique.

1.2.1 Les rythmes d'acquisition

L'évolution du lexique précoce se manifeste par une augmentation du nombre de mots produits par l'enfant. En moyenne, à l'âge de 15 mois les enfants produisent 10 mots, à 20 mois 50 mots et à 24 mois 200 mots (Kern, 2019 ; Nelson, 1973, 1975). L'augmentation du lexique de production est caractérisée par des progrès non linéaires, bien identifiables et localisés dans le temps (Bassano, 2000). La production des premiers mots appelés « conventionnels » fait son apparition aux alentours de 12 mois. La progression des premiers mots est très lente et c'est au milieu de la deuxième année que l'accroissement des mots devient important, il s'agit de « l'explosion lexicale », un brusque accroissement se produisant le plus souvent autour de 18-20 mois (Bates et al., 1994 ; Bloom, 1973). Au moment où l'enfant atteint un stock d'environ 50 mots, le rythme d'acquisition et d'apprentissage s'accélère et l'enfant peut produire entre 4 et 10 nouveaux mots par jour (Bassano, 2000 ; Kern, 2019). C'est durant cette période — et plus encore, à la suite de cette période, qu'émergent les premières combinaisons de mots permettant à l'enfant de produire des énoncés à plusieurs termes (Martel et Dodane, 2017).

Plusieurs explications sont données dans la littérature pour expliquer le phénomène d'explosion lexicale, reposant à la fois sur des facteurs endogènes et exogènes. Les explications liées à des facteurs endogènes mettent en évidence que l'explosion lexicale est liée aux changements dans le développement cognitif de

l'enfant alors que, pour les facteurs exogènes, l'explosion lexicale est liée à l'exposition à la langue de l'enfant (Kern, 2019).

Les facteurs endogènes peuvent être décrits de plusieurs manières. Dans un premier temps, l'enfant améliore sa capacité à catégoriser non verbalement les objets du monde (Kern, 2019). Dans un second temps, il a été mis en évidence qu'en grandissant l'enfant prend conscience que tout objet se trouvant dans son monde porte une étiquette, il s'agit du phénomène de « naming insight ». Il réalise également que toutes les choses du monde qu'il a nommées appartiennent à des catégories (Reznick et al., 1992). Par ailleurs, au cours de la deuxième année de l'enfant, on observe un lien temporel entre l'explosion lexicale et le développement de compétences cognitives de l'enfant (Florin, 2010). C'est à partir de ce moment que l'enfant, de manière implicite et inconsciente, s'appuie sur plusieurs principes cognitifs pour acquérir de nouveaux mots. Ces principes, décrits par des chercheurs comme Clark (1993), MacWhinney (2004), et Tomasello (2003), permettent à l'enfant d'interpréter et de retenir des significations lexicales, même dans des contextes où la référence de l'objet ou du concept n'est pas clairement établie. On distingue quatre principes principaux : le principe de dénomination catégorielle, le principe de conventionnalité, le principe d'exclusivité mutuelle et le principe de contraste. Le principe de dénomination catégorielle stipule que l'enfant attribue un nouveau mot à une catégorie d'objets correspondant au niveau de base auquel l'objet appartient, en fonction de ses propriétés communes. Par exemple, lorsqu'un enfant observe un chien, un chat ou bien un cheval, il peut les désigner par le terme générique « animaux », car ce terme regroupe les caractéristiques les plus représentatives des objets de cette catégorie (Clark, 1993). Le principe de conventionnalité renvoie à l'idée que l'enfant apprend que les formes linguistiques sont socialement convenues et qu'elles doivent correspondre à celles utilisées par les adultes dans son environnement. Cette recherche de la « forme correcte » est un processus implicite mais essentiel dans l'acquisition du langage, car elle guide l'enfant vers les mots et les constructions grammaticales de sa communauté linguistique (Tomasello, 2003). Le principe d'exclusivité mutuelle désigne la capacité de l'enfant à faire l'hypothèse qu'un nouveau mot correspond à un objet pour lequel il n'a pas encore de terme. En d'autres termes, si un enfant connaît déjà un mot pour un objet donné, il interprétera un nouveau mot comme désignant un objet différent, qu'il n'a pas encore nommé (Markman, 1990). Et enfin, le principe de contraste postule que l'enfant, lorsqu'il rencontre un nouveau mot, suppose qu'il doit nécessairement se distinguer d'un autre mot déjà connu, dans le but d'élargir son lexique et de mieux organiser les catégories lexicales. Par exemple, si l'enfant connaît le mot « chat », un nouveau mot comme « chaton » serait perçu comme un terme qui contraste avec celui déjà appris (Clark, 1993). Ces

principes ne sont pas consciemment appliqués par l'enfant, mais émergent à travers ses interactions avec son environnement linguistique. Ils sont des mécanismes inconscients qui, en guidant l'interprétation et l'attribution de sens aux nouveaux mots, permettent à l'enfant de constituer progressivement un lexique cohérent et fonctionnel (Bassano, 2000 ; Clark, 1993 ; Kern, 2019).

Comme évoqué précédemment, les explications liées à des facteurs exogènes soutiennent l'idée que les rythmes d'acquisition des mots dépendent de l'exposition de l'enfant à la langue de son environnement. Ces facteurs mettent en avant le phénomène de « leverage learning », autrement dit l'apprentissage par levier. Il implique que le début de l'acquisition est lent pour l'enfant. Chaque mot acquis fournit des informations aidant à l'acquisition des mots futurs (Mitchell et McMurray, 2009).

1.2.2 La structuration du premier lexique

De nombreux travaux se sont penchés sur la structuration du lexique, autrement dit sur la nature des mots que les jeunes enfants produisent, et plus précisément sur le développement des classes de mots (noms et verbes notamment) dans le langage de l'enfant. Ces études ont observé que très tôt, les enfants peuvent produire une grande variété d'éléments lexicaux (Bassano, 2000 ; Kern, 2019).

Parmi ces éléments, les premiers à être utilisés sont ceux permettant l'échange dans les situations de communication et d'interaction avec autrui comme l'adverbe *non*, les routines sociales comme *allo*, *au revoir*, les onomatopées, les mots relationnels, les interjections, les formules figées comme *ça y est* ou *a plus...* On les appelle les éléments sociopragmatiques, et ils sont considérés comme fondamentaux lors de l'émergence du langage. De par leur fonction pragmatique, ces éléments sont fréquemment produits durant l'acquisition du lexique précoce, néanmoins, ils restent difficiles à catégoriser au moyen des catégories lexicales adultes (Bassano, 2000). En revanche, d'autres types d'éléments au sein du premier lexique peuvent être plus facilement comparés aux catégories lexicales adultes. Ils sont scindés en deux catégories : les « termes à fonction référentielle » et les « termes à fonction prédicative ». Les termes à fonction « référentielle » renvoient aux objets qui sont inanimés et aux individus qui eux sont animés, par exemple les noms comme *biberon*, *chaussure* ou encore *maman*, *oiseau*. Les termes à fonction

« prédicative » réfèrent quant à eux aux actions avec les verbes, aux propriétés et qualités des objets ou des individus pour les adjectifs. On pourra donner comme exemples les verbes *vouloir* ou *manger* et les adjectifs *beau* ou *petit* fréquemment produits dans les premiers énoncés de l'enfant (Bassano, 2000 ; Bornstein et al., 2004 ; Kern et Gayraud, 2010). Les termes de ces deux catégories sont aussi appelés mots référentiels ou de contenu lexical.

Outre ces mots référentiels, le lexique du jeune enfant comprend également des mots fonctionnels (ou mots « grammaticaux »), tels que les déterminants, les pronoms, les prépositions ou les conjonctions. En ce qui concerne l'ordre de production, les mots référentiels sont généralement produits avant les mots fonctionnels (Brown, 1973 ; Kern 2019). Parmi les mots référentiels, les noms et plus particulièrement les noms d'objets sont produits avant les prédicats, notamment les verbes (Kern et Gayraud, 2010). L'étude de Parisse & Le Normand (2000) confirme ces constats. En s'appuyant sur un corpus de 27 enfants francophones âgés de 2 ans en situation de jeu symbolique, les auteurs ont montré que les enfants produisent un plus grand nombre de noms que de verbes. Cependant, l'ordre d'apparition des éléments dépend tout de même de la langue en acquisition. Prenons l'exemple du coréen, les enfants acquièrent en premier les verbes et par la suite les noms. Cela s'explique par les caractéristiques structurelles et sémantiques du coréen notamment parce que les verbes sont placés à la fin des énoncés et seraient donc plus saillants (Choi et Gopnik, 1995).

Parmi les nombreuses études menées sur le développement lexical de l'enfant, l'une d'entre elles a proposé un modèle en quatre stades permettant de généraliser les régularités du développement lexical (Bates et al., 1988, 1994 ; Caselli et al., 1995). Le stade des tout premiers mots est appelé le stade des éléments sociopragmatiques et ludiques. Le stade suivant se nomme le stade de la référence et compte un usage prédominant de noms communs ; c'est à cette période que l'enfant atteint un stock de 50 mots. Vient ensuite le stade de la prédication ; l'enfant dépasse le seuil des 100 mots et produit davantage de verbes et d'adjectifs encodant les significations relationnelles. Le dernier stade, celui de la grammaire, est caractérisé par un stock lexical de plus de 400 mots.

Malgré l'intérêt indubitable de ces recherches qui visent à généraliser les rythmes d'acquisition et la nature du premier lexique, il est important de souligner que le développement lexical ne peut être étudié sans considérer la question de la variation individuelle. C'est cet aspect que nous aborderons au sein de la section qui suit.

1.2.3 La variation individuelle

Les premiers mots n'apparaissent pas au même âge chez tous les enfants ni au même rythme. Avant 12 mois, la variabilité est peu visible, mais c'est après la première année que les profils des enfants se différencient. À l'âge de 24 mois, les enfants produisent en moyenne 300 mots, les 10 % les plus avancés produiront plus de 500 mots et les 10 % les moins avancés compteront eux moins de 100 mots dans leur lexique (Bassano, 2000).

Le modèle universel d'émergence des différentes classes lexicales a été remis en cause. Nelson (1973) a par exemple mis en évidence la variation « stylistique » relative à l'acquisition du lexique. Elle a distingué les enfants selon deux « styles » différents : le style « référentiel » et le style « expressif ». Un enfant avec un style « référentiel » acquiert dans un premier temps des mots faisant référence à des objets, et donc les noms. De ce fait, la fonction référentielle est mise en avant. Un enfant avec un style « expressif » a un discours plus varié c'est-à-dire qu'il exprime ses routines quotidiennes, des histoires, des récits. Pour lui, le langage a une fonction plus sociale et instrumentale. À la suite de Nelson, d'autres chercheurs ont proposé une analyse plus fine de ces différents styles d'acquisition lexicale. Afin d'expliquer au mieux ces différences, Bates et ses collègues (1988, 1994, 1995) ont établi deux « dimensions » : la dimension « analytique » et la dimension « holistique ». Les enfants s'insérant dans la dimension « analytique » de l'acquisition ont généralement un style « référentiel », c'est-à-dire que, lors de leur apprentissage, ils produisent des « unités » courtes et bien segmentées. À l'inverse, les enfants s'insérant dans la dimension « holistique » ont un style « expressif », ils produisent des séquences plus longues, mais leurs constructions sont plus approximatives.

Concernant l'organisation du lexique à 20 mois, trois profils lexicaux différents ont été identifiés dans la littérature. Il y a les enfants avec un profil référentiel, pour qui la production de noms est dominante. D'autres enfants possèdent un profil paralexical avec un grand nombre d'items paralexicaux tels que les interjections (*oh !, ah !, coucou !*), les prophrases *oui* et *non* et les formules syncrétiques (*s'il-te-plaît, ça-y-est*). Enfin, le troisième profil lexical concerne les enfants qui produisent en plus grand nombre les prédicats ; avec notamment différentes catégories de mots comme des noms, des adjectifs ou des interjections (Bassano, 1998 ; Kern, 2019).

La variabilité peut aussi être causée par l'environnement social de l'enfant. En effet, l'acquisition du lexique est rendue possible grâce à l'environnement ainsi qu'à l'expérience sociale (Bassano, 2000). L'enfant adapte les fonctionnalités du langage en fonction de celui des adultes qui l'entourent. Si les adultes utilisent le langage référentiel pour se référer au monde et le catégoriser (dénomination de nouveaux objets), alors l'enfant assimilera que la fonction principale du langage est de référer et d'étiqueter des objets. À l'inverse, si les adultes utilisent le langage à des fins de régulation sociale (notamment saluer, remercier, interdire), alors l'enfant assimilera qu'il s'agit de la fonction primaire du langage (Tomasello et al., 1983,1986). Bruner (1983) a montré l'importance des routines sociales structurées comme le bain, les jeux, les repas pour l'émergence des premiers mots. De plus, le comportement des adultes entourant l'enfant est important pour l'acquisition du lexique, car il fournit un ensemble d'indices pragmatiques. Les parents jouent un rôle indispensable dans le développement lexical et c'est grâce au processus d'imitation des conduites des adultes que l'enfant va pouvoir s'approprier de nouveaux mots et les catégoriser (Ogren et al., 2021).

Le genre de l'enfant est également un facteur causant des différences interindividuelles. Les filles et les garçons diffèrent concernant l'acquisition du lexique. Les filles sont généralement plus précoces que les garçons. Dans un premier temps, la compréhension des mots chez les filles est légèrement plus avancée que chez les garçons (Fenson et al., 1994). Par la suite, elles produisent leurs premiers mots plus tôt que les garçons, de même pour le stade des 50 premiers mots, elles l'atteignent avant les garçons et enfin, la combinaison des premiers mots arrive plus tôt chez les filles que les garçons (Kern, 2019). Dans une étude de 2004, Bornstein et ses collègues ont observé des enfants (filles et garçons) âgés de 20 mois durant une session de jeux. Ils ont enregistré et observé les enfants sur une durée de 10 minutes. Ils ont constaté que sur cette même durée de temps, les filles produisaient plus de mots que les garçons : 82 mots contre 54. Cette étude montre que les filles sont plus précoces que les garçons concernant les premières productions.

L'étude de Bouchard et ses collègues (2009) confirment que le genre de l'enfant révèle des différences lors du développement du langage. En effet, les filles montrent de meilleures compétences linguistiques que les garçons dès le plus jeune âge, avec l'apparition des premiers mots qui se fait plus tôt et de manière plus rapide. L'acquisition des formes grammaticales débute légèrement plus tôt chez les filles. Leurs énoncés sont plus complexes sur le plan syntaxique contenant un plus grand nombre de formes grammaticales. Ainsi, il existe une supériorité linguistique des filles sur les garçons à un jeune âge (ibid, 2009).

1.2.4 Le rôle de l'adulte dans le processus d'acquisition

L'enfant possède un accès direct au langage grâce aux échanges quotidiens qu'il entretient avec son entourage, c'est un processus d'apprentissage à partir de l'expérience (Hoff et al., 2002). L'attention conjointe est un facteur clé et fondamental dans le développement lexical précoce des enfants. Ce phénomène désigne la capacité des individus à partager et coordonner leur attention sur un objet ou un événement spécifique, en orientant leur regard et leurs actions vers un même référent (Tomasello, 2003). L'attention conjointe permet à l'enfant de relier des mots à des objets ou des actions, facilitant ainsi l'apprentissage du langage. L'acquisition précoce du langage se déroule dans des contextes de routines entre parents et enfants, qui sont des interactions récurrentes et familières pour l'enfant (Bruner, 1983). Tomasello et Farrar (1986) ont montré l'importance de l'attention conjointe dans l'acquisition précoce du langage en analysant deux perspectives distinctes, que l'on peut qualifier de « macro » et « micro ». Le niveau « macro » fait référence à l'analyse des périodes prolongées d'interactions entre adultes et enfants, tandis que le niveau « micro » se concentre sur les moments spécifiques où adultes et enfants coordonnent leurs attentions sur un élément précis du langage, en se focalisant conjointement sur son référent intentionnel. Dans cette étude, les auteurs ont observé que, dans des situations d'attention conjointe, les discussions entre adultes et enfants étaient plus longues et que les phrases produites par l'enfant étaient plus longues, tandis que celles de l'adulte étaient plus courtes. Ces interactions favorisent donc le langage dirigé de l'adulte vers l'enfant. De plus, au niveau « micro », lorsqu'un adulte présente une étiquette d'objet à l'enfant, cela attire son attention et, dans ces épisodes d'attention conjointe, l'enfant apprend de nouveaux mots (ibid, 1986 ; Bruner, 1983).

De plus, les enfants produisent leurs premiers mots grâce aux échanges avec les adultes, mais ces premières productions sont souvent modifiées par rapport à la cible adulte. Autrement dit, l'enfant utilise fréquemment des formes non conventionnelles que les parents peuvent reformuler sans nécessairement s'attarder sur l'erreur. En effet, le plus souvent, les parents ne signalent pas directement ces formes non conventionnelles, mais leur reformulation permet de poursuivre le dialogue de manière fluide sans marquer l'erreur (De Pontonx et al., 2019). Lorsque l'enfant produit une forme non conventionnelle, la reformulation de l'adulte permet souvent à l'enfant de produire un énoncé plus conforme à la norme, sans pour autant interrompre le processus d'échange. Cependant, même si la reformulation peut jouer un rôle important dans le développement du langage, elle n'est pas un facteur essentiel à l'acquisition du langage.

Les enfants peuvent apprendre de nouveaux mots, de nouvelles constructions et énoncés complexes, sans qu'il n'y ait systématiquement de reformulation de la part des adultes (Veneziano, Sinclair et Berthoud, 1990). Les sollicitations des adultes, notamment à travers les répétitions et les corrections implicites, permettent aux enfants de s'approprier des formes plus attendues, en favorisant des productions plus adaptées au contexte conversationnel (De Pontonx et al., 2019).

La parole des adultes fournit de nombreuses informations aux enfants lors de l'apprentissage de leur langue maternelle. Le discours que les enfants entendent autour d'eux est très répétitif. En effet, bon nombre de modèles produits par les enfants sont présents dans le discours des adultes ; ils utilisent des constructions similaires (Cameron-Faulkner et al., 2003). Cette répétition permet aux enfants de construire progressivement des représentations syntaxiques semblables à celles des adultes (Bybee, 1995). Bien que l'étude de Cameron-Faulkner et ses collègues (2003) ne soit pas spécifiquement centrée sur le développement du lexique, elle illustre bien l'approche basée sur l'usage du langage, soulignant l'importance de la fréquence des constructions dans l'acquisition des structures linguistiques. Par ailleurs, les enfants sont exposés à une grande quantité de langage chaque jour ; ils entendent environ 7000 énoncés, dont 2000 questions, 1500 fragments d'énoncés, 1000 copules et 400 énoncés complexes. C'est grâce à la répétition fréquente de ces constructions que l'enfant acquiert progressivement le langage, y compris les formes syntaxiques et lexicales (Cameron-Faulkner et al., 2003).

Dans cette seconde partie de notre revue de littérature, nous avons pu constater que le développement du lexique et notamment les rythmes d'acquisition et la structuration du premier lexique ont été largement décrits dans la littérature. Dans la section suivante, nous traiterons des travaux ayant cherché à comprendre quels facteurs pouvaient influencer la production des formes par les enfants. Notre travail portant spécifiquement sur les noms, nous avons réduit notre revue de littérature à cette catégorie. Dans un premier temps, il sera question des travaux portant sur les caractéristiques sémantiques du nom. Ensuite, notre attention se portera sur les recherches concernant la longueur du nom, la fréquence, suivies de celles sur la valence émotionnelle et le niveau de développement langagier.

1.3 Le développement du nom

1.3.1 Les caractéristiques sémantiques du nom

Parmi les divers traits sémantiques qui caractérisent le nom, nous considérerons, dans notre étude, l'opposition distinguant les noms animés des noms inanimés, mais également la distinction entre noms concrets et noms abstraits.

La distinction entre animé et inanimé dans la langue française repose sur une catégorisation lexicale qui attribue à certains noms des caractéristiques de vivacité ou de sensibilité (Benveniste, 1966). Un nom est dit animé s'il désigne un être vivant, c'est-à-dire un individu possédant une forme de vie ou de conscience, qu'il s'agisse d'êtres humains (ex. « homme », « femme »), d'animaux (ex. « chat », « chien »), ou parfois de certains êtres vivants comme les plantes ou les micro-organismes (Joly, 2009). Cette catégorie inclut également des noms désignant des êtres doués d'*agency*, c'est-à-dire capables d'agir de manière autonome sur leur environnement (Corbin, 2004). À l'inverse, un nom est inanimé lorsqu'il désigne un objet ou un concept qui ne possède pas de vie, de mouvement propre ou de conscience. Les noms inanimés incluent des objets matériels (ex. « table », « voiture », « chaise ») ainsi que des abstractions, des idées ou des concepts (ex. « liberté », « amour », « joie »). Ces objets ou concepts sont généralement considérés comme dénués de sensibilité ou de vie propre (Garrigues, 2010). En français, cette distinction entre animé et inanimé est particulièrement importante pour des phénomènes syntaxiques tels que l'accord en genre et en nombre, ainsi que la formation de certaines expressions grammaticales, comme les verbes d'état ou les adjectifs qui se rapportent à des entités animées ou inanimées (Corblin, 2004). Par exemple, dans une phrase comme « Le chat est vivant », « chat » est animé, tandis que dans « La table est en bois », « table » est inanimée. Il est également à noter que cette distinction peut parfois être floue, notamment dans des contextes figurés où des objets inanimés peuvent être personnifiés (ex. « La voiture crie sous la pluie »), mais elle reste une base fondamentale de la syntaxe et de la sémantique du français (Jouitteau, 2011). La personnification de certains objets inanimés dans des contextes métaphoriques est un phénomène que l'on retrouve dans de nombreuses langues, et il peut rendre la distinction entre animé

et inanimé moins nette (Tournier, 1998). Néanmoins, cette distinction reste essentielle pour comprendre certaines régularités grammaticales en français et dans d'autres langues (Vellard, 2007).

Dans des ouvrages de grammaire de référence, les noms concrets sont définis comme des noms dont la référence est perçue par les sens : la vue, l'ouïe, le goût, l'odorat, le toucher ; alors que les noms abstraits regroupent les noms de propriété (ex. faiblesse), de processus (ex. destruction) et de relation (ex. infériorité), généralement dérivés de verbes et d'adjectifs (i.a. Riegel et al., 1994).

Chez les enfants, avant 20 mois, les noms animés concrets sont fréquents, mais leur proportion diminue progressivement par la suite : les enfants désignent d'abord des êtres humains de leur entourage comme « maman », « papa », « bébé », et ensuite des animaux tels que « lapin », « nounours », « canard » ou « chat » (Kern, 2007), principalement utilisés lors d'activités manipulant ces référents.

Les noms inanimés renvoient à des noms d'objets concrets. Pour cette catégorie, l'enfant produit des noms référant à des parties de son corps comme « cheveux », « pied », « nez », « main », des aliments tels que « pomme », « pain », « carotte », « eau », des vêtements « chausson », « culotte », « chapeau » ou encore des objets, des lieux liés aux situations de la vie quotidienne, par exemple « bain », « fourchette », « chambre » (Bassano, 1998, 2005). Ainsi, entre 1 an et 2 ans, une proportion importante de noms produits par les enfants sont des noms inanimés concrets. Cela renforce l'idée que ces derniers représentent l'élément central de la catégorie nominale des enfants.

De plus, concernant la production de noms abstraits, ils sont rares et n'apparaissent qu'à l'âge de 18 – 20 mois. Ils font d'abord référence à des séquences d'actions ou des situations comme « dodo », « histoire » et ultérieurement, à l'expression des émotions et des sentiments comme « peur », « envie », « faim ». Les noms abstraits augmentent non seulement avec l'âge, mais également grâce à l'expérience de l'enfant (Bassano, 1998, 2005). Ce phénomène s'explique par le fait que les noms abstraits ne sont pas connectés au système imagé de l'enfant contrairement aux noms concrets. Ainsi, le système imagé fournissant de l'information supplémentaire, les noms concrets sont plus susceptibles d'être reconnus plus rapidement que les noms abstraits qui ne sont pas connectés au système imagé, d'où l'importance de l'expérience de la part de l'enfant (Ferrand, 2001). Cette idée est également soutenue par des recherches suggérant que les concepts concrets sont associés à des représentations perceptuelles qui facilitent leur traitement. De plus, l'expérience sensorimotrice semble jouer un rôle essentiel dans le traitement des concepts abstraits, comme l'indiquent les travaux de Borghi et Pecher (2011).

Pour conclure, les noms animés et les noms inanimés sont produits en nombre égal jusqu'à l'âge de 18 mois, selon des études qui indiquent qu'à cet âge, les enfants produisent généralement entre 50 et 100 mots, avec une répartition équitable entre noms animés (par exemple, « maman », « papa », « chien ») et noms inanimés (par exemple, « balle », « voiture », « chaise ») (Bates, 1976 ; Fenson et al., 1994). Cependant, à partir de 2 ans, des différences commencent à apparaître dans les productions des enfants. En effet, la production de noms animés stagne alors que celle de noms inanimés augmente. Cette augmentation des noms inanimés fait partie du phénomène d'explosion lexicale observé à cet âge. Ainsi, à 2 ans et demi, l'enfant produit majoritairement des noms inanimés, tandis que la production des noms animés diminue (Bassano, 1998, 2005).

1.3.2 La longueur du nom

On observe une dominance des productions monosyllabiques lors de la production des premiers mots. Gayraud et Kern (2007) ont mené une étude dans laquelle elles mettent en perspective la longueur des énoncés et des premiers mots d'enfants répartis en quatre groupes d'âge (24, 30, 46 et 87 mois). Elles émettent l'hypothèse selon laquelle, plus les enfants sont jeunes plus les noms qu'ils produisent sont courts, ce qui expliquerait la prédominance des productions monosyllabiques dans les premières étapes du développement. Bien que la fréquence des productions monosyllabiques diminue avec l'âge, cette diminution n'est en fait pas significative. De plus, les autrices ont observé une répartition similaire des noms dissyllabiques dans tous les groupes d'âge (24, 30, 46 et 87 mois). Enfin, à 24 mois, seuls des noms monosyllabiques et dissyllabiques sont acquis, tandis que l'acquisition des noms trisyllabiques se produit beaucoup plus tardivement. En effet, un changement significatif est observé à l'âge de 46 mois pour les noms trisyllabiques (ibid, 2007).

1.3.3 La fréquence du nom

La fréquence fait partie des facteurs qui peuvent influencer l'acquisition des noms. La fréquence d'occurrence d'un mot renvoie au nombre de fois où un mot est entendu par l'enfant (Kern, 2019). Cette fréquence favorise la production des mots : les mots les plus fréquents sont produits plus tôt que les autres. Par exemple, dans le cas où l'enfant est confronté à deux mots nouveaux, le mot le plus fréquemment entendu et produit par l'entourage est appris en premier (Ambridge et al., 2015 ; Goodman et al., 2008). Plusieurs études ont montré que la fréquence affecte la catégorie des mots que les enfants apprennent. Par exemple, l'anglais est une langue qui compte plus de noms que de verbes ; de ce fait les enfants américains, apprenant l'anglais, utilisent une forte proportion de noms dans leurs productions précoces. Cela amène à constater que la fréquence à laquelle les enfants entendent des mots appartenant à des catégories syntaxiques différentes peut avoir une influence sur les productions de certains mots (Goodman et al., 2008). De plus, comme nous avons pu le relever plus tôt (cf. section 1.2.4), on retrouve un lien entre la production des premiers mots de l'enfant et les productions de son entourage. En effet, la production des premiers mots des enfants est fortement corrélée avec la production des mères (Tomasello et al., 1983,1986) ; plus la mère produit un mot fréquemment plus l'enfant va le produire à son tour (Ogren et al., 2021).

Cependant, la fréquence est une variable qui présente certaines limites. L'hypothèse selon laquelle plus le mot est fréquent, plus l'enfant l'apprend tôt, peu importe les catégories lexicales n'est pas toujours exacte. En effet, par exemple, dans le discours des adultes, les mots de classes fermées sont fréquents alors qu'ils ne le sont pas dans celui de l'enfant (Goodman et al., 2008).

1.3.4 La valence émotionnelle

Les mots, en tant qu'unités complexes, se définissent par un ensemble de caractéristiques formelles et sémantiques (Syssau et al., 2005). Ils se distinguent les uns des autres sur un grand nombre de dimensions, dont celle de la valence émotionnelle. Elle est considérée comme une dimension fondamentale dans la signification du mot (Syssau et al., 2009). La valence émotionnelle se définit comme le degré avec lequel sa référence est plaisante ou déplaisante. Elle fait partie du « bagage sémantique » du mot et est déterminée par l'impact émotionnel que celui-ci peut avoir sur un individu (Bonin et al., 2003).

Les mots qui désignent des émotions sont appelés mots dénotatifs (Syssau et al., 2005), et leur valence peut être positive ou négative, en fonction de la charge émotionnelle associée. Par exemple, des mots tels que « joie » ou « amour » véhiculent une valence positive, tandis que des termes comme « douleur » ou « colère » sont associés à une valence négative. Cependant, certains mots peuvent être considérés comme ayant une valence neutre, lorsqu'ils ne suscitent aucune réponse émotionnelle marquée, comme c'est le cas pour des mots tels que « table » ou « voiture » (Syssau et al., 2005).

La valence émotionnelle des mots joue un rôle clé dans leur traitement cognitif. De nombreuses études ont montré que la valence d'un mot explique entre 50 et 60 % de la variance dans la manière dont sa signification est appréhendée (Bonin et al., 2003). En effet, les mots à forte valence émotionnelle ont tendance à capter l'attention de manière plus marquée que les mots à valence neutre, cet effet étant particulièrement prononcé pour les mots à valence négative. Des recherches ont montré que les mots négatifs, comme « peur » ou « échec », activent plus rapidement les mécanismes attentionnels que les mots positifs, ce qui peut être lié à des mécanismes de survie adaptatifs (Pratto et John, 1991).

Les premiers mots à valence émotionnelle commencent à émerger dans les conversations des enfants âgés de 18 à 36 mois. Ces mots, souvent liés à des expériences émotionnelles fondamentales comme le plaisir, la douleur ou la détresse, sont utilisés par les jeunes enfants pour exprimer leurs états émotionnels et interagir avec leur environnement social (Dunn et al., 1987 ; Ogren et al., 2021 ; Syssau et al., 2009). À l'âge de 2 ans, la production de mots à valence émotionnelle est particulièrement stimulée lors des jeux ou dans les interactions sociales avec les parents et les proches. La majorité de ces premiers mots émotionnels ont

une valence positive, ce qui reflète la tendance des jeunes enfants à acquérir en priorité des mots liés à des expériences agréables ou sécurisantes (Dunn et al., 1987).

1.3.5 Le niveau de développement langagier

Le niveau de développement langagier constitue un indicateur corrélé au développement des noms, car les enfants ayant un niveau langagier plus avancé, mesuré par des indicateurs tels que la longueur moyenne des énoncés (MLU), ont tendance à développer un lexique plus riche en noms (Bates et al., 1994). Le MLU est une mesure de base dans l'étude de l'acquisition du langage, qui permet d'évaluer la complexité des énoncés produits par l'enfant. Il est calculé en déterminant le nombre moyen de mots ou de morphèmes par énoncé produit par l'enfant (Brown, 1973). Il sert à situer le niveau de développement langagier des enfants et à comparer ces derniers entre eux sur la base de leur production langagière (Parisse et Le Normand, 2006).

En outre, le MLU est corrélé à l'âge de l'enfant, ainsi qu'aux interactions sociales qu'il entretient avec son entourage. Les enfants qui sont plus engagés dans des interactions sociales complexes et variées, notamment avec des adultes, ont tendance à produire des énoncés plus longs et plus complexes, ce qui se reflète dans un MLU plus élevé. Ces interactions permettent à l'enfant de développer davantage de structures syntaxiques et morphologiques, favorisant ainsi l'enrichissement de son lexique, notamment en ce qui concerne les noms (Le Normand et al., 2013). Cette corrélation est particulièrement forte avant l'âge de 4 ans, période durant laquelle l'acquisition des structures grammaticales de base est en plein développement. Cependant, après 4-5 ans, cette corrélation diminue, car les enfants commencent à atteindre une certaine maturité langagière et la progression du MLU devient moins directement liée à l'âge (Parisse et Le Normand, 2006)¹.

¹ La fiabilité du MLU a été remise en question, car la longueur moyenne des énoncés varie au sein des groupes d'âge et il s'agit d'une mesure sensible aux influences pragmatiques. Par exemple, la structure des énoncés peut varier en fonction du contexte de communication, du type d'interaction ou des émotions de l'enfant. Ces facteurs peuvent rendre le MLU moins fiable comme indicateur du développement langagier.

1.4 Question de recherche

Les travaux présentés dans notre revue de littérature amènent à constater que de nombreux facteurs ont déjà été considérés dans les travaux en acquisition pour expliquer le développement lexical, et en particulier le développement du nom : les caractéristiques sémantiques du nom (caractère animé/inanimé, concrétude), sa longueur, sa fréquence, la valence émotionnelle qui y est associée et le niveau de développement langagier. Notre travail s'inscrit dans la continuité des recherches récentes en offrant une vue d'ensemble sur les avancées réalisées ces dernières années. L'objectif principal est d'intégrer simultanément ces différents facteurs dans les analyses, afin d'apporter une nouvelle perspective et de mieux comprendre leurs interactions. Ainsi, nous nous interrogeons dans notre étude sur l'influence de l'ensemble de ces facteurs sur l'acquisition du nom : un ou plusieurs de ces facteurs contribuent-ils à expliquer l'acquisition des noms chez l'enfant ? Si oui, lesquels ?

Dans notre étude, d'autres facteurs propres aux individus et aux contextes d'interaction ont également été pris en compte : le genre de l'enfant (fille/garçon) et le nombre d'occurrences des noms dans le langage adressé à et entendu par l'enfant. Étant donné que les enfants sont en interaction avec un adulte de leur entourage dans les corpus choisis pour notre étude, constate-t-on une influence du langage de l'adulte sur les productions de l'enfant ? De plus, le genre de l'enfant joue-t-il un rôle important ?

CHAPITRE 2

MÉTHODE

Ce chapitre de méthode est divisé en deux grandes parties. Dans la première partie, nous présentons le corpus utilisé pour l'étude, en détaillant sa composition et sa provenance. Ensuite, nous présentons les locuteurs et locutrices du corpus sélectionné. De plus, nous décrivons la norme ValEmo utilisée tout au long du travail. Dans la seconde partie, nous décrivons les variables prises en considération dans l'étude ainsi que l'analyse appliquée à nos données (norme et corpus). De plus, nous expliquons le modèle statistique utilisé lors de l'analyse.

2.1 Corpus et participants

Le corpus utilisé pour ce travail est extrait de la base de données Child Language Data Exchange System (CHILDES). Cette base de données a été créée en 1984 et a d'abord été élaborée par les chercheurs Brian MacWhinney et Catherine Snow (MacWhinney, 2000). Il s'agit d'une base de données accessible en ligne sur la production langagière de l'enfant. On y trouve des corpus d'enfants, le plus souvent en interaction avec un adulte. Les corpus disponibles couvrent plus de trente langues différentes (Morgenstern et al., 2007). Tous les corpus accessibles sont transcrits et associés à des fichiers audio ou vidéo. Le plus souvent, les données sont des enregistrements d'interactions et de conversations spontanées entre l'enfant et son entourage.

Dans le cadre de notre travail, nous avons fait le choix d'utiliser des corpus existants afin de constituer notre ensemble final de données. Le tableau suivant présente les corpus utilisés, le nombre de séances sélectionnées pour chacun de ces corpus ainsi que les références des travaux présentant ces données.

Tableau 2.1 Corpus de CHILDES sélectionnés pour notre étude

Corpus	Nombre de séances	Référence
Lyon	30	Katherine Demuth et Harriet Jisa (Demuth & Tremblay, 2008)
Paris	40	Aliyah Morgenstern (Leroy, Mathiot & Morgenstern, 2009; Morgenstern & Parisse, 2007; Morgenstern & Sekali, 2009; Morgenstern, 2006; Morgenstern & al., 2009)
Yamaguchi	10	Naomi Yamaguchi (Yamaguchi, 2012)
York	10	Bernadette Plunkett et Cécile De Cat (De Cat & Plunkett, 2002 ; Plunkett, 2002).
Pauline	10	Isabelle Maillochon (Bassano & Maillochon, 1994 ; Bassano, Maillochon & Eme, 1998 ; Bassano, 2000 ; Bassano et al., 2004).

Notre étude se base sur dix suivis longitudinaux extraits de ces corpus. Plusieurs études ayant montré des différences développementales entre filles et garçons (Bouchard et al., 2009), nous avons retenu autant de corpus de filles (5 ; Anaïs, Marie, Anne, Anae et Pauline) que de garçons (5 ; Théophile, Léonard, Antoine, Adrien et Théotime). Tous les enfants vivent en France métropolitaine, trois d'entre eux sont originaires de la ville de Lyon (Marie, Théotime et Anaïs), cinq de Paris (Anae, Léonard, Théophile, Antoine et Adrien), une de Rouen (Pauline) et la dernière participante du sud de la France, et plus spécifiquement de la région des Pyrénées (Anne). Tous les enfants sélectionnés ont pour L1 le français parlé en France, de même que leurs parents.

La littérature sur le traitement de la valence émotionnelle chez les enfants (déjà citée section 1.3.4 ; cf. Ogren, 2021) a montré qu'à partir de 24 mois les enfants associent plus facilement les noms à valence émotionnelle à des situations de la vie quotidienne. Pour cette raison, nous avons fait le choix de ne

sélectionner que des séances après 24 mois, nous donnant ainsi davantage d’opportunités de saisir l’impact de la valence sur la production des noms chez ces enfants. Dans les séances choisies, les enfants étaient en présence d’un ou plusieurs interlocuteurs : mère, père, frère(s)/sœur(s) et autre(s) membre(s) de la famille (grand(s)— parent(s), oncle(s) ou tante(s), cousin(s)).

Tableau 2.2 Âge des participants enfants

Enfants	Âges
Adrien	De 3 ;03 à 4 ;03
Anaé	De 3 ;08 à 5 ;10
Anaïs	De 2 ;06 à 3 ;00
Anne	De 2 ;06 à 3 ;05
Antoine	De 2 ;09 à 5 ;11
Léonard	De 2 ;00 à 3 ;02
Marie	De 2 ;11 à 4 ;00
Pauline	De 2 ;11 à 4 ;03
Théophile	De 3 ;06 à 4 ;11
Théotime	De 2 ;06 à 3 ;00

En complément du tableau 2.2, le tableau suivant présente de manière plus précise les âges des enfants pour chacune des séances.

Tableau 2.3 Âge des enfants par séance

Âge des enfants										
	Adrien	Anae	Anaïs	Anne	Antoine	Léonard	Marie	Pauline	Théophile	Théotime
Séance 1	3;03.12	3;08.09	2;06.21	2;06.02	3;00.21	2;00.24	2;11.14	2;01.17	3;06.10	2;06.17
Séance 2	3;05.14	3;10.00	2;09.16	2;06.18	3;07.01	2;01.26	3;01;02	2;02.05	3;07.08	2;07.03
Séance 3	3;08.05	4;00.12	2;09.16	2;07.01	4;02.19	2;03.00	3;02.08	2;02.19	3;08.15	2;07.28
Séance 4	3;09.09	4;01.17	2;09.26	2;07.16	4;05.16	2;03.26	3;02.23	2;02.29	3;10.00	2;08.12
Séance 5	3;10.13	4;04.10	2;10.20	2;09.15	4;08.02	2;04.27	3;05.21	2;03.20	3;11.15	2;08.25
Séance 6	3;11.13	4;05.23	2;10.20	2;11.02	5;02.13	2;07.10	3;06.19	2;04.03	4;01.24	2;09.16
Séance 7	4;00.15	4;08.08	3;00.00	3;00.02	5;04.28	2;07.27	3;07.16	2;04.17	4;03.29	2;10.01
Séance 8	4;01.12	5;01.20	3;00.00	3;02.04	5;07.11	2;11.03	3;09.04	2;05.07	4;06.20	2;10.12
Séance 9	4;02.15	5;07.06	3;00.15	3;03.16	5;11.18	2;11.29	3;11.08	2;05.20	4;09.07	2;10.15

Séance 10	4;03.26	5;10.30	3;00.15	3;05.04	6;03.08	3;02.25	4;00.05	2;06.13	4;11.11	3;00.03
--------------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

Dans notre corpus, parmi toutes les activités dans lesquelles les participants sont impliqués, les séances de jeu sont les plus nombreuses. Au cours de ces temps ludiques, les enfants parlent et manipulent des objets et des jouets. Parmi ces jouets, on trouve des jeux de construction (ex. des cubes en bois), des jeux symboliques (ex. jeux avec des figurines) et des casse-têtes. On trouve aussi dans nos données des temps de lecture pendant lesquels les participants manipulent des livres imagés et illustrés. Dans certaines séances, les enfants sont impliqués dans des activités manuelles comme le dessin, la peinture ou l'apprentissage de l'écriture de certaines lettres de l'alphabet. Enfin, les enfants sont aussi enregistrés lors des temps de repas, de goûter ou lorsqu'ils aident au rangement des courses ou à la préparation des repas.

2.1.1 La norme utilisée pour l'étude

Dans le cadre de notre travail, nous avons par ailleurs eu recours à une base de valence émotionnelle. Nous avons utilisé la norme ValEmo, une base de données qui fournit une évaluation émotionnelle (valence émotionnelle) de 600 noms de la langue française. Rappelons que la valence émotionnelle d'un mot est définie comme le degré avec lequel sa référence est plaisante ou déplaisante (Bonin et al., 2003). Syssau et Monnier (2009) ont élaboré cette norme, car, auparavant, aucune base de données, évaluant l'émotivité associée aux mots, n'existait pour les enfants francophones.

Pour construire cette norme, il a été demandé à 350 enfants français d'évaluer la valence émotionnelle des 600 noms préalablement choisis. Les enfants étaient répartis en trois groupes :

- 85 élèves de maternelle (44 filles et 41 garçons) avec un âge moyen de 5.8² ans.
- 127 élèves de CE1 (75 filles et 52 garçons) avec un âge moyen de 7.8 ans.
- 138 élèves de CM1 (71 filles et 67 garçons) avec un âge moyen de 9.8 ans.

² La notation d'âge utilisée est celle disponible dans les références de la base (Syssau, A., & Monnier, C. (2009). Children's emotional norms for 600 French words. *Behavior Research Methods*, 41(1), 213-219.)

Les 600 noms constituant la base sont issus de deux bases de données, 191 proviennent uniquement de la base de données de Bonin et al. (2003), 130 de celle de Syssau et Front (2005) et 271 mots sont présents dans les deux bases. Les noms de la norme ValEmo ont été présentés à haute voix aux enfants par l'expérimentateur. Concernant les enfants de 9 ans, le corpus a été présenté dans sa version complète (600 mots). Il a été réduit à 588 mots lorsqu'il a été présenté aux enfants de 7 ans et à 200 mots dans la version présentée aux enfants de 5 ans. L'évaluation de la valence émotionnelle des mots a été réalisée à l'aide d'une échelle à trois points (positif, négatif, neutre). Afin de simplifier la tâche d'évaluation pour les enfants, les trois points ont été étiquetés à l'aide de dessins : bouche heureuse pour la valence positive, bouche triste pour la valence négative et bouche droite pour la valence neutre. Les enfants de 9 ans ont évalué les 600 noms répartis en huit ensembles de 75 noms, et chaque enfant évaluait deux ensembles lors de deux sessions. Pour les enfants de 7 ans, les 588 noms ont été divisés en huit ensembles. Chaque enfant a évalué deux ensembles : un de 74 noms et l'autre de 73 noms, lors de deux sessions. Et enfin, les enfants de 5 ans, ont évalué les 200 noms répartis en quatre ensembles de 50 noms. Chaque enfant a évalué deux ensembles de noms lors de deux sessions distinctes.

Après cette évaluation, afin d'examiner la fiabilité des réponses des enfants, deux analyses de corrélations ont été effectuées pour observer la significativité et les variations de pourcentage des variables. Par la suite, les chercheurs ont constitué deux groupes de noms émotionnels distincts : les noms à fort accord (*high-agreement*) et les noms à faible accord (*low-agreement*). Un nom à fort accord (HA) est évalué comme positif, neutre ou négatif à 60 % ou plus par les enfants alors qu'un nom à faible accord (LA) l'est à 40 % ou moins. La distribution des noms LA et HA varie avec l'âge. À 5 et 9 ans, plus de 50 % des noms ont reçu la même évaluation par plus de 60 % des enfants, c'est-à-dire que plus de la moitié des noms était à fort accord (HA). Alors qu'à l'âge de 7 ans, les enfants étaient plus en désaccord dans les évaluations, moins de 50 % des noms étaient à fort accord (HA). La majorité des noms à fort accord (HA) était des noms avec une valence émotionnelle positive. Toutefois, une variation selon le genre a été observée à l'âge de 7 et de 9 ans. Sur les 588 noms évalués à 7 ans, 353 noms sont évalués comme nom à fort accord (HA) pour les filles contre 252 par les garçons. Cette tendance se confirme également à 9 ans, où parmi les 600 noms, 394 noms ont été évalués comme nom HA par les filles, contre 319 seulement pour les garçons. De plus, les noms à valence neutre augmentent significativement entre 5 et 7 ans tandis que les noms à valence positive diminuent. Le même phénomène se poursuit entre 7 et 9 ans. Malgré quelques variations de genre, au cours de cette étude, les chercheurs n'ont trouvé aucune différence significative entre garçons et filles.

Tableau 2.4 Proportion de noms à fort accord et faible accord selon l'âge et le genre (Syssau et Monnier, 2009)

Age	Sexe	Words				Total (Corpus Size)
		High Agreement		Low Agreement		
		No.	%	No.	%	
5 years	Girls	116	58	84	42	200
	Boys	111	56	89	44	200
	All	104	52	96	48	200
7 years	Girls	353	60	235	40	588
	Boys	252	43	336	57	588
	All	261	44	327	56	588
9 years	Girls	394	66	206	34	600
	Boys	319	53	281	47	600
	All	340	57	260	43	600

Malgré un ensemble de mots limité — 600 noms — cette base est pertinente pour mon travail. Dans un premier temps, contrairement aux nombreuses normes de la littérature qui sont évaluées par des adultes, dans la norme ValEmo, ce sont des enfants âgés de 5, 7 et 9 ans qui ont évalué la valence émotionnelle des mots. Bien que ces enfants soient plus âgés, ils restent les plus proches du groupe d'enfants de mon

étude. Ce critère est important pour mon étude. De plus, cette norme prend en compte la différence de genre (fille/garçon), aspect qui sera également pris en considération dans mon travail³.

2.2 Les variables et l'analyse du corpus

Dans notre travail, une variable dépendante a été définie, soit le nombre d'apparitions dans les corpus, de chacun de ces noms chez l'enfant. Les variables indépendantes, propres aux noms de l'étude, sont identiques pour tous. Ont été répertoriés les caractéristiques sémantiques du nom, c'est-à-dire le caractère animé/inanimé et la concrétude, la longueur du nom, la fréquence du nom, la valence émotionnelle, le nombre d'occurrences produites par les adultes présents lors de l'enregistrement, le niveau de développement langagier ainsi que le genre de l'enfant. Ces variables sont prises en compte dans l'étude afin de contrôler leur effet et de mieux comprendre la distribution des données. Dans notre étude, nous avons défini deux variables aléatoires⁴ principales : le participant (l'enfant) et l'item (le nom).

Le traitement de nos données a commencé par une sélection des informations à extraire du fichier de la norme ValEmo. Le fichier ValEmo provient du site Lexique (New, Pallier et al., 2004). Il se présente sous la forme d'un tableau Excel de 28 colonnes. Les différentes colonnes indiquent le nombre de fois où le nom est identifié comme positif, négatif ou neutre en fonction de l'âge de l'enfant (5, 7 et 9 ans) et ensuite selon le genre (fille, garçon). Dans le cadre de notre travail, pour chaque nom, nous avons relevé les scores de valence des trois âges différents : à 5 ans, à 7 ans et à 9 ans. Les valences positives, neutres et négatives de chaque mot ont été mesurées pour chaque âge. Nous avons calculé la moyenne des valences positives pour chaque nom à travers ces trois âges, puis procédé de la même manière pour les valences neutres et négatives. Après avoir obtenu ces trois moyennes pour chaque nom, nous avons déterminé la valence dominante en sélectionnant la moyenne la plus élevée. Cette valence dominante nous a permis d'attribuer

³ Dans un premier temps, nous avons utilisé une autre base évaluant l'émotivité (norme EMA), mais elle s'avérait moins pertinente en termes de nombre de noms de la base aussi observés dans les corpus.

⁴ Nous avons choisi de ne pas considérer les sessions comme une variable aléatoire, car celles-ci peuvent coïncider avec le MLU des enfants. L'inclusion des sessions comme variable aléatoire n'apportait pas de bénéfice supplémentaire au modèle, car elle risquait d'introduire une redondance avec le MLU.

une étiquette « positif », « négatif » ou « neutre » à chaque nom du corpus. Le tableau suivant illustre des exemples d'attribution pour six mots présents dans notre corpus.

Tableau 2.5 Exemples d'attribution de valence pour six noms

Noms	5 ans			7 ans			9 ans			Moyenne obtenue			Valence retenue
	Négative	Positive	Neutre	Négative	Positive	Neutre	Négative	Positive	Neutre	Négative	Positive	Neutre	
bain	12	77	12	0	78	22	8	76	16	7	77	17	positive
cadeau	7	81	12	0	100	0	0	94	6	2	92	6	positive
chaise	29	33	38	9	28	63	3	32	65	14	31	55	neutre
biberon	60	14	26	44	19	36	30	22	49	45	18	37	négative
moustique	51	26	23	69	13	19	71	6	23	64	15	22	négative
ped	52	29	19	35	13	52	5	11	84	31	18	52	neutre

Après avoir effectué ce traitement sur la norme ValEmo, nous nous sommes concentrées sur le traitement du corpus. Les 10 sessions des 10 participants retenus ont été extraites de la base Childes. Les noms issus de la norme ValEmo ont ensuite été croisés avec ceux des 10 sessions des 10 participants⁵. Les noms présents à la fois dans la norme et dans le corpus d'interactions ont ainsi formé notre ensemble final de données. Un tableur Excel a été créé afin de regrouper les données extraites de la base, les noms de notre corpus et par la suite toutes les autres variables. Les noms étudiés sont tous des noms communs et ils sont au nombre de 419.

Les noms sélectionnés pour notre corpus sont en grande partie le reflet des thèmes observés dans les interactions adultes-enfants et les types d'activités dans lesquels les participants sont engagés. La diversité des activités, allant des jeux symboliques aux séances de lecture en passant par les tâches ménagères comme le rangement des courses et les repas, entraîne une variété de lexique produit. Ainsi, les noms utilisés au cours de ces interactions sont directement influencés par les objets et actions auxquels ils sont associés. Les jeux de construction, par exemple, induisent l'utilisation de noms liés aux objets matériels et à la manipulation (comme « cube », « tour » ou « voiture »), tandis que les séances de lecture favorisent l'usage de termes, associés de façon préférentielle à la narration et aux personnages (« histoire », « image »). De même, les contextes quotidiens, comme le repas et le rangement, apportent un lexique lié à la vie domestique comme « assiette », « fourchette » ou « nourriture ». La liste des noms étudiés dans ce travail est donc en cohérence avec les activités observées et les thématiques traitées dans les interactions.

2.2.1 Les caractéristiques sémantiques du nom

Les variables relatives aux caractéristiques sémantiques du nom (cf.1.3.1), qu'il s'agisse du caractère animé/inanimé ou de la concrétude, sont des facteurs binaires, autrement dit, le mot est animé ou inanimé, et il est concret ou abstrait. Ces variables ont été codées manuellement.

⁵ Ce travail de mise en correspondance a été effectué par Grégoire Winterstein que je remercie.

Concernant le caractère animé/inanimé, lors de l'annotation des noms, chaque nom a été catégorisé de manière binaire comme animé ou inanimé. Cette annotation a été effectuée en tant que caractéristique lexicale du nom, sans prendre en compte le contexte spécifique de l'énonciation ni la situation de communication dans laquelle le nom pouvait être utilisé. Autrement dit, l'annotation s'est basée sur la nature intrinsèque du nom et non sur les circonstances concrètes de son usage dans des énoncés particuliers (Joly, 2009). Par exemple, des noms comme « chat » et « bébé » ont été annotés comme animés, tandis que des noms comme « ballon » et « camion » ont été classifiés comme inanimés, indépendamment du contexte dans lequel ils apparaissent dans un discours ou une conversation. Cette approche, bien que permettant une analyse lexicale cohérente, ne prend pas en compte les variations que pourrait introduire la situation d'énonciation. Une telle distinction purement lexicale permet de poser une base solide pour l'analyse, mais elle ne reflète pas toute la richesse et la flexibilité du langage dans son usage quotidien (Duchêne, 2013).

Les noms inanimés sont en plus grand nombre que les noms animés. Parmi les 419 noms, 345 sont identifiés comme des noms inanimés contre seulement 74 noms animés.

Tableau 2.6 Répartition des noms animés et des noms inanimés

Noms inanimés	Noms animés
82,34 % (345 noms)	17,66 % (74 noms)

Pour la concrétude, le phénomène est similaire, c'est-à-dire que les noms concrets sont largement majoritaires par rapport aux noms abstraits. Les noms concrets représentent 90,69 % des noms et les noms abstraits ne représentent eux que 9,31 % des noms. En comparaison, sur les 419 noms retenus dans notre corpus, 380 sont des noms concrets contre seulement 39 qui sont abstraits.

Tableau 2.7 Répartition des noms concrets et des noms abstraits

Noms concrets	Noms abstraits
90,69 % (380 noms)	9,31 % (39 noms)

2.2.2 La longueur du nom

La variable de la longueur du nom est mesurée en nombre de syllabes à l'oral de la cible standard du mot, quelle que soit la production de l'enfant. Elle a été comptabilisée et codée manuellement. Elle regroupe quatre modalités : les noms monosyllabiques, les noms bisyllabiques, les noms trisyllabiques et les noms quadrisyllabiques. Les noms monosyllabiques et bisyllabiques sont largement majoritaires ; ils représentent 44,87 % et 45,11 % des noms. À l'inverse, les noms trisyllabiques et quadrisyllabiques représentent une proportion minoritaire de noms (8,59 % et 1,43 %). En valeurs absolues, 188 noms sont monosyllabiques, 189 bisyllabiques, 36 trisyllabiques et enfin seulement 6 sont quadrisyllabiques.

Tableau 2.8 Répartition des noms selon leur longueur

Noms monosyllabiques	Noms bisyllabiques	Noms trisyllabiques	Noms quadrisyllabiques
44,87 % (188)	45,11 % (189)	8,59 % (36)	1,43 % (6)

2.2.3 La fréquence

La fréquence du nom a été relevée au moyen du logiciel Antidote. L'indice de fréquence fourni par cet outil est affiché pour la plupart des mots. Il s'agit d'un nombre entre 0 (rare) et 100 (fréquent), calculé logarithmiquement et qui indique la fréquence relative des mots dans un corpus de textes de plusieurs milliards de mots (Druide informatique, 2016). Il est important de noter que, dans Antidote, ces fréquences sont calculées à partir de corpus écrits, notamment des textes littéraires, journalistiques et d'autres types de productions écrites récentes. Ces corpus ont été sélectionnés pour refléter l'usage moderne et diversifié du français écrit en contexte académique et médiatique (Ségéral, 2008). Les productions écrites étant souvent des indicateurs fiables de la fréquence et de la diversité des noms (Brybaert et New, 2009), nous avons retenu cet indice de fréquence pour notre travail. Cependant, ces fréquences ne tiennent pas compte des caractéristiques de l'oral, et encore moins des caractéristiques des interactions adulte-enfant et il faudra en tenir compte lors du traitement des résultats.

Concernant l'annotation, pour chaque mot, nous avons relevé le nombre correspondant à la fréquence et codé cet indice manuellement dans notre fichier Excel. Dans notre corpus, la fréquence la plus basse est 34, il s'agit des noms « arrosoir » et « marelle », et la fréquence la plus haute est 77 pour le nom « année ».

2.2.4 La valence émotionnelle

Comme souligné auparavant (cf. 1.2), les informations relatives à la valence émotionnelle proviennent de la norme ValEmo. Concernant la valence des noms de notre corpus, elle n'est pas répartie équitablement. Autrement dit, les noms à valence positive sont nettement plus nombreux que ceux à valence neutre, et davantage que ceux à valence négative. Les noms à valence positive représentent 57,76 % du corpus, soit 242 noms, les noms à valence neutre, 32,94 % avec 138 noms et enfin seulement 9,31 % des noms sont identifiés avec une valence négative, soit 39 noms.

Tableau 2.9 Répartition des noms selon leur valence émotionnelle

Valence positive	Valence neutre	Valence négative
57,76 % (242 noms)	32,94 % (138 noms)	9,31 % (39 noms)

2.2.5 Le nombre d'occurrences produites par les adultes

Pour chacun des noms de l'ensemble de noms constitué, nous avons comptabilisé le nombre d'occurrences produites par les adultes, soit le nombre d'apparitions du mot chez l'adulte dans l'échange avec l'enfant. Nous avons identifié et comptabilisé le nombre de fois où le mot a été produit par les adultes entourant l'enfant pour chaque séance choisie. Le nom le plus produit par les adultes est « main », il a été produit 24 fois dans la même séance ; il s'agit de la séance d'Adrien à 3 ;08.

2.2.6 Le niveau de développement langagier

Concernant la variable du niveau de développement langagier, définie par la longueur moyenne des énoncés (MLU) (cf. 1.3.5), elle a été regroupée en neuf modalités, par tranches de 0,5, afin de faciliter la répartition des enfants sur un même continuum (2,5 ; 3 ; 3,5 ; 4 ; 4,5 ; 5 ; 5,5 ; 6 ;7). Dans un premier temps, les longueurs moyennes des énoncés ont été calculées grâce à la commande MLU de CLAN. Ensuite, nous avons codé manuellement, pour chaque enfant, la modalité de MLU qui lui correspondait dans la grille d'analyse. Le tableau suivant présente la production des noms de notre étude selon le groupe MLU.

Tableau 2.10 Production des noms selon le groupe MLU

MLU 2,5	MLU 3	MLU 3,5	MLU 4	MLU 4,5	MLU 5	MLU 5,5	MLU 6	MLU 7
0,24 %	6,92 %	19,57 %	29,83 %	14,80 %	4,30 %	5,97 %	17,90 %	0,48 %
1 nom	29 noms	82 noms	125 noms	62 noms	18 noms	25 noms	75 noms	2 noms

Contrairement au tableau précédent, décrivant la répartition des noms selon le groupe MLU, le tableau suivant illustre la répartition des séances selon le groupe MLU des enfants. Dans ce tableau, nous remarquons que le groupe 1, représentant un MLU de 2,5, et le groupe 9, représentant un MLU de 7, ne comptent qu'un seul enfant. Nous avons conscience que la présence d'une seule séance pour ces deux MLU peut avoir un impact sur le traitement des résultats. Toutefois, nous avons choisi de ne pas éliminer ces groupes extrêmes, car leur exclusion pourrait fausser l'analyse en supprimant une portion potentiellement informative des données. Le groupe de MLU bas (comme 2,5) peut refléter les stades de développement précoce du langage, tandis que le groupe à MLU plus élevé (comme 7) peut indiquer les enfants dans des phases plus avancées du développement langagier. En les incluant, nous permettons une représentation complète de la diversité des compétences langagières dans notre échantillon, ce qui nous permet de mieux comprendre la variation naturelle des productions lexicales à différents niveaux de développement. Éliminer ces extrêmes risquerait de simplifier les résultats et de ne pas rendre compte des différences réelles observées dans le développement du lexique de l'enfant.

Tableau 2.11 Répartition des séances des enfants selon leur groupe MLU

Groupe 1	MLU 2,5	Léonard, séance 3
Groupe 2	MLU 3	Adrien, séance 1 Antoine, séance 1 Anaïs, séances 3, 4, 8, 9, 10 Léonard, séances 2, 4, 5, 6 Théotime, séance 10
Groupe 3	MLU 3,5	Anae, séances 1, 6, 7, 10 Anne, séances 1, 2, 4, Théophile, séance 1 Adrien, séance 2 Anaïs, séances 2, 5, 7 Théotime, séances 2, 8,
Groupe 4	MLU 4	Antoine, séances 2, 3, 4, 5, 6, 8 Marie, séances 2, 8, 9 Adrien, séances 3, 4 Théotime, séances 3, 4, 9 Anae, séances 4, 5, 9 Théophile, séance 6 Léonard, séances 7, 8

Groupe 5 MLU 4,5
Adrien, séances 5, 8, 9
Anne, séance 6,
Marie, séance 6
Anae, séance 8
Anne, séance 8
Léonard, séances 9, 10

Groupe 6 MLU 5
Marie, séance 4
Anne, séances 7, 9, 10
Théophile, séances 7, 8
Antoine, séance 9

Groupe 7 MLU 5,5
Théophile, séance 2
Marie, séance 5
Adrien, séance 6, 7

Groupe 8 MLU 6
Anae, séances 2, 3
Antoine, séances 7, 10
Adrien, séance 10

Groupe 9 MLU 7
Théophile, séance 9

2.2.7 Le genre

Pour terminer, la variable du genre comptabilise la proportion d'enfants filles et d'enfants garçons dans notre corpus. Elle est équitable : 50 % des participants sont des filles et 50 % des participants sont des garçons. Concernant la production des noms chez l'enfant, le corpus complet montre également une égalité entre les garçons et les filles, respectivement 50,60 % et 49,40 %. Le codage a également été effectué de manière manuelle⁶.

2.2.8 Le modèle mixte généralisé de type « Poisson »

Le modèle mixte généralisé de type « Poisson » est le modèle statistique que nous avons utilisé, dans notre étude, afin d'effectuer les analyses.

Le modèle mixte généralisé de type « Poisson » est un modèle statistique utilisé pour analyser des données de comptage tout en prenant en compte la structure hiérarchique ou les corrélations intragroupes. Ces données sont des entiers non négatifs utilisés pour représenter des événements se produisant de manière indépendante sur un intervalle de temps ou dans un espace donné. Dans notre cas, il s'agit de comptabiliser le nombre d'apparitions d'un nom chez l'enfant lors de séances précises. De plus, ce modèle intègre des effets fixes et des effets aléatoires dans l'analyse des données (Foulley, 1992). Les effets fixes sont utilisés pour modéliser l'impact des variables explicatives observables sur la variable dépendante. En revanche, les effets aléatoires, eux, modélisent la variation non observée entre les groupes (Rivest, 2013).

D'ailleurs, la partie Poisson du modèle gère les données de comptage en supposant que la variable dépendante suit une distribution de « Poisson ». Cela signifie que la probabilité de chaque résultat est déterminée par une moyenne liée aux prédicteurs par une fonction log-linéaire. Quant aux effets fixes, ils

⁶ Le codage n'a pas fait l'objet d'un accord interjuge, mais toutes les catégories utilisées dans le cadre de mon mémoire (variables 2.2.1 à 2.2.7) ont été validées par la direction de recherche.

représentent les contributions des variables explicatives observables et les effets aléatoires, eux, servent à modéliser les variations non expliquées au sein des groupes (Winter et Bürkner, 2021).

Ainsi, le modèle mixte généralisé de type « Poisson » est une approche statistique utile pour analyser des données de comptage tout en tenant compte des effets fixes et aléatoires. Ce modèle est également pertinent pour des études longitudinales lors d'analyses de mêmes unités, comme c'est le cas dans notre étude.

CHAPITRE 3

RÉSULTATS

Les résultats sont présentés au sein des quatre sections qui suivent. La section 3.1 présente les analyses descriptives, offrant un aperçu général des données. La section 3.2 présente les résultats du modèle statistique utilisé, montrant l'effet des variables présentées précédemment (cf. section 1.3) sur la production lexicale chez l'enfant. Enfin, la section 3.3 inclut les résultats d'interactions du modèle entre différentes variables.

3.1 Les analyses descriptives

Comme mentionné précédemment (cf. 2.2), 419 noms différents ont été analysés dans le corpus. Les données produites par les enfants représentent, quant à elles, 3028 occurrences de ces noms. Le tableau 12 présente le nombre d'occurrences de noms et d'énoncés pour chaque enfant.

Tableau 3.1 Répartition du nombre d'occurrences et d'énoncés produits par enfant

Enfants	Nombre d'occurrences	Nombre d'énoncés produits
Adrien	263 noms	5184 énoncés produits
Anae	562 noms	6684 énoncés produits
Anaïs	194 noms	3683 énoncés produits
Anne	269 noms	2966 énoncés produits
Antoine	397 noms	6384 énoncés produits
Léonard	331 noms	4483 énoncés produits
Marie	239 noms	2572 énoncés produits
Pauline	126 noms	2231 énoncés produits
Théophile	369 noms	5671 énoncés produits
Théotime	278 noms	3176 énoncés produits

Ce tableau met en évidence une forte variabilité dans le nombre total d'énoncés produits par les enfants, qui varie de 2231 pour Pauline à 6684 pour Anae. En revanche, le nombre de noms utilisés reste dans un intervalle plus restreint, allant de 126 pour Pauline à 562 pour Anae. Ainsi, bien que la quantité d'énoncés produits diffère considérablement, l'écart dans l'usage des noms est bien moins marqué.

Concernant le tableau 13, il illustre la distribution du nombre d'occurrences de noms et d'énoncés produits en fonction des groupes MLU des enfants.

Tableau 3.2 Répartition du nombre d'occurrences et d'énoncés produits par groupe MLU

Groupe MLU	Nombre d'occurrences	Nombre d'énoncés produits
Groupe MLU 2	37 noms	1046 énoncés produits
Groupe MLU 2,5	55 noms	803 énoncés produits
Groupe MLU 3	426 noms	7225 énoncés produits
Groupe MLU 3,5	524 noms	6910 énoncés produits
Groupe MLU 4	752 noms	10554 énoncés produits
Groupe MLU 4,5	431 noms	5629 énoncés produits
Groupe MLU 5	280 noms	4072 énoncés produits
Groupe MLU 5,5	164 noms	2508 énoncés produits
Groupe MLU 6	308 noms	3735 énoncés produits
Groupe MLU 7	51 noms	552 énoncés produits

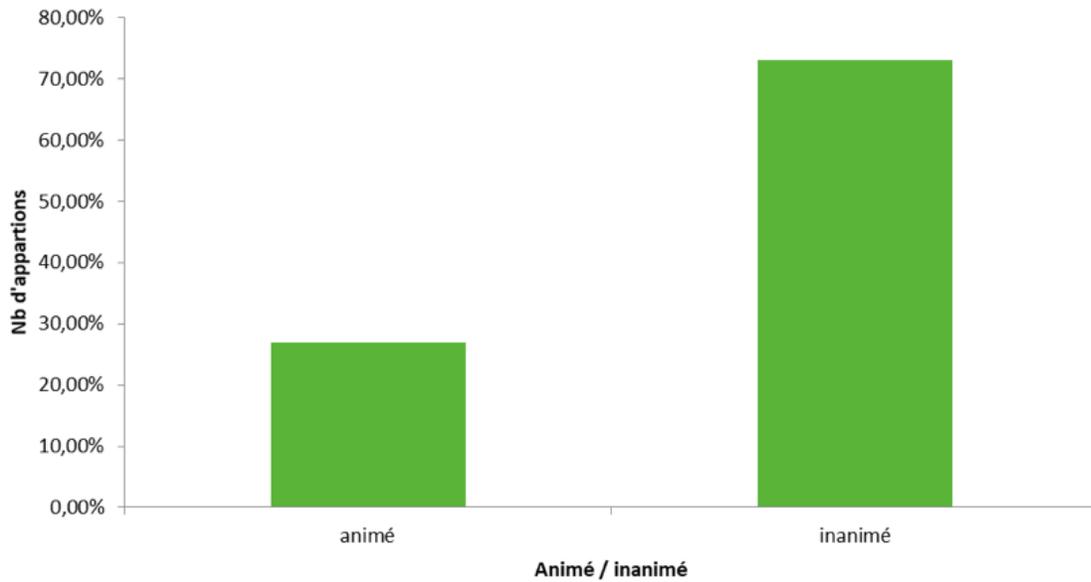
Comme évoqué précédemment (cf. 1.2.1), plusieurs facteurs ont été considérés pour expliquer la production des noms chez l'enfant : le caractère animé/inanimé des référents associés aux noms, leur concrétude, la longueur du nom, la valence émotionnelle portée par les noms et le niveau de développement langagier des enfants. Rappelons que dans notre travail, nous nous interrogeons sur l'influence potentielle de ces variables sur la production des noms. Ainsi, nous allons regarder la distribution des noms dans notre corpus en fonction de chacun de ces facteurs, d'abord pris séparément.

3.1.1 Les caractéristiques sémantiques du nom

3.1.1.1 Le caractère animé/inanimé des référents associés aux noms

Concernant la variable du caractère animé/inanimé des référents, on constate que les noms inanimés sont plus nombreux que les noms animés. Parmi les 3028 occurrences de noms, 2324 sont identifiés comme des noms inanimés contre seulement 704 comme des noms animés.

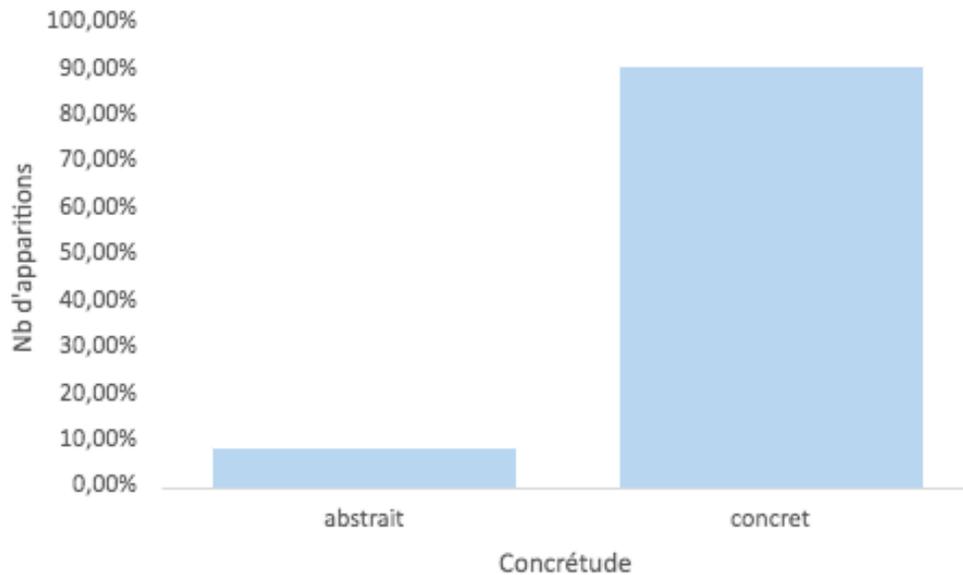
Figure 3-1 Répartition des occurrences des noms en fonction du caractère animé ou inanimé des référents



3.1.1.2 La concrétude

Pour la concrétude, la distribution est similaire c'est-à-dire que les noms renvoyant à des référents concrets sont largement majoritaires par rapport aux noms renvoyant à des référents abstraits. Les noms concrets représentent 90,52 % des données alors que les noms abstraits ne représentent que 9,48 %. En comparaison, sur les 3028 occurrences produites par les enfants, 2741 sont des noms concrets contre seulement 287 qui sont abstraits.

Figure 3-2. Répartition des occurrences des noms en fonction de la concrétude



3.1.2 La longueur du nom

Comme vu précédemment (cf. 2.2.2), la variable relative à la longueur du mot couvre quatre types de noms distincts : les noms monosyllabiques, les noms bisyllabiques, les noms trisyllabiques et les noms quadrisyllabiques. Les noms monosyllabiques et bisyllabiques sont largement majoritaires, représentant respectivement 48,94 % et 44,88 % des données. À l'inverse, les noms trisyllabiques et quadrisyllabiques sont minoritaires, ne représentant que 5,78 % et 0,40 % des occurrences du corpus. En valeurs absolues, 1482 noms sont monosyllabiques, 1359 bisyllabiques, 175 trisyllabiques et enfin seulement 12 sont quadrisyllabiques. De plus, il est intéressant de noter que ces différents types de noms produisent également une quantité variée d'énoncés. Les noms monosyllabiques ont généré 42782 énoncés, tandis que ceux bisyllabiques en ont produit 42825, ce qui montre leur utilisation fréquente dans la langue. En revanche, les noms trisyllabiques ont engendré 35473 énoncés, et les noms quadrisyllabiques, bien que limité en nombre, ont contribué avec 6329 énoncés. Cette répartition souligne la prépondérance quantitative des noms monosyllabiques et bisyllabiques.

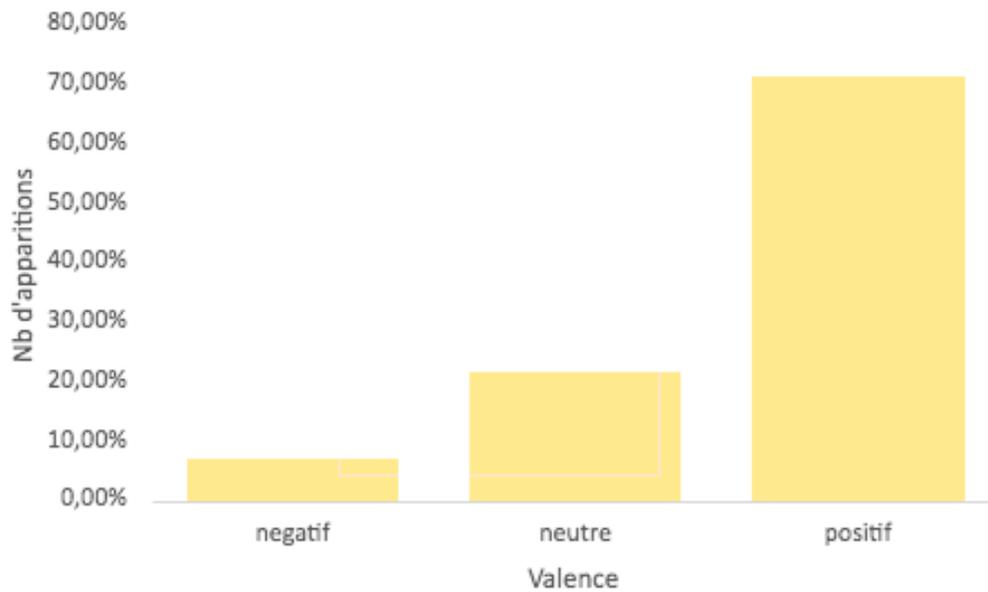
Tableau 3.3 Répartition des occurrences selon leur longueur (en nombre de syllabes)

Types de noms	Nombre de noms	Pourcentage	Nombre d'énoncés produits
Monosyllabique	1482 noms	48,94 %	42782 énoncés produits
Bisyllabique	1359 noms	44,88 %	42825 énoncés produits
Trisyllabique	175 noms	5,78 %	35473 énoncés produits
Quadrisyllabique	12 noms	0,40 %	6329 énoncés produits

3.1.3 La valence émotionnelle

Comme souligné au chapitre précédent (cf. section 2.2.4, du chapitre 2), les informations relatives à la valence émotionnelle proviennent de la norme ValEmo. Les noms relevés à la fois dans la norme ValEmo et dans les corpus ne se distribuent pas de façon équilibrée. Autrement dit, les noms à valence positive sont nettement plus nombreux que ceux à valence neutre, et davantage encore que ceux à valence négative. Les noms à valence positive représentent 68,10 % des occurrences de noms, soit 1482 noms, les noms à valence neutre, 24,74 % (749 noms) et les noms à valence négative, seulement 7,17 %, soit 217 noms.

Figure 3-3. Répartition des occurrences des noms selon leur valence émotionnelle



La disproportion entre valence positive, neutre et négative n'est pas surprenante considérant le déséquilibre déjà présent dans la norme ValEmo, utilisée pour l'étude. Cette dernière compte en effet un nombre plus élevé de noms à valence positive que de noms à valence neutre et négative (cf. section 1.1.3).

3.1.4 Le niveau de développement langagier

Comme mentionné plus haut (cf. 1.2.1), les séances des corpus observés ont été classées selon dix groupes de niveau de développement langagier (MLU 2 ; 2,5 ; 3 ; 3,5 ; 4 ; 4,5 ; 5 ; 5,5 ; 6 ;7). Le tableau suivant présente la répartition des occurrences de noms ainsi que leur pourcentage pour chaque sous-groupe MLU.

Tableau 3.4 Répartition des occurrences de noms selon la longueur moyenne des énoncés

Groupe MLU	Nombre d'occurrences
Groupe MLU 2	1,22 % (37 noms)
Groupe MLU 2,5	1,82 % (55 noms)
Groupe MLU 3	14,07 % (426 noms)
Groupe MLU 3,5	17,31 % (524 noms)
Groupe MLU 4	24,83 % (752 noms)
Groupe MLU 4,5	14,23 % (431 noms)
Groupe MLU 5	9,25 % (280 noms)
Groupe MLU 5,5	5,42 % (164 noms)
Groupe MLU 6	10,17 % (308 noms)
Groupe MLU 7	1,68 % (51 noms)

3.2 Les analyses du modèle sans interactions

Dans cette section, nous présentons les résultats obtenus à partir du modèle mixte généralisé de type « Poisson » (cf. 2.2.8) que nous avons utilisé pour analyser nos données. Ce modèle a convergé avec succès, confirmant sa validité pour notre analyse. Le coefficient R-carré (coefficient de détermination) obtenu est de 0.228450, ce qui indique une performance acceptable pour ce modèle (en ANNEXE B). Les sections qui vont suivre répondent aux questions de recherche qui concernent l'influence des variables considérées dans notre étude sur la production des noms. Nous présentons les résultats de régression montrant l'effet de nos variables indépendantes sur le nombre d'apparitions du nom chez l'enfant.

Les résultats sont présentés dans le tableau 15 ci-dessous.

Tableau 3.5 Distribution des variables dans le modèle sans interaction

Fixed Effect Omnibus tests			
Variables	X²	df	p
Groupe MLU	25.3250509	1.0000	< .001
Valence	0.8638316	1.0000	0.353
Fréquence_Antidote	2.7162853	1.0000	0.099
Concrétude	1.7187142	1.0000	0.190
Animé/inanimé	4.2068864	1.0000	0.040
Longueur (nbr syllabes)	2.0789744	1.0000	0.149
Occurrences adultes	349.3441524	1.0000	< .001
Sexe	0.0039592	1.0000	0.950

Les résultats de l'analyse révèlent que trois variables se distinguent par leur contribution significative à la variation du modèle, mesurée par le test chi-carré (X^2). En particulier, la variable des occurrences adultes présente un X^2 très élevé de 349.3441524 avec une valeur de p inférieure à 0.001, indiquant qu'elle contribue de manière significative et prépondérante à la variation de la variable dépendante (VD). De même, la variable du groupe MLU affiche un X^2 de 25.3250509 avec une valeur p également inférieure à 0.001, soulignant son importance dans la manière dont les données se distribuent. La variable animé/inanimé possède quant à elle un X^2 de 4.2068864 avec une valeur de p de 0.040, ce qui la rend significative malgré une contribution relativement moindre par rapport aux deux autres premières variables. Ainsi, ces trois variables — occurrences adultes, groupe MLU et caractère animé/inanimé — jouent un rôle clé et significatif dans la variation du modèle, comme en témoignent leurs valeurs de X^2 respectives.

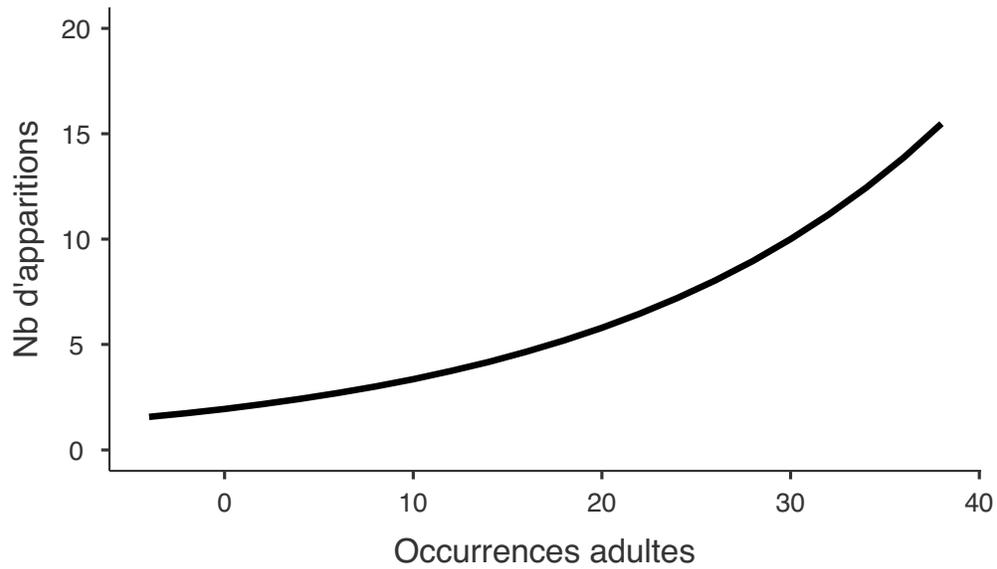
Les sections suivantes détaillent l'effet de chacune de nos variables qui s'avèrent significatives.

3.2.1 L'effet de la variable *occurrences adultes*

Comme évoqué précédemment (cf. 3.2), la régression montre un effet de la variable *occurrences adulte* : $X^2 = 349.3441524$ $p < 0.001$. Cela signifie que la variable occurrences adulte a un effet significatif sur la variable du nombre d'apparitions du mot chez l'enfant.

La figure 3.4 montre une corrélation positive entre le nombre d'apparitions du mot chez l'enfant et les occurrences produites par les adultes. Autrement dit, lorsque le nombre d'occurrences chez l'adulte augmente alors le nombre d'apparitions d'un nom chez l'enfant augmente également.

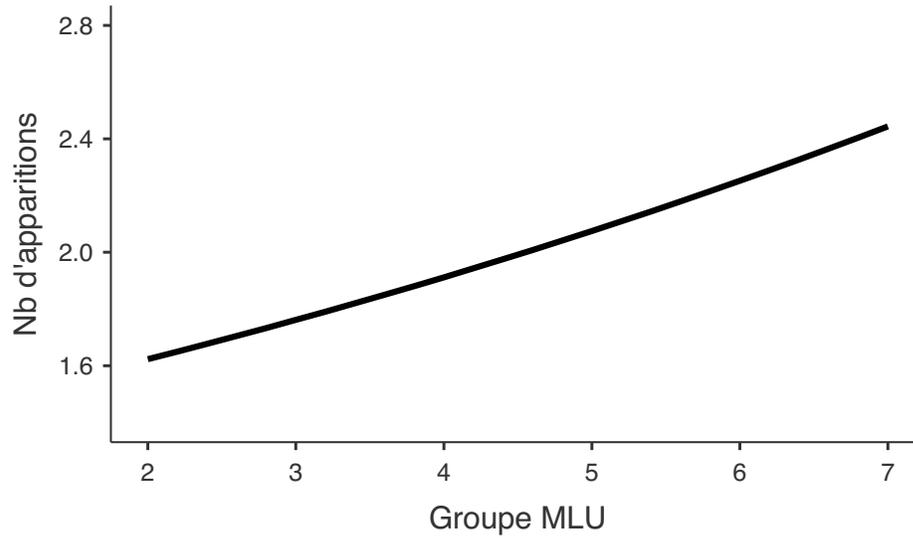
Figure 3-4. Nombre d'apparitions en fonction des occurrences adultes



3.2.2 L'effet du niveau de développement langagier (groupe MLU)

La régression indique un effet significatif de la variable groupe MLU : $X^2 = 25.3250509$ $p < 0.001$. La figure 5 ci-dessous présente la distribution du nombre d'apparitions des noms dans le discours des enfants, en fonction du groupe MLU. Conformément à ce qui pouvait être envisagé, le nombre d'apparitions augmente à mesure que le niveau de développement langagier augmente.

Figure 3-5. Nombre d'apparitions en fonction du niveau de développement langagier (groupe MLU)

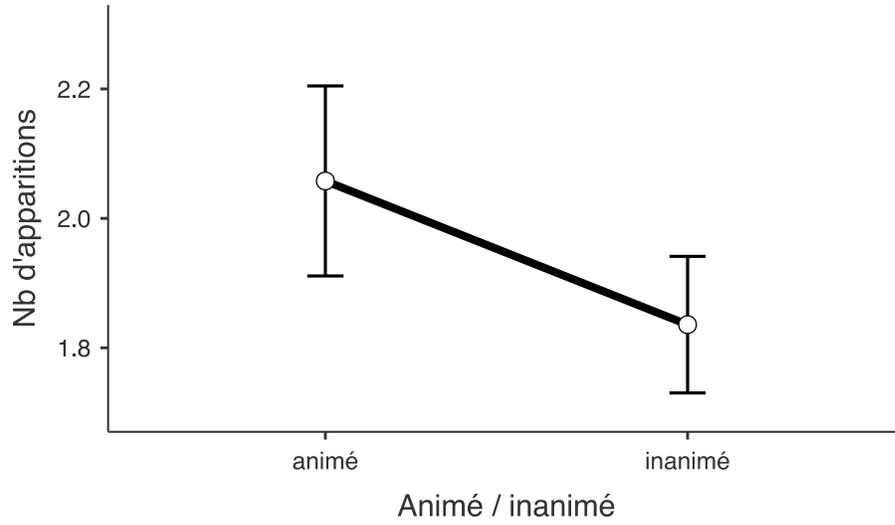


3.2.3 L'effet du caractère animé/inanimé du référent associé au nom

La régression montre un effet moindre par rapport aux deux autres variables, mais cette variable est tout de même associée à une valeur significative : $X^2 = 4.2068864$ $p < 0.040$. Cela montre que la variable du caractère animé/inanimé a un effet significatif sur la variable du nombre d'apparitions du nom chez l'enfant.

La figure 3.6 présente le nombre d'apparitions du nom en fonction du caractère animé/inanimé du référent associé au nom. Nous pouvons noter que les noms animés présentent un nombre d'apparitions plus élevé que les noms inanimés.

Figure 3-6. Moyenne du nombre d'apparitions en fonction du caractère animé/inanimé



3.2.4 L'effet des variables aléatoires

Dans cette section consacrée à l'effet des variables aléatoires, nous examinons l'influence de l'enfant (participant) et du nom (parmi les 419 noms retenus) sur la production lexicale. Pour cela, nous avons analysé le coefficient de corrélation interclasse (ICC), qui mesure la part de variance attribuable à chaque variable aléatoire.

Les résultats du tableau 16 montrent que l'ICC pour les noms (items) est de 0.131756, ce qui indique qu'il existe des différences entre les noms en termes de fréquence d'apparition. Certains noms sont donc produits plus fréquemment que d'autres, en tant que mot, par les enfants. En revanche, l'ICC pour les participants (enfant) est de 0.021031, suggérant que la variation entre les enfants est relativement faible. Autrement dit, tous les enfants présentent un nombre d'apparitions de noms relativement équivalent, et les différences observées sont principalement dues aux caractéristiques des noms eux-mêmes plutôt qu'aux différences entre les participants. Ainsi, les résultats indiquent que les variations dans le nombre d'apparitions sont principalement déterminées par les variables indépendantes, tandis que les variables aléatoires n'exercent qu'une influence marginale.

Tableau 3.6 Distribution des variables aléatoires dans le modèle sans interaction

Random Components

Groups	Name	SD	Variance	ICC
Items	(Intercept)	0.250673	0.0628370	0.131756
Enfant	(Intercept)	0.094317	0.0088956	0.021031
Residuals		1.000000	1.000000	

3.3 L'analyse du modèle avec les interactions

Comme indiqué précédemment (cf. 3.2), nous avons utilisé le modèle mixte généralisé de type « Poisson » pour notre analyse. Ce modèle a également convergé avec les interactions, et le coefficient R-carré de 0.236046 indique une performance acceptable (en ANNEXE C). Dans cette section, nous présentons les résultats des analyses avec interactions, de la variable du niveau de développement langagier (groupe MLU). Après avoir effectué des tests avec les variables significatives, c'est la variable du niveau de développement langagier (cf.1.2.2) qui a révélé le plus de résultats probants. Ainsi, elle sera mise en interaction avec d'autres variables afin d'examiner comment ces interactions influencent le nombre d'apparitions du nom chez l'enfant.

Trois variables se distinguent par leur contribution significative, mesurée par le chi-carré (X^2). La variable du niveau de développement langagier (groupe MLU) présente un X^2 de 18.8325922, $p < 0.01$, la variable *occurrences adulte* affiche un X^2 de 351.4954108, $p < 0.01^7$, et la variable du caractère animé/inanimé est associée à une valeur significative : $X^2 = 5.2735774$, $p < 0.022$.

⁷ Malgré la significativité de la variable *occurrences adultes*, les interactions entre cette variable et les autres variables n'ont pas révélé d'informations pertinentes pour notre travail

3.3.1 L'interaction entre le niveau de développement langagier (groupe MLU) et le caractère animé/inanimé des noms

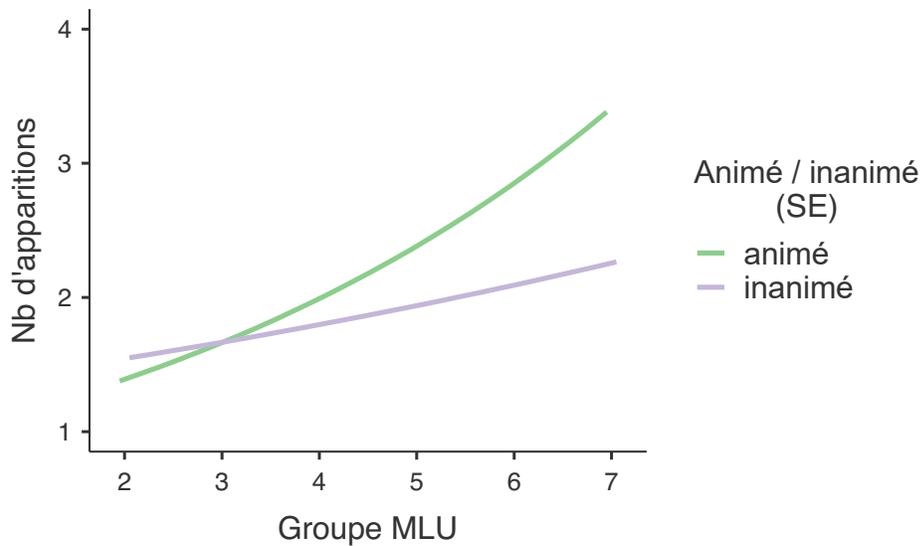
Comme le montre le tableau 16 ci-dessous, l'interaction entre les variables *niveau de développement langagier* (groupe MLU) et *caractère animé/inanimé* est significative avec un X^2 de 11.4250974, $p < 0.01$.

Tableau 3.7 Distribution des variables dans le modèle avec les interactions

Fixed Effect Omnibus test

	X^2	df	p
Groupe MLU	18.8325922	1.0000	< .001
Valence	1.0494838	1.0000	0.306
Fréquence_Antidote	0.5986918	1.0000	0.439
Concrétude	1.3344005	1.0000	0.248
Animé/inanimé	5.2735774	1.0000	0.022
Longueur (nbr syllabes)	6.0171968	1.0000	0.014
Occurrences adultes	351.4954108	1.0000	< .001
Sexe	0.0059905	1.0000	0.938
Groupe MLU * Animé/inanimé	11.4250974	1.0000	< .001
Groupe MLU * Concrétude	0.6143618	1.0000	0.433
Groupe MLU * Valence	0.4749077	1.0000	0.491
Groupe MLU * Fréquence_Antidote	0.0427098	1.0000	0.836
Groupe MLU * Longueur (nbr syllabes)	4.2745011	1.0000	0.039

Figure 3-7. Moyenne du nombre d'apparitions en fonction du niveau de développement langagier (groupe MLU) et du caractère animé/inanimé



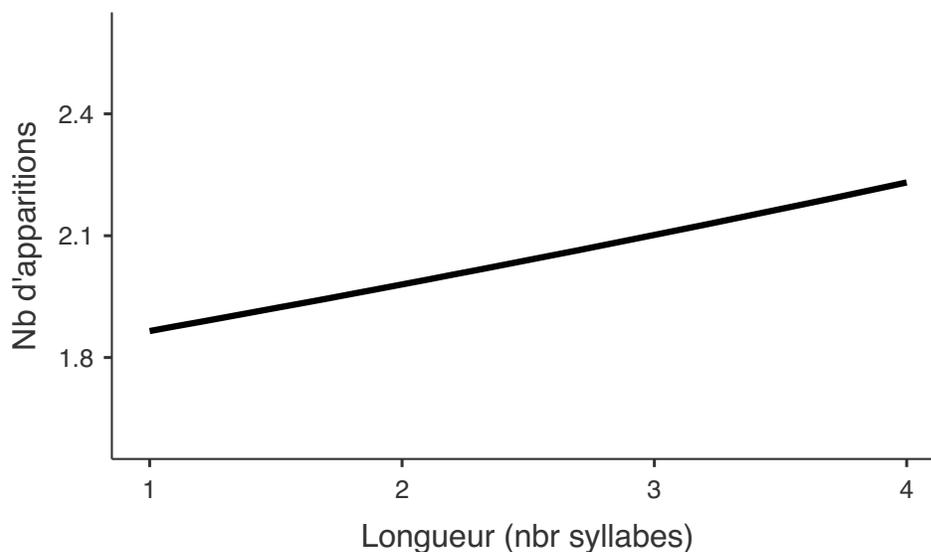
La figure 3.7 illustre le nombre d'apparitions des noms selon le niveau de développement langagier considéré (groupe MLU) et le caractère animé/inanimé des noms. Qu'il s'agisse des noms animés ou inanimés, les résultats révèlent que, le nombre d'apparitions augmente lorsque le MLU augmente. Plus précisément, en deçà d'un MLU de 3, les noms inanimés sont produits plus fréquemment que les noms animés. En revanche, lorsque l'enfant atteint un MLU de 3, les noms animés augmentent significativement.

3.3.2 L'interaction entre le niveau de développement langagier (groupe MLU) et la longueur des noms

Pour rappel, les résultats présentés dans le tableau 6 ont montré que l'interaction entre les variables *niveau de développement langagier* (groupe MLU) et *longueur du nom* est significative avec un X^2 de 4.2745011, $p < 0.039$.

La figure 3.8 qui présente le nombre d'apparitions du nom en fonction de la longueur du nom montre que le nombre d'apparitions augmente en fonction du nombre de syllabes.

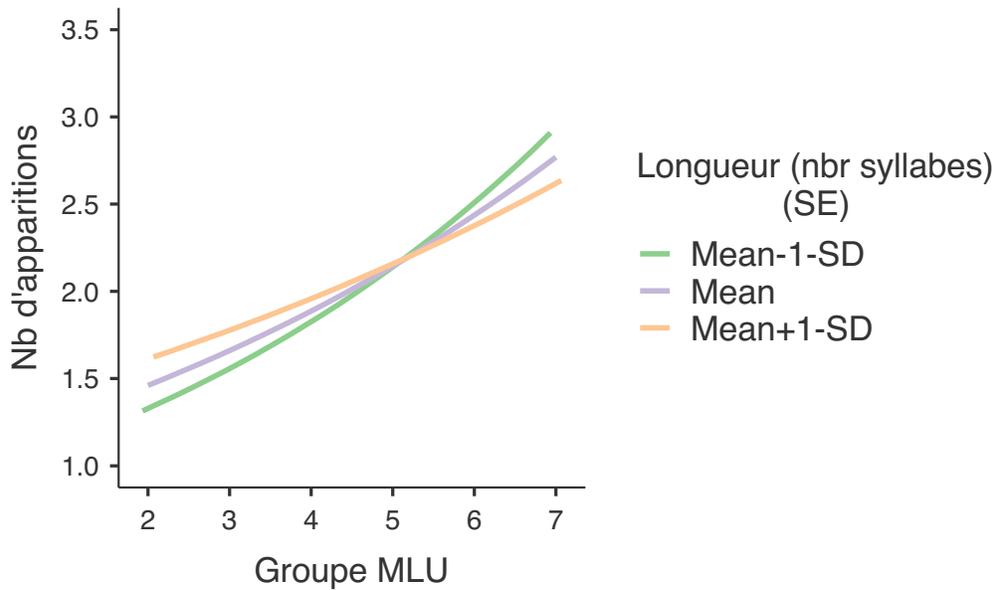
Figure 3-8. Nombre d'apparitions du nom en fonction de la longueur du nom



La figure 3.9 représente le nombre d'apparitions selon le niveau de développement langagier et la longueur du nom⁸. À la lecture du graphique, on observe que plus le MLU augmente, plus les noms longs sont fréquents.

⁸ Dans le cadre de notre travail, nous avons mesuré le nombre de syllabes des cibles et non le nombre de syllabes des productions.

Figure 3-9. Moyenne du nombre d'apparitions en fonction du niveau de développement langagier (groupe MLU) et de la longueur du nom



Afin d'illustrer ces résultats, nous avons élaboré le tableau 18, dans lequel deux exemples de noms sont présentés pour chaque enfant, en distinguant les noms les plus courts et les plus longs. Ces noms ont ensuite été associés à leur groupe MLU. Ce tableau reflète la tendance observée dans les résultats, à savoir que plus le groupe MLU augmente plus fréquents sont les noms longs.

Tableau 3.8 Relations entre le niveau de développement langagier (groupe MLU) et la longueur des noms

Enfant	Groupe MLU	Nom	Longueur du nom
Adrien	3	eau	monosyllabique
Adrien	5,5	hélicoptère	quadrisyllabique
Anae	3,5	boîte	monosyllabique
Anae	4	médicament	quadrisyllabique
Anaïs	3	biche	monosyllabique
Anaïs	3	thermomètre	trisyllabique
Anne	3,5	bain	monosyllabique
Anne	5	marelle	bisyllabique
Antoine	4	rat	monosyllabique
Antoine	6	bibliothèque	quadrisyllabique
Léonard	3	œuf	monosyllabique
Léonard	4,5	silence	bisyllabique
Marie	4	corde	monosyllabique
Marie	5,5	collier	bisyllabique
Théophile	5	poing	monosyllabique

Théophile	7	vérité	trisyllabique
Théotime	3,5	orteil	bisyllabique
Théotime	4	locomotive	quadrisyllabique

3.3.3 L'effet des variables aléatoires

Les résultats obtenus (cf. tableau 18) pour l'effet des variables aléatoires présentent des similitudes avec les résultats observés dans la partie précédente (cf. 3.2.4). Pour rappel, nos variables aléatoires sont l'enfant et le nom. Nous avons examiné le coefficient de corrélation interclasse (ICC) pour ces variables. Pour la variable du nom, l'ICC est de 0.134744, ce qui indique que certains noms apparaissent plus fréquemment que d'autres. Tandis que pour la variable de l'enfant, l'ICC est de 0.022785 suggérant que la fréquence de production de noms par les enfants est relativement homogène. En d'autres termes, les enfants ne diffèrent pas de manière significative concernant le nombre d'apparitions des noms. Ainsi, les variations dans le nombre d'apparitions sont principalement influencées par les variables indépendantes, et non par les variables aléatoires.

Tableau 3.9 Distribution des variables aléatoires dans le modèle avec les interactions

Random Components

Groups	Name	SD	Variance	ICC
Items	(Intercept)	0.253937	0.0644839	0.134744
Enfant	(Intercept)	0.098259	0.0096548	0.022785
Residuals		1.000000	1.000000	

CHAPITRE 4

DISCUSSION

Dans ce travail, nous avons observé comment différentes variables, telles que les caractéristiques sémantiques du nom (le caractère animé/inanimé, concret/abstrait du nom), sa longueur, sa fréquence, la valence émotionnelle qui y est associée, de même que le nombre d'occurrences du nom produites par les adultes, le niveau de développement langagier des enfants ainsi que le genre des enfants, sont susceptibles d'influencer l'acquisition du nom. Les sections qui suivent mettent en perspective les résultats que nous avons obtenus avec les données de la littérature et offrent différentes pistes de réflexion.

4.1 Les variables qui influencent l'acquisition du nom chez l'enfant

Nous rappelons les questions de recherche auxquelles nous avons tenté de répondre au travers de ce mémoire. Nous nous interrogeons tout d'abord sur l'influence du langage adulte sur les productions des enfants : en quoi le discours des adultes peut-il affecter l'acquisition des noms chez l'enfant ? Dans cette partie, nous allons examiner les variables qui se sont avérées pertinentes lors des résultats telles que la variable *occurrences adultes*, la variable *niveau de développement langagier* et la variable *caractère animé/inanimé*.

4.1.1 L'effet de la variable *occurrences adultes* sur le nombre d'apparitions du nom chez l'enfant

Comme constaté plus haut, la variable *occurrences adulte* a une influence sur le nombre d'apparitions du nom chez l'enfant (cf. 3.2.1). Ce résultat permet de répondre en partie à notre deuxième question de recherche : la variable *occurrences adultes* est un des facteurs explicatifs de la production du nom chez l'enfant. Cependant, bien que cette corrélation soit pertinente, elle ne permet pas de conclure directement à une relation causale entre l'usage des noms par les adultes et leur production par les enfants. En d'autres termes, si la fréquence d'utilisation d'un nom par l'adulte semble favoriser son usage par l'enfant, cela ne signifie pas nécessairement que cet usage par les adultes cause directement la production du nom par les enfants. Les enfants peuvent également orienter les échanges en fonction de leurs propres intérêts, ce qui pourrait également contribuer à la fréquence d'apparition de certains noms dans le discours. Cette interaction bidirectionnelle entre adultes et enfants a été soulignée par plusieurs chercheurs, comme Vygotsky (1978), qui met en avant le rôle actif de l'enfant dans la construction du discours. De plus, les travaux de Tomasello (2003) soutiennent l'idée d'une construction linguistique basée sur l'usage et non sur une transmission linéaire, où l'enfant joue un rôle central dans la direction des échanges, influençant ainsi la sélection des noms utilisés.

Nos résultats s'inscrivent néanmoins dans la continuité des travaux existants sur l'acquisition du langage à partir de l'expérience. Hoff et al. (2002) ont montré que l'enfant possède un accès direct au langage grâce aux échanges quotidiens avec son entourage. Ce processus d'apprentissage basé sur l'expérience est conditionné par la fréquence des formes produites dans le discours adulte auquel l'enfant est exposé quotidiennement. La variable *occurrences adultes* souligne l'importance de la continuité du discours entre adultes et enfants. En effet, les noms fréquemment mentionnés par les adultes sont plus susceptibles d'être repris par les enfants, ce qui crée une continuité dans le discours, mais aussi dans l'apprentissage lexical. Toutefois, il est essentiel de noter à nouveau que cette continuité ne résulte pas uniquement d'une influence unidirectionnelle des adultes sur les enfants. L'enfant joue un rôle actif en influençant également les thèmes de conversation et en orientant les échanges avec les adultes, ce qui pourrait moduler l'apparition de certains noms dans les interactions (Tomasello, 2003). Ainsi, les enfants ne sont pas simplement récepteurs passifs de l'input linguistique mais agissent activement sur la dynamique de la conversation, contribuant à l'usage des noms et à la construction du discours.

Dans ce cadre, les théories basées sur l'utilisation du langage dans l'interaction quotidienne (De Pontonx et al., 2019 ; Veneziano et al., 1990) indiquent que les enfants extraient des structures relativement simples du discours adulte, et que ces structures sont renforcées par les interactions répétées entre les interlocuteurs. Cette idée rejoint les travaux de De Villiers (1985), Naigles et Hoff-Ginsberg (1998), ou Theakston, Lieven, Pine et Rowland (2001), qui montrent que les premières formes verbales et syntaxiques acquises par les enfants sont souvent celles les plus fréquemment utilisées par leurs adultes. Cela suggère que les interactions entre les enfants et les adultes ne se limitent pas à une simple exposition, mais aident à renforcer l'acquisition des formes, des sons, du lexique, ou des structures — morphologiques, syntaxiques — des enfants. De ce point de vue, nos résultats confirment que les répétitions dans les échanges adultes-enfants jouent un rôle central dans l'acquisition lexicale et syntaxique. Comme l'a souligné Clark (2006), la répétition contribue à renforcer l'apprentissage de mots nouveaux, et cela semble également valable pour les noms : le discours adulte est riche en répétitions, facilitant ainsi la production des mêmes noms par les enfants. Il est d'ailleurs intéressant de noter que plusieurs modèles produits par les enfants sont déjà présents dans le discours adultes, illustrant ainsi l'interaction dynamique entre les deux groupes de locuteurs. Un exemple extrait de notre corpus montre cette dynamique : dans une séance, le père d'Adrien produit le nom « voiture », que l'enfant répète immédiatement après lui. L'exemple 1 illustre la répétition. Ce phénomène de répétition constitue un aspect important de l'acquisition du langage, mais il convient de souligner que cette répétition peut aussi être influencée par des facteurs interactifs et non uniquement par l'exposition à un discours adulte.

(1) Reprise d'un nom par l'enfant (corpus d'Adrien ; 4;03)

Père : parce_que moi je crois que le , la voiture verte je crois qu'elle veut aller à l'hôtel !

Adrien : celle-là ?

Père : ouais c'est c'est c'est celle-ci la verte ou pas ?

Père : non c'est laquelle la verte ?

Adrien secoue la tête négativement

Adrien : Oje sais pas si c'est celle-là .

Père : ben si c'est celle-ci !

Adrien : Oil Oy en a deux voitures rouges !

L'exemple 2 montre également ce phénomène. La mère d'Anae mentionne le nom « perroquet » à 2 reprises, et l'enfant le répète.

(2) Reprise d'un mot par l'enfant (corpus Anae ; 4;05)

Mère : ben pourquoi Oil y a deux perroquets ?

Anae : [pas kə] 'parce que'.

Anae : [j a sa mamã k ε ale l jεʁʁe] 'Oil y a sa maman qu(i) est allé le chercher'.

Mère : ah le perroquet aussi il a retrouvé sa maman.

Anae se retourne vers le chien.

Anae : que sa maman.

Mère : ah alors que le singe il a retrouvé qui ?

Anae : [tulmɔ̃d] 'tout_l(e)_monde'.

Mère : +< toute sa famille et qu'est-ce qu'i(ls) mangent pour fêter ça ?

Anae : [dε fɔwi] 'des fruits'.

Mère : des fruits.

Anae : et Oil y a aussi les perroquets .

Ces exemples mettent en évidence les résultats observés dans la section 3.2.1. À mesure que le nombre d'occurrences d'un nom chez l'adulte augmente, sa fréquence d'apparition chez l'enfant semble suivre la même trajectoire. Toutefois, comme évoqué précédemment, cette corrélation ne permet pas de conclure de manière définitive à une influence causale directe du lexique adulte sur l'acquisition des noms par les enfants. Il est tout à fait envisageable que l'utilisation de ces noms par les enfants ne soit pas uniquement le reflet d'un modèle adulte, mais aussi un phénomène dynamique où les enfants influencent à leur tour le discours. Par exemple, les enfants peuvent orienter les thèmes de conversation, incitant ainsi les adultes à utiliser certains noms (Tomasello, 2003). De plus, même dans les échanges entre adultes, on peut observer l'utilisation de noms communs, ce qui souligne la complexité des interactions et des facteurs en jeu. Ainsi, bien que les adultes jouent un rôle important dans la mise en œuvre de certains noms, les enfants ne se contentent pas de les reproduire passivement ; ils participent activement à la construction et à la circulation de ces noms au sein des échanges verbaux (Bergelson et Swingley, 2012 ; Snow, 2010).

4.1.2 L'effet de la variable *niveau de développement langagier* sur le nombre d'apparitions du nom chez l'enfant

Au vu des résultats obtenus concernant la variable liée au niveau de développement langagier (cf. 3.2.2), il apparaît qu'elle influence le nombre d'apparitions du nom chez l'enfant. Ce résultat permet lui aussi de répondre — au moins en partie — à notre première question de recherche : la variable du niveau de développement langagier est l'un des facteurs explicatifs de la production du nom chez l'enfant. Toutefois, il convient d'approfondir cette analyse afin d'examiner plus en détail la portée et les limites de cette variable dans le cadre de notre étude.

Pour rappel, le niveau de développement langagier est évalué à travers la longueur moyenne des énoncés en morphèmes (MLU), comme indiqué à la section 1.3.5. Cette mesure est non seulement corrélée à l'âge, mais également à la qualité et à la fréquence des interactions que les enfants entretiennent avec leur entourage (Hoff, 2006).

Les recherches antérieures (Le Normand et al., 2013) ont montré que la longueur moyenne des énoncés est une mesure fiable pour les enfants de moins de quatre ans, dans la mesure où leur contexte de

production est relativement contrôlé. De ce fait, les séances durant lesquelles les participants sont âgés de deux à quatre ans se situent dans cette tranche d'âge. Nos résultats confirment que les enfants dont le développement langagier est avancé sont en mesure de produire des noms en plus grande quantité et des noms plus longs. De plus, la longueur moyenne des énoncés est aussi impactée par la fréquence des interactions que l'enfant entretient avec les adultes. En effet, si les adultes interagissent régulièrement avec l'enfant en utilisant un lexique varié, par exemple des noms différents, cela contribue à renforcer le niveau de développement lexical de l'enfant (Le Normand et al., 2006 ; Rowe, 2012).

Cependant, une analyse critique de ces résultats s'impose. D'une part, bien que le MLU soit un bon indicateur du développement langagier, son interprétation devient plus complexe chez les enfants de plus de quatre ans. Dans notre corpus, certaines séances impliquent des enfants dépassant cet âge, ce qui soulève des interrogations quant à la fiabilité de cet indice dans ces cas spécifiques (cf.1.3.5). En effet, à partir de quatre ans, la structure grammaticale des énoncés se complexifie et le MLU ne reflète plus uniquement l'ampleur du lexique mais aussi des acquisitions lexicales plus élaborées (Klee et Fitzgerald, 1985). Il est donc possible que le MLU, dans ce contexte, ne capture pas pleinement les variations du développement lexical et surestime la corrélation entre la longueur des énoncés et l'utilisation des noms.

D'autre part, le MLU ne permet pas de distinguer clairement l'impact des interactions langagières sur l'acquisition des noms. Par exemple, un enfant exposé à un environnement langagier riche, mais dont la structure syntaxique des énoncés reste relativement simple, pourrait présenter un MLU plus faible alors que son lexique (de noms en particulier) est potentiellement étendu (Tomasello, 2003). Cette limitation souligne la nécessité de coupler le MLU avec d'autres indicateurs, tels que la diversité lexicale ou l'analyse qualitative des contextes d'utilisation des noms.

En conclusion, bien que nos résultats confirment l'importance du niveau de développement langagier dans la production des noms, il est essentiel de nuancer cette interprétation en tenant compte des limites du MLU, notamment pour les enfants plus âgés. Des études futures pourraient approfondir cette question en combinant plusieurs indices du développement lexical et en explorant plus finement l'influence des interactions adultes-enfants sur l'enrichissement du lexique (Snow, 1999).

4.1.3 L'effet de la variable *caractère animé/inanimé du nom* — sur le nombre d'apparitions du nom chez l'enfant

La variable du caractère animé/inanimé sur le nombre d'apparitions du nom chez l'enfant, déjà évoquée précédemment (cf. 3.2.3), apporte en partie une réponse à notre première question de recherche. En effet, cette variable est un des facteurs explicatifs de l'acquisition du nom chez l'enfant.

En effet, nos résultats confirment une tendance observée dans la littérature : avant 20 mois, les noms animés dominant dans le lexique de l'enfant (Kern, 2007). Cette observation peut s'expliquer par le fait que les noms animés, souvent associés à des individus ou des animaux, sont directement liés à des interactions sociales et émotionnelles, ce qui en fait des éléments privilégiés dans les premières étapes de l'acquisition lexicale. Cependant, à partir de 2 ans, nos résultats semblent s'écarter de certaines études, comme celles de Bassano (1998, 2005), qui suggèrent que les noms animés stagnent, tandis que les noms inanimés commencent à augmenter. Cette divergence pourrait être expliquée par la méthodologie de notre étude, qui a limité l'analyse aux noms présents à la fois dans ValEvo et dans nos corpus. Ce choix pourrait restreindre la diversité des noms étudiés et expliquer, en partie, l'écart avec les résultats rapportés par d'autres recherches.

De plus, il est essentiel de considérer d'autres facteurs contextuels qui peuvent influencer ces résultats. Par exemple, les types de jeux auxquels les enfants sont exposés peuvent avoir un impact sur la fréquence des noms animés ou inanimés. Si un enfant joue régulièrement avec des figurines ou des jouets inanimés, il est probable que cela favorise l'acquisition de noms inanimés (Kern, 2007). En outre, le style parental peut aussi jouer un rôle. Des parents qui utilisent un langage plus descriptif, orienté vers des actions impliquant des agents animés, pourraient naturellement encourager l'acquisition de noms animés (Tardif, 1996).

En résumé, bien que nos résultats s'inscrivent dans une tendance générale observée dans la littérature, plusieurs éléments méthodologiques et contextuels doivent être pris en compte pour expliquer les divergences avec d'autres recherches. Une analyse plus approfondie, prenant en compte ces variables, serait nécessaire pour mieux comprendre les mécanismes sous-jacents à l'acquisition du lexique chez l'enfant.

4.2 Les variables non pertinentes dans le cadre de notre étude

Notre étude avait pour objectif d'examiner huit variables caractérisant le nom : les caractéristiques sémantiques du nom (le caractère animé/inanimé et concret/abstrait du nom), sa longueur, sa fréquence, la valence émotionnelle qui y est associée, le nombre d'occurrences du nom produites par les adultes, le niveau de développement langagier des enfants et le genre des enfants. Les résultats obtenus montrent que seulement quatre de ces variables se sont révélées significatives.

Tout d'abord, les analyses sans tenir compte des interactions (cf. 4.1), ont révélé trois variables significatives : les variables *occurrences adultes*, *niveau de développement langagier* et *caractère animé/inanimé*. Ensuite, en intégrant les interactions dans le modèle (cf. 4.3), une quatrième variable est devenue significative : la variable *longueur du nom*. En revanche, les autres variables n'ont pas révélé d'effets significatifs dans notre étude. C'est le cas des variables de la fréquence, du caractère concret/abstrait, de la valence émotionnelle et du genre.

Concernant la fréquence, souvent considérée comme une variable centrale dans l'acquisition lexicale, il est important de souligner les limites que la littérature met en évidence à ce sujet. En effet, l'idée selon laquelle les mots plus fréquents sont acquis plus tôt par l'enfant n'est pas toujours confirmée de manière systématique. Par exemple, bien que les mots fréquemment utilisés dans le discours des adultes soient généralement acquis plus rapidement, cela ne s'applique pas toujours aux mots employés par les enfants. Certains termes très fréquents dans le discours adulte, comme les mots des classes fermées (articles, prépositions...) ne figurent pas nécessairement parmi les mots les plus courants dans le lexique de l'enfant (Goodman et al., 2008). Cela suggère que la fréquence d'un mot, en dehors du contexte d'utilisation spécifique par l'enfant, pourrait ne pas être un indicateur aussi déterminant qu'on pourrait le supposer, notamment au début du développement lexical. Cependant, les fréquences utilisées dans notre étude proviennent de corpus écrits (Antidote), qui ne reflètent pas nécessairement les usages oraux, encore moins ceux propres aux interactions adulte-enfant. Cette limite méthodologique invite à nuancer la portée de l'indice de fréquence dans le cadre de l'acquisition précoce du lexique.

En ce qui concerne les caractéristiques concret/abstrait, de nombreuses études suggèrent que les noms concrets sont généralement acquis plus tôt que les noms abstraits, en raison de leur lien plus direct avec

les expériences sensorielles (Bates et al., 1994). Cependant, dans notre étude, cette distinction ne semble pas jouer un rôle majeur, ce qui pourrait indiquer que d'autres facteurs, tels que certaines dimensions des interactions sociales – comme la fréquence et la qualité des échanges verbaux avec les adultes, les reformulations, les épisodes d'attention conjointe, ou encore les routines langagières partagées – ont un impact plus important que la nature des noms (concrets ou abstraits). Cette absence de différence pourrait également indiquer que, dans les premiers stades du développement lexical, les enfants privilégient des critères autres que la concrétude ou l'abstraction du nom, comme son utilité dans les échanges quotidiens (Tomasello, 2003).

De même, bien que la valence émotionnelle du nom soit théoriquement un facteur important, elle n'a pas montré d'effets significatifs. Des recherches antérieures indiquent que les enfants pourraient être plus enclins à apprendre des mots associés à des émotions fortes ou positives (Stern et al., 1985), mais nos résultats n'ont pas confirmé cette hypothèse. Il est possible que la valence émotionnelle des mots n'intervienne de manière décisive que dans des contextes particuliers, tels que des interactions émotionnellement intenses, ou à un stade plus avancé du développement langagier, lorsque les enfants deviennent plus conscients des dimensions affectives du langage. Par exemple, certains auteurs suggèrent que la charge émotionnelle d'un mot pourrait jouer un rôle plus déterminant lorsque l'enfant atteint un certain niveau de développement affectif.

Et enfin, concernant le genre, il est souvent suggéré que les filles et les garçons pourraient développer leur lexique de manière différente, en raison de facteurs biologiques, cognitifs et sociaux (Leaper, 2002). Toutefois, nos résultats n'ont révélé aucune différence significative liée au genre, ce qui laisse penser que d'autres variables, telles que l'environnement socio-éducatif ou la qualité des interactions adultes-enfants, pourraient avoir un impact plus important que le genre dans l'acquisition du lexique. Cette absence d'effet pourrait également refléter la complexité des interactions entre ces facteurs sociaux et linguistiques, soulignant ainsi que le genre seul ne suffit pas à expliquer les variations observées dans l'acquisition lexicale (Snow, 2010).

Ainsi, bien que certaines de ces variables (fréquence, caractère concret/abstrait, valence émotionnelle, genre) soient souvent citées comme des facteurs influençant l'acquisition lexicale, nos résultats suggèrent qu'elles ne jouent pas un rôle primordial à ce stade particulier du développement. Cela ne remet pas en question leur importance dans d'autres contextes ou à des étapes plus avancées du développement

langagier. Des recherches supplémentaires pourraient approfondir la manière dont ces variables interagissent avec des facteurs contextuels, sociaux et cognitifs.

4.3 Les variables qui influencent l'acquisition du nom chez l'enfant : interactions entre le niveau de développement langagier, le caractère animé/inanimé et la longueur du nom

En plus des questions évoquées précédemment (4.1), nous rappelons notre interrogation sur l'influence de la longueur du nom, facteur inclus dans notre étude. Nous nous demandons si cette variable joue un rôle dans l'acquisition des noms lors de productions précoces. A première vue, on pourrait supposer que la longueur des noms a un impact direct sur leur acquisition, en particulier dans les premières étapes du développement lexical, où la complexité phonologique et la mémorisation des mots peuvent influencer la production langagière (Dromi, 1996). Cependant, comme le montrent nos résultats, cette hypothèse n'a pas été confirmée par des preuves empiriques, ce qui soutient l'idée que d'autres facteurs jouent un rôle plus déterminant à ce stade de développement. Par ailleurs, certains chercheurs suggèrent que l'influence de la longueur des noms pourrait être plus marquée dans des contextes plus avancés de l'acquisition, lorsque les enfants commencent à produire des phrases plus complexes (Goldberg, 2007).

Comme vu plus haut (cf. 3.3), le niveau de développement langagier prime, et cela quelles que soient les caractéristiques du nom. Ce constat est appuyé par des études antérieures qui montrent que, avec la progression de la maîtrise du langage, la capacité générale à traiter et produire des structures plus complexes semble prédominer sur des aspects purement phonologiques ou lexicaux (Lecerf, 2011). En effet, les enfants ayant un développement langagier plus avancé démontrent une capacité à utiliser des noms, qu'ils soient longs ou courts, dans des contextes variés et avec une précision plus grande (Pine et al., 1999). Cela suggère que, bien que la longueur des noms puisse théoriquement influencer l'acquisition lexicale, son impact semble modéré, par rapport au stade de développement langagier de l'enfant.

Ainsi, dans cette deuxième partie de la discussion, nous nous concentrerons sur les variables significatives en relation avec le niveau de développement langagier.

4.3.1 L'effet de l'interaction entre le niveau de développement langagier et le caractère animé/inanimé du référent associé au nom

L'interaction entre le *niveau de développement langagier* et le *caractère animé/inanimé du nom*, que nous avons déjà mentionnée précédemment (cf. 3.3.1), montre une influence significative sur le nombre d'apparitions du nom chez l'enfant. En effet, le niveau de développement langagier, souvent mesuré par la longueur moyenne des énoncés (MLU), est un indicateur clé du développement lexical de l'enfant. Plusieurs études, comme celles de Parisse et Le Normand (2007), ont montré que l'augmentation du MLU est liée à l'élargissement du lexique de l'enfant. Par ailleurs, l'étude de Batista et Colleta (2011) suggère qu'à mesure que les énoncés deviennent plus complexes, les enfants ajoutent davantage de nouveaux « éléments », ce qui est souvent associé à l'enrichissement lexical.

Nos résultats suivent cette tendance générale : nous observons qu'à mesure que le MLU augmente, le nombre de noms produits par les enfants augmente également, peu importe s'ils sont animés ou inanimés. Cependant, une divergence importante se fait sentir à partir de 2 ans. Avant 18 mois, les enfants utilisent des noms animés et inanimés en proportions relativement similaires. Mais après 2 ans, une différence apparaît : la production de noms animés stagne, tandis que celle des noms inanimés commence à augmenter. Cette tendance est en accord avec les travaux de Bassano (1998, 2005), qui montrent une prédominance des noms animés dans les premières étapes du développement lexical, suivie par une augmentation de la production de noms inanimés vers 2 ans et demi. Cependant, nos résultats diffèrent de ceux de Bassano, qui suggère que l'augmentation du MLU devrait s'accompagner d'une augmentation continue des noms animés. Cette différence peut être expliquée par plusieurs facteurs. Premièrement, notre étude a limité l'analyse aux noms présents dans la norme ValEmo et dans nos corpus, ce qui pourrait avoir restreint la diversité des noms observés. De plus, il est possible que les facteurs contextuels, tels que la nature des interactions sociales — notamment la fréquence des échanges verbaux avec les adultes, les épisodes d'attention conjointe, les reformulations, ou encore la participation de l'enfant à des routines langagières et des jeux symboliques — jouent un rôle important dans l'acquisition et la production des noms.

La saturation du lexique en noms animés, observée avant l'âge de 2 ans, pourrait conduire à une augmentation des noms inanimés, favorisant ainsi une diversification lexicale (Bassano, 1998, 2005). Cette évolution pourrait également s'expliquer par l'influence des contextes éducatifs et familiaux, où les objets

inanimés sont fréquemment nommés (Bornstein et al., 2004). En effet, certaines études montrent que l'exposition à un environnement riche en interactions verbales influence directement la répartition des catégories lexicales acquises par l'enfant (Snow, 1995). De plus, la fréquence d'usage des noms inanimés dans le discours parental pourrait faciliter leur intégration et leur apprentissage progressif (Golinkoff et Hirsh-Pasek, 2006). Ces éléments suggèrent que l'environnement et les interactions jouent un rôle déterminant dans l'acquisition des noms, en particulier au moment où les enfants enrichissent et structurent davantage leur langage.

En résumé, bien que nos résultats confirment que le développement langagier (mesuré par le MLU) est lié à une augmentation des noms produits, des différences notables apparaissent dans la production des noms animés et inanimés, qui ne correspondent pas toujours à la littérature. Ces divergences soulignent la nécessité d'explorer davantage les influences contextuelles, ainsi que les variations méthodologiques.

4.3.2 L'effet de l'interaction entre le niveau de développement langagier et la longueur du nom

Comme mentionné plus haut, l'interaction entre les variables *niveau de développement langagier* et *longueur du nom* a une influence sur le nombre d'apparitions du nom chez l'enfant (cf. 3.3.2). Ce résultat est en accord avec plusieurs recherches précédentes. En effet, il est établi que les jeunes enfants produisent d'abord des noms courts, souvent monosyllabiques, dans les premières étapes de leur développement lexical. Gayraud et Kern (2007) soulignent que les premiers noms des enfants sont généralement courts, ce qui facilite leur production. À mesure que les enfants grandissent, leur lexique devient plus complexe et inclut des noms plus longs. Par exemple, à 46 mois, de nombreux enfants commencent à utiliser des mots trisyllabiques. Nos résultats renforcent cette observation : nous constatons qu'à mesure que le niveau de développement langagier (mesuré par le MLU) augmente, la longueur des noms produits par les enfants tend également à augmenter. Par exemple, Adrien produit le nom monosyllabique « ours » lorsque son MLU est de 3, puis produit le nom trisyllabique « kangourou » avec un MLU de 6. De même, Léonard passe de « pipe », un nom monosyllabique, à « tonneau », un nom bisyllabique, en augmentant son MLU de 3 à 4 (cf. 3.3.2 ; tableau 19).

Cependant, cette relation entre le MLU et la longueur des noms mérite une analyse plus nuancée. Bien que nous observions une tendance générale où l'augmentation du MLU s'accompagne d'une utilisation de noms plus longs, cette relation n'est pas nécessairement uniforme. Autrement dit, l'augmentation du MLU ne se traduit pas toujours par une augmentation directe de la longueur des noms produits. D'autres facteurs, tels que le développement des compétences grammaticales (par exemple, l'utilisation de mots grammaticaux), peuvent également influencer la complexité des énoncés des enfants (Bates et Goodman, 1997).

Il est aussi possible que des facteurs contextuels, comme les interactions sociales et les jeux dans lesquels les enfants sont impliqués, jouent un rôle clé. Par exemple, un enfant qui joue avec des jouets ou participe à des activités qui incluent des objets ou des animaux pourrait être plus exposé à des mots plus longs (Hoff, 2006). Cela suggère que l'environnement dans lequel l'enfant évolue peut influencer la production de mots plus longs, et non seulement le développement langagier en lui-même.

En conclusion, bien que nos résultats confirment une relation entre le développement langagier (mesuré par le MLU) et la longueur des noms, cette relation semble être influencée par d'autres facteurs.

4.4 Les variables aléatoires

Pour rappel, dans notre étude, nous avons pris en compte deux variables aléatoires : le participant (enfant) et le nom (parmi les 419 noms retenus). Les résultats révèlent des différences entre les noms en termes de fréquence d'apparition, certains étant produits plus fréquemment que d'autres. Cette différence peut être expliquée par plusieurs facteurs. D'une part, certains enfants produisent un plus grand nombre de noms que d'autres, ce qui crée une variation entre les participants. D'autre part, la fréquence d'apparition de ces noms dans l'environnement quotidien des enfants joue également un rôle important. En effet, les enfants ont plus de chances de reprendre les noms qu'ils entendent souvent dans leurs interactions avec les adultes, à la maison, en service de garde ou à l'école. Cette idée est soutenue par des travaux de Snow (2010), qui montrent que la quantité et la qualité de l'exposition linguistique jouent un rôle clé dans le développement lexical des enfants.

Cependant, la variation entre les enfants est relativement faible dans notre étude. En effet, tous les participants ont produit un nombre similaire de noms. Cela contraste avec ce que l'on observe souvent dans la littérature, où des études montrent une plus grande variation entre les enfants. Par exemples, plusieurs recherches indiquent que le développement lexical des enfants peut être très différent d'un individu à l'autre, influencé par des facteurs tels que l'environnement social, les interactions adultes-enfants et même les caractéristiques individuelles des enfants (Hoff, 2006). Cette faible variation dans notre étude pourrait s'expliquer par plusieurs raisons. D'abord, l'échantillon de participants est relativement petit (seulement dix enfants), ce qui peut limiter la diversité des résultats observés. De plus, il est possible que les enfants viennent d'environnements sociaux ou culturels similaires, ce qui pourrait réduire les différences dans leurs productions lexicales. En outre, les critères de sélection des noms pourraient aussi limiter cette variation, si les noms choisis ne couvrent pas suffisamment de contextes d'utilisation différents. Cette homogénéité de l'échantillon est également soulignée par des chercheurs comme Bates et al. (1994), qui mettent en avant l'importance de la diversité de l'échantillon pour mieux comprendre la variabilité dans l'acquisition lexicale.

Il est également important de noter que notre étude se concentre principalement sur la fréquence des noms produits par les enfants. Or, cette fréquence pourrait ne pas refléter l'ensemble de l'exposition de l'enfant aux noms. Par exemple, un enfant peut entendre un nom fréquemment sans forcément le produire, ou au contraire, il peut produire un nom qu'il entend rarement. Une étude prenant en compte non seulement la production, mais aussi la compréhension aux mots entendus, pourrait fournir une image plus complète de l'acquisition lexicale. Cette distinction entre production et réceptivité a été abordée par Tomasello (2003), qui suggère que les enfants apprennent d'abord à comprendre un mot avant de le produire.

En conclusion, bien que la fréquence d'apparition des noms chez les enfants de notre échantillon varie peu, cette observation doit être relativisée. Des facteurs tels que la taille de l'échantillon, le contexte d'interaction et la manière dont la fréquence des noms est mesurée, doivent être considérés pour mieux comprendre les résultats. Une analyse plus approfondie de ces éléments pourrait enrichir la compréhension des mécanismes d'acquisition lexicale chez l'enfant.

4.5 Les limites de l'étude

L'une des principales limites de notre étude réside dans l'utilisation de la norme ValE_{mo}, qui, rappelons-le, ne comporte que 600 noms. De plus, une fois cette norme mise en corrélation avec nos corpus, seuls 419 noms ont été retenus pour l'analyse, représentant les termes communs entre la norme ValE_{mo} et ceux dans nos données. Ce processus crée un biais, car nous ne traitons pas l'intégralité des données des corpus. En effet, nous avons effectué une présélection en ne prenant en considération que les noms présents à la fois dans les corpus d'enfants et dans la norme ValE_{mo}. Par conséquent, notre traitement des données est restreint à ces noms et non à l'intégralité des noms produits par les enfants dans les suivis longitudinaux retenus.

L'ajout de noms supplémentaires à l'étude pourrait avoir un impact important sur les résultats, en permettant une meilleure représentativité des données et en offrant une vision plus complète du phénomène étudié. En intégrant tous les noms présents dans les corpus, y compris ceux qui n'ont pas été initialement pris en compte, l'analyse pourrait bénéficier d'une plus grande précision, en évitant que certains noms influents ne soient négligés. Pour des travaux futurs, les étapes nécessaires afin d'effectuer ce travail commenceraient par l'établissement d'une liste exhaustive de tous les noms présents dans les corpus, suivie d'une identification des noms manquants. Une fois ces noms supplémentaires répertoriés, nous pourrions procéder à la même annotation que celle effectuée pour les noms de la présente étude, et aux mêmes traitements statistiques. L'ajout de ce nouvel ensemble de données pourrait révéler de nouvelles dynamiques ou renforcer les conclusions initiales.

L'autre limite de l'étude concerne notre approche de l'influence adulte. Si la variable "occurrences adultes" semble influencer la production des noms, l'utilisation fréquente d'un nom par les adultes ne garantit pas qu'il sera repris par les enfants. Il est également possible que les enfants orientent eux-mêmes les échanges en influençant les thèmes de conversation, incitant ainsi indirectement les adultes à employer certains noms plutôt que d'autres. Ce phénomène souligne l'importance de considérer le dialogue comme une co-construction, dans laquelle l'enfant joue un rôle actif. De plus, nous nous sommes focalisées uniquement sur le nombre d'occurrences et sur la fréquence de productions des noms par les adultes, sans explorer en profondeur leurs actions dans le dialogue. En effet, nous n'avons pas examiné comment l'adulte se comporte dans le dialogue, comment il étaye, reprend les productions erronées de l'enfant ou

l'aide à produire des formes conventionnelles. Notre perspective reste très quantitative, laissant de côté des aspects interactionnels qui auraient pu enrichir davantage notre analyse du processus d'acquisition du nom.

CONCLUSION

Dans le cadre de ce mémoire, nous avons cherché à explorer l'acquisition du nom chez des enfants âgés de deux à cinq ans. Nous nous sommes intéressées précisément à huit variables caractérisant le nom telles que les caractéristiques sémantiques du nom (le caractère animé/inanimé et concret/abstrait du nom), sa longueur, sa fréquence, la valence émotionnelle qui y est associée, le nombre d'occurrences du nom produites par les adultes, le niveau de développement langagier des enfants et le genre des enfants. Pour ce faire, nous avons sélectionné les noms communs entre les corpus CHILDES des enfants choisis et la norme VAEMo.

Ce travail s'est appuyé sur des données extraites de la base CHILDES, une base de données rassemblant des corpus d'interactions généralement spontanées entre des enfants et des adultes (MacWhinney, 2000). Dix suivis longitudinaux francophones ont été extraits, cinq filles et cinq garçons. Ce travail consiste à identifier les facteurs prédominants qui influencent la production des noms chez les enfants.

Les résultats de notre étude ont montré que quatre variables ont une influence sur la production des noms chez l'enfant. Les quatre variables sont la variable *occurrences adultes*, la variable *niveau de développement langagier*, la variable *caractère animé/inanimé* et la variable *longueur du nom*. Ces résultats révèlent que le rôle de l'adulte est important pour expliquer la production des noms par l'enfant : il influence le nombre d'apparitions du nom chez l'enfant, car un mot fréquemment produit dans le discours adulte sera également produit par l'enfant. Le niveau de développement langagier a également montré une forte corrélation avec la diversité lexicale. En effet, les enfants, avec un niveau de développement langagier avancé, produisent des noms en plus grand nombre, mais également des noms plus longs. La variable *caractère animé/inanimé* confirme en partie les affirmations de la revue de littérature : les noms animés prédominent dès le plus jeune âge. Cependant, contrairement aux prévisions de la littérature qui suggèrent une augmentation des noms inanimés avec l'âge, notre étude montre que les noms animés restent plus nombreux que les noms inanimés, même à mesure que l'enfant grandit. La variable *longueur du nom* prouve que les noms monosyllabiques dominent aux premières étapes du développement, avec l'âge, les enfants produisent des noms plus longs.

Les autres variables, c'est-à-dire la variable *concret/abstrait* du nom, la variable *fréquence*, la variable *valence émotionnelle* et la variable *genre* de l'enfant, n'ont pas révélé d'effets significatifs. De même pour les variables aléatoires (l'enfant et le nom). Elles n'influencent pas le nombre d'apparitions des noms, suggérant que ces facteurs n'ont pas d'impact direct sur la production des noms chez les enfants.

Bien que quatre de nos variables fournissent une réponse positive à nos questions de recherche, des études ultérieures devraient être menées sur un plus large échantillon de noms afin de confirmer la significativité des variables. Comme mentionné dans les limites de l'étude, le nombre total de noms, limité à 419, restreint la portée des résultats. Il serait aussi pertinent de se tourner vers un approfondissement des productions adultes afin d'évaluer qualitativement le rôle des interactions. Cela viendrait enrichir les connaissances que nous avons du rôle de l'adulte sur la production des noms chez l'enfant.

ANNEXE A :
CORPUS

Noms communs entre la norme ValEmo et nos corpus	
A à J	K à Z
abeille	kangourou
agneau	lacet
aiguille	lait
aile	lampe
ami	langue
amour	lapin
ampoule	larme
ananas	lavabo
âne	léopard
année	lettre
antenne	lézard
araignée	limace
arbre	lion
argent	lionceau
arme	lit
arrosoir	livre
aspirateur	locomotive
assiette	louche
autruche	loup
avion	loupe
baguette	luge
baignoire	lumière
bain	lune
balai	main
balançoire	maïs
baleine	maison
balle	manteau
ballon	mare
banane	marelle
banc	mariage
barbe	marteau
barre	masque
bassine	matin

bateau	médaille
bébé	médicament
bec	méduse
biberon	mensonge
bibliothèque	mer
biche	mètre
biscuit	micro
boîte	miroir
bol	monde
bombe	montagne
bonbon	montre
bonnet	moto
bosse	mouche
botte	moulin
bouche	moustache
bouée	moustique
bougie	mouton
bouteille	mur
bouton	muscle
branche	musique
bras	natte
brique	neige
brosse	nez
brouette	niche
bruit	nid
bulle	nœud
bureau	noix
bus	note
cadeau	nuage
cafetière	nuit
cage	œil
cahier	œuf
caisse	oiseau
calendrier	ongle
calme	orange
camion	oreille
canapé	orteil
canard	os
canne	ours

capuche	pain
carnet	paix
carotte	panda
carte	panier
casque	pantalon
casquette	papier
cassette	papillon
ceinture	paquet
cercle	parapluie
cerf	pays
cerise	pêche
chaise	peinture
chameau	pelle
champignon	perroquet
chant	peur
chapeau	phoque
chasse	piano
chat	pièce
château	pied
chaussette	pigeon
chaussure	pile
cheminée	pinceau
chemise	pion
chenille	pipe
cheval	piscine
cheveux	placard
chèvre	planète
chien	plante
chocolat	plat
chou	pluie
ciel	plume
cigarette	pneu
cintre	poche
circuit	poing
ciseau	poire
citron	poisson
citrouille	pomme
cloche	pont
clou	porte

clown	pot
coccinelle	poteau
cochon	poubelle
cœur	pouce
coffre	poule
colère	poulet
collier	poupée
coq	poussière
coquillage	poussin
corde	prise
corne	puzzle
cou	quille
couleur	raisin
couronne	rat
cours	râteau
couteau	règle
crabe	reine
crayon	renard
crêpe	requin
crevette	rêve
crocodile	rire
croix	robe
cube	robinet
cuillère	robot
cuisine	rose
cygne	roue
dauphin	route
dent	sandwich
dieu	sang
dinosaure	sanglier
disque	santé
docteur	sapin
doigt	seau
domino	selle
douche	serpent
dragon	sieste
drap	silence
eau	singe
échelle	sirène

éclair	ski
école	soleil
écran	sommeil
écureuil	sort
église	soupe
éléphant	souris
enfant	souvenir
épée	squelette
erreur	stylo
escalier	table
escargot	tableau
essence	tabouret
étiquette	tambour
étoile	tampon
famille	tartine
fantôme	tasse
fatigue	taupe
fée	téléphone
fenêtre	temps
fête	tente
feu	terre
feuille	tétine
filles	thermomètre
flèche	tigre
fleur	titre
force	toilette
fouet	toit
fourchette	tomate
fourmi	tonneau
fraise	tortue
framboise	toupie
fromage	tournevis
fruit	tracteur
fusée	train
gant	transport
garage	travail
garçon	trompette
gâteau	tronc
genre	trou

gilet	tunnel
girafe	tuyau
gomme	vache
gourde	valise
goutte	vélo
grenouille	vent
grimace	vérité
guerre	verre
guitare	vie
hache	vieux
hélicoptère	vin
hérisson	violon
heure	vitesse
hibou	voiture
idée	volant
île	volcan
indien	voyage
jambe	wagon
jambon	zèbre
jardin	
jeu	
jupe	

ANNEXE B :

RÉSULTATS COMPLETS DES ANALYSES SANS INTERACTIONS

Résultats

Generalized Mixed Models

Model Info

Info	Value	Comment
Model Type	Poisson	Model for count data
Call	glm	Nb d'apparitions ~ 1 + Groupe MLU + Valence + Fréquence_ Antidote + Concrétude + Animé / inanimé + Longueur (nbr syllabes) + Occurrences adultes + Sexe + (1 Enfant) + (1 Items)
Link function	log	Coefficients are in the log(y) scale
Distribution	Poisson	Model for count data
LogLikel.	-4254.854407	Unconditional Log-Likelihood
-2*LogLikel.	8509.708814	Unconditional absolute deviance
Deviance	2487.399313	Conditional relative deviance
R-squared	0.094792	Marginal
R-squared	0.228450	Conditional
AIC	8531.710000	Less is better
BIC	8594.755226	Less is better
Residual DF	2268.000000	
Chi-squared/DF	1.400118	Overdispersion indicator
Converged	yes	
Optimizer	bobyqa	

[3]

Model Results

Fixed Effect Omnibus tests

	X ²	df	p
Groupe MLU	25.3250509	1.0000	<.001
Valence	0.8638316	1.0000	0.353
Fréquence_ Antidote	2.7162853	1.0000	0.099
Concrétude	1.7187142	1.0000	0.190
Animé / inanimé	4.2068864	1.0000	0.040
Longueur (nbr syllabes)	2.0789744	1.0000	0.149
Occurrences adultes	349.3441524	1.0000	<.001
Sexe	0.0039592	1.0000	0.950

Fixed Effects Parameter Estimates

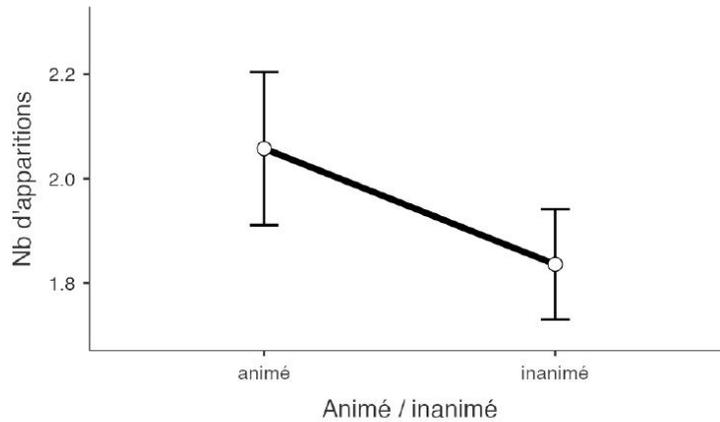
Names	Effect	Estimate	SE	exp(B)	95% Exp(B) Confidence Interval		z	p
					Lower	Upper		
(Intercept)	(Intercept)	0.3204880	0.0893277	1.37780	1.15651	1.64143	3.587780	<.001
Groupe MLU	Groupe MLU	0.0818776	0.0162701	1.08532	1.05126	1.12049	5.032400	<.001
Valence1	positif - negatif	0.0740398	0.0796619	1.07685	0.92118	1.25882	0.929425	0.353
Fréquence_ Antidote	Fréquence_ Antidote	0.0060976	0.0036997	1.00612	0.99885	1.01344	1.648116	0.099
Concrétude1	concret - abstrait	0.1096447	0.0836346	1.11588	0.94717	1.31464	1.310997	0.190
Animé / inanimé1	inanimé - animé	-0.1141697	0.0556635	0.89211	0.79990	0.99494	-2.051070	0.040
Longueur (nbr syllabes)	Longueur (nbr syllabes)	0.0545265	0.0378167	1.05604	0.98060	1.13729	1.441865	0.149
Occurrences adultes	Occurrences adultes	0.0545992	0.0029212	1.05612	1.05009	1.06218	18.690750	<.001
Sexe1	garçon - fille	-0.0042649	0.0677803	0.99574	0.87187	1.13721	-0.062922	0.950

Random Components

Groups	Name	SD	Variance	ICC
Items	(Intercept)	0.250673	0.0628370	0.131756
Enfant	(Intercept)	0.094317	0.0088956	0.021031
Residuals		1.000000	1.0000000	.

Note. Number of Obs: 2279 , groups: Items 281, Enfant 10

Effects Plots



Références

[1] The jamovi project (2023). *jamovi*. (Version 2.4) [Computer Software]. Retrieved from <https://www.jamovi.org>.

[2] R Core Team (2022). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2023-04-07).

[3] Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from <https://gamlj.github.io/>.

ANNEXE C :

RÉSULTATS COMPLETS DES ANALYSES AVEC INTERACTIONS

Résultats

Generalized Mixed Models

Model Info

Info	Value	Comment
Model Type	Poisson	Model for count data
Call	glm	Nb d'apparitions ~ 1 + Groupe MLU + Valence + Fréquence_Antidote + Concrétude + Animé / inanimé + Longueur (nbr syllabes) + Occurrences adultes + Sexe + Groupe MLU:Animé / inanimé + Groupe MLU:Concrétude + Groupe MLU:Valence + Groupe MLU:Fréquence_Antidote + Groupe MLU:Longueur (nbr syllabes) + (1 Enfant) + (1 Items)
Link function	log	Coefficients are in the log(y) scale
Distribution	Poisson	Model for count data
LogLikel.	-4246.945695	Unconditional Log-Likelihood
-2*LogLikel.	8493.891391	Unconditional absolute deviance
Deviance	2466.851458	Conditional relative deviance
R-squared	0.099265	Marginal
R-squared	0.236046	Conditional
AIC	8525.890000	Less is better
BIC	8617.595263	Less is better
Residual DF	2263.000000	
Chi-squared/DF	1.383501	Overdispersion indicator
Converged	yes	
Optimizer	bobyqa, Nelder_Mead, nloptwrap	

Note. The model did not converge. Results may be misleading or uninterpretable.

Note. Model failed to converge with max|grad| = 0.0368386 (tol = 0.002, component 1)

Note. Model is nearly unidentifiable: very large eigenvalue - Rescale variables?

[3]

Model Results

Fixed Effect Omnibus tests

	X ²	df	p
Groupe MLU	18.8325922	1.0000	<.001
Valence	1.0494838	1.0000	0.306
Fréquence_Antidote	0.5986918	1.0000	0.439
Concrétude	1.3344005	1.0000	0.248
Animé / inanimé	5.2735774	1.0000	0.022
Longueur (nbr syllabes)	6.0171968	1.0000	0.014
Occurrences adultes	351.4954108	1.0000	<.001
Sexe	0.0059905	1.0000	0.938
Groupe MLU * Animé / inanimé	11.4250974	1.0000	<.001
Groupe MLU * Concrétude	0.6143618	1.0000	0.433
Groupe MLU * Valence	0.4749077	1.0000	0.491
Groupe MLU * Fréquence_Antidote	0.0427098	1.0000	0.836
Groupe MLU * Longueur (nbr syllabes)	4.2745011	1.0000	0.039

Fixed Effects Parameter Estimates

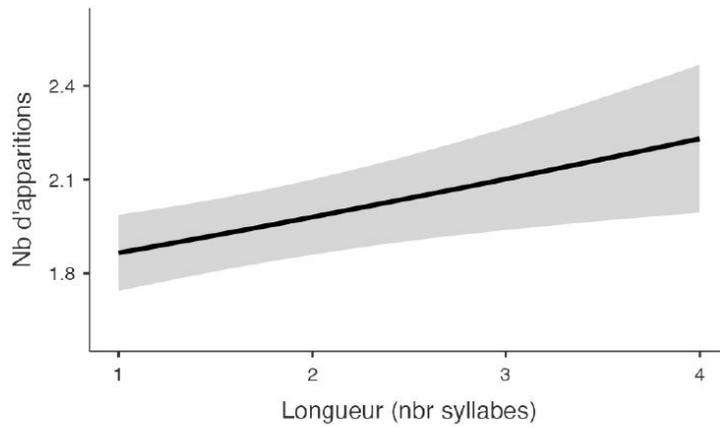
Names	Effect	Estimate	SE	exp(B)	95% Exp(B) Confidence Interval		z	p
					Lower	Upper		
(Intercept)	(Intercept)	-0.3120995	0.2248012	0.73191	0.47109	1.13712	-1.388336	0.165
Groupe MLU	Groupe MLU	0.2084134	0.0480253	1.23172	1.12107	1.35329	4.339653	<.001
Valence1	positif - negatif	0.2366123	0.2309667	1.26695	0.80568	1.99232	1.024443	0.306
Fréquence_ Antidote	Fréquence_ Antidote	0.0078539	0.0101504	1.00788	0.98803	1.02814	0.773752	0.439
Concrétude1	concret - abstrait	0.2711619	0.2347392	1.31149	0.82785	2.07766	1.155163	0.248
Animé / inanimé1	inanimé - animé	0.3249631	0.1415081	1.38398	1.04876	1.82634	2.296427	0.022
Longueur (nbr syllabes)	Longueur (nbr syllabes)	0.2679805	0.1092461	1.30732	1.05534	1.61947	2.452998	0.014
Occurrences adultes	Occurrences adultes	0.0550537	0.0029365	1.05660	1.05053	1.06270	18.748211	<.001
Sexe1	garçon - fille	-0.0054194	0.0700189	0.99460	0.86705	1.14090	-0.077398	0.938
Groupe MLU * Animé / inanimé1	Groupe MLU * inanimé - animé	-0.1036677	0.0306700	0.90152	0.84893	0.95738	-3.380103	<.001
Groupe MLU * Concrétude1	Groupe MLU * concret - abstrait	-0.0387032	0.0493782	0.96204	0.87329	1.05980	-0.783812	0.433
Groupe MLU * Valence1	Groupe MLU * positif - negatif	-0.0346669	0.0503050	0.96593	0.87524	1.06602	-0.689135	0.491
Groupe MLU * Fréquence_ Antidote	Groupe MLU * Fréquence_ Antidote	-4.6199e-4	0.0022355	0.99954	0.99517	1.00393	-0.206663	0.836
Groupe MLU * Longueur (nbr syllabes)	Groupe MLU * Longueur (nbr syllabes)	-0.0495652	0.0239737	0.95164	0.90796	0.99743	-2.067487	0.039

Random Components

Groups	Name	SD	Variance	ICC
Items	(Intercept)	0.253937	0.0644839	0.134744
Enfant	(Intercept)	0.098259	0.0096548	0.022785
Residuals		1.000000	1.0000000	.

Note. Number of Obs: 2279 , groups: Items 281, Enfant 10

Effects Plots



Références

- [1] The jamovi project (2023). *jamovi*. (Version 2.4) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2022). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2023-04-07).
- [3] Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from <https://gamlj.github.io/>.

RÉFÉRENCES

- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of child language*, 42(2), 239-273.
- Bassano, D. & Maillolchon, I. (1994). Early grammatical and prosodic marking of utterance modality in French: a longitudinal case study. *Journal of Child Language*, 21, 649-675.
- Bassano, D. (1998). L'élaboration du lexique précoce chez l'enfant français : structure et variabilité. *Enfance*, 51(4), 123-153.
- Bassano, D. (1998). Sémantique et syntaxe dans l'acquisition des classes de mots : l'exemple des noms et des verbes en français. *Langue française*, 26-48.
- Bassano, D., Maillolchon, I., & Eme, E. (1998). Developmental changes and variability in early lexicon: A study of French children's naturalistic productions, *Journal of Child Language*, 25, 493-531.
- Bassano, D. (2000). Early development of nouns and verbs in French: Exploring the interface between the lexicon and grammar. *Journal of Child Language*, 27, 521-559.
- Bassano, D. (2000). La constitution du lexique : le développement lexical précoce. *L'acquisition du langage*, 1, 137-168.
- Bassano, D., Laaha, S., Maillolchon, I., & Dressler, W.U. (2004). Early acquisition of verb grammar and lexical development: Evidence from periphrastic constructions in French and Austrian German. *First Language*, 24(1), 33-70.
- Bassano, D. (2005). Production naturelle précoce et acquisition du langage. L'exemple du développement des noms. *Lidil. Revue de linguistique et de didactique des langues*, (31), 61-84.
- Bates, E. (1976). *Language and Context: The Acquisition of Pragmatics*. Academic Press.
- Bates, E., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. Academic Press.
- Bates, E., Bretherton, I. et Snyder, L. (1988), *From first words to grammar: individual differences and dissociable mechanisms*, New York, Cambridge University Press.
- Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P. S., Reznick, J. S., Reilly, J. et Hartung, J. (1994), Developmental and stylistic variation in the composition of early vocabulary, *Journal of Child Language*, 21 (1), 85-123.
- Bates, E., Bretherton, I., & Snyder, L. (1994). *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge University Press.

- Bates, E., Dale, P. et Thal, D. (1995), Individual differences and their implications for theories of language development, in P. Fletcher et B. MacWhinney (Eds), *The handbook of child language* (p. 96-151), Oxford, UK, Basil Blackwell.
- Batista, A., & Colletta, J.M. (2011). Etude longitudinale des productions multimodales d'enfants français âgés de 18 mois à 3 ans et demi (41 mois). *LIDIL-Revue de linguistique et de didactique des langues*.
- Benveniste, E. (1966). *Problèmes de linguistique générale, Tome 1*. Gallimard.
- Bergelson, E., & Swingley, D. (2012). *The influence of input on early word learning*. *Journal of Child Language*, 39(1), 134-157
- Bloom L. (1973). *One word at a time*. Mouton : Janua Linguarum, The Hague.
- Bloom, L. (2000). How children learn words. *Cambridge University Press*.
- Blitman. (2015). *Le langage est-il inné ?* Presses universitaires de Franche-Comté.
- Bonin, P., Méot, A., Aubert, L. F., Malardier, N., Niedenthal, P., & Capelle-Toczek, M. C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L'année Psychologique*, 103(4), 655-694.
- Borghi, A. M., & Pecher, D. (2011). The role of sensorimotor experiences in abstract concepts: A critical review. In *The Cambridge Handbook of Cognitive Science* (pp. 299-311). Cambridge University Press.
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., Pêcheux, M-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child development*, 75(4), 1115-1139.
- Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M. C., & Deneault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied psycholinguistics*, 30(4), 685-707.
- Brown, R. (1973), *A first language: the early stages*, Cambridge, MA, Harvard University Press.
- Bruner, J. S. (1983). *Le développement de l'enfant : savoir faire, savoir dire*. Presses Universitaires de France
- Brysbaert, M., & New, B. (2009). *Moving beyond Kucera and Francis: A critical review of current word frequency norms and the introduction of a new and improved word frequency measure for American English*. *Behavior Research Methods*, 41(4), 977-990.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, 10(5), 425-455.
- Bybee, J. (2006). *From usage to grammar: The mind's response to repetition*. *Language*, 82(4), 711-733.

- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction-based analysis of child directed speech. *Cognitive science*, 27(6), 843-873.
- Caselli, M. C., Bates, E., Casadio, P., Fenson, L., Fenson, J., Sanderl, L. et Weir, J. (1995), A crosslinguistic study of early lexical development, *Cognitive Development*, 10, 159-199.
- Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of child language*, 22(3), 497-529.
- Chomsky, N. (1959/1969). Un compte-rendu du « comportement verbal » de B.F. Skinner. *Langages*.
- Chomsky, N. (1965/1971). *Aspects de la théorie syntaxique*. Paris : Éditions du Seuil.
- Chomsky, N. (1980). Rules and representations. *Behavioral and brain sciences*, 3(1), 1-15.
- Chomsky, N. (1991). *Théorie du gouvernement et du liage : les conférences de Pise*. Paris : Ed. du Seuil.
- Chomsky, N. (1996). *World orders, old and new*. Columbia University Press.
- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge : Cambridge University Press.
- Clark, E. V. (2006). La répétition et l'acquisition du langage. *La linguistique*, 42(2), 67-80.
- Clark, E. V. (2006). *Language acquisition and language change: Cognitive aspects of language development*. Cambridge University Press.
- Corbin, D. (2004). *Le lexique de l'animé et de l'inanimé en français : enjeux et représentations*. Presses Universitaires de Rennes.
- De Cat, C. & B. Plunkett (2002). QU'est ce qu'i (l) dit, celui+Là?: notes méthodologiques sur la transcription d'un corpus francophone. *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache / Romance Corpus Linguistics : Corpora and Spoken Language*. Tübingen: Narr. CD— rom.
- Demuth, K. & A. Tremblay (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 35, 99-127
- De Pontonx, S., Leroy-Collombel, M., & Morgenstern, A. (2019). How mother and child co-(re) construct non-conventional productions in spontaneous interaction. *First language*, 39(2), 220-242.
- De Villiers, J. (1985). Learning how to use verbs: Lexical coding and the influence of input. *Journal of ChildLanguage*, 12, 587-596
- Dodane, C., Martel, K., Yamaguchi, N., & Morgenstern, A. (2017). Émergence des premiers mots. *Le Langage de l'Enfant. De l'Éclosion à l'Explosion*, 41-58
- Dromi, E. (1996). *Early lexical development*. Oxford University Press.

- Druide informatique. (2016). *Antidote 9 - Manuel d'utilisation*. Druide Informatique
- Duchêne, A. (2013). *Les relations entre le lexique et la situation d'énonciation : une étude sur les noms animés et inanimés*. *Journal of French Linguistics*, 24(2), 45–63.
- Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental psychology*, 23(1), 132.
- Fenson, L., Dale, P. S., Reznick, J., Bates, E., Thal, D. & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, 59(5), 1-173.
- Ferrand, L. (2001). Normes d'associations verbales pour 260 mots « abstraits ». *L'Année psychologique*, 101(4), 683-721.
- Florin, A. (2010). Le développement du lexique et l'aide aux apprentissages. *Enfances & Psy*, (2), 30-41.
- Foulley, J. L., & Im, S. (1992). A marginal quasi-likelihood approach to the analysis of Poisson variables with generalized linear mixed models. *Genetics Selection Evolution*, 25(1), 101-107.
- Garrigues, J. (2010). *Syntaxe et sémantique des noms animés et inanimés en français*. *Le Français Moderne*, 4(4), 91–110.
- Gayraud, F. & Kern, S. (2007). Caractéristiques phonologiques des noms en fonction de l'âge d'acquisition. *Enfance*, 59, 324-338.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *BBN report; no. 4854*.
- Goldberg, A. E. (2007). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Golinkoff, R. M., & Hirsh-Pasek, K. (2006). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, 10(10), 397-403.
- Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of child language*, 35(3), 515-531.
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child development*, 73(2), 418-433.
- Hoff, E. (2006). *How social contexts support and shape language development*. *Developmental Review*, 26(1), 55-88.

- Jouitteau, M. (2011). *Le rôle de la distinction animé/inanimé en syntaxe et en sémantique*. *Revue de Linguistique Française*, 35(2), 120-135.
- Joly, C. (2009). *La classification animée/inanimée : perspectives linguistiques et psycholinguistiques*. *Revue de Linguistique*, 38(3), 123–136.
- Kern, S. (2007). Lexicon development in French-speaking infants. *First Language*, 27(3), 227-250.
- Kern, S., & Gayraud, F. (2010). *Inventaire Français du Développement Communicatif*. Grenoble : Éditions La Cigale
- Kern, S. (2019). Des premiers mots à l'émergence de la grammaire. *Le développement du langage chez le jeune enfant. Théorie, clinique, pratique*, 85-111.
- Klee, T., & Fitzgerald, M. D. (1985). The relation of length of utterance to grammatical complexity in normal and language-disordered groups. *Journal of Speech and Hearing Research*, 28(2), 259-264.
- Leaper, C. (2002). *Parent-child interactions and gender differences in the development of language*. *Developmental Psychology*, 38(5), 520-531.
- Lecerf, P. (2011). *L'acquisition du langage chez l'enfant*. Editions de la Sorbonne.
- Le Normand, M. T., Moreno-Torres, I., Parisse, C., & Dellatolas, G. (2013). How do children acquire early grammar and build multiword utterances? A corpus study of French children aged 2 to 4. *Child Development*, 84(2), 647-661.
- Leroy, M., Mathiot, E., & Morgenstern, A. (2009). Pointing gestures and demonstrative words: Deixis between the ages of one and three. *Studies in Language and Cognition Cambridge Scholars Publishing*. Editors: Jordan Zlatev, Marlene Johansson Falck, Carita Lundmark and Mats André. 386-404.
- Lieven, E. (2006). *Usage-based approaches to language acquisition*. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (5th ed., pp. 329-358).
- MacWhinney, B. (2000). The CHILDES Project. *Computational Linguistics*, 26(4), 657-657.
- Markman, E. M. (1990). *Constraints children place on word meanings*. *Cognitive Science*, 14(1), 57-77.
- McNeill, D. (1970). *The acquisition of language*. New-York : Harper.
- Mitchell, C., & McMurray, B. (2009). On leveraged learning in lexical acquisition and its relationship to acceleration. *Cognitive Science*, 33(8), 1503-1523.
- Morgenstern A. (2006) Un JE en construction. Ontogenèse de l'auto-désignation chez l'enfant. Bibliothèque de Faits de langues. Ophrys

- Morgenstern, A., & Parisse, C. (2007). Codage et interprétation du langage spontané d'enfants de 1 à 3 ans. *Corpus*, (6), 55-78.
- Morgenstern, A., Sekali, M. (2009). What can child language tell us about prepositions? A contrastive corpus-based study of cognitive and social-pragmatic factors. *Studies in Language and Cognition*, Cambridge Scholars Publishing. Editors: Jordan Zlatev, Marlene Johansson Falck, Carita Lundmark and Mats André. 261-275
- Morgenstern A. with the collaboration of Benazzo, S., Leroy, M., Mathiot, E., Parisse. C., & Sekali, M. (2009). *L'enfant dans la langue. De l'observation du naturaliste à l'analyse du linguiste.* (Presses de la Sorbonne Nouvelle)
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1-135.
- Naigles, L., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before others? *Journal of Child Language*, 25, 95–120
- Nelson, K. (1975). The nominal shift in semantic-syntactic development. *Cognitive psychology*, 7(4), 461-479.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : Une nouvelle base de données lexicales françaises. *Méthodes, instruments et ordinateurs de recherche sur le comportement*, 36 (3), 516-524.
- Ogren, M., & Sandhofer, C. M. (2021). Emotion words in early childhood: A language transcript analysis. *Cognitive development*, 60, 101122.
- Osgood C. E., Suci G. J. — (1955) Factor analysis of meaning, *Journal of Experimental Psychology*, 50, 325-338.
- Parisse, C., & Le Normand, M. T. (2000). How children build their morphosyntax: The case of French. *Journal of Child Language*, 27(2), 267-292.
- Parisse, C., & Le Normand, M. T. (2006). Une méthode pour évaluer la production du langage spontané chez l'enfant de 2 à 4 ans. *Glossa*, (97), 20-41.
- Parisse, C., & Le Normand, M. T. (2007). Une méthode pour évaluer la production du langage spontané chez l'enfant de 2 à 4 ans. *Glossa*, (97), 10-30.
- Pine, J. M., Lieven, E. V. M., & Rowland, C. F. (1999). *The role of lexical frequency in the acquisition of syntax.* *Journal of Child Language*, 26(1), 73-92.
- Plunkett, B. (2002). Null Subjects in child French interrogatives: A view from the York Corpus. *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache / Romance corpus linguistics : Corpora and spoken language*, 441-452.

- Pratto F., John O. P. — (1991) Automatic vigilance : The attention-grabbing power of negative information, *Journal of Personality and Social Psychology*, 61, 381-391.
- Reznick, J. S., & Goldfield, B. A. (1992). Rapid change in lexical development in comprehension and production. *Developmental psychology*, 28(3), 406.
- Riegel, M., Pellat, J.-C., & Rioul, R. (1994). *Grammaire méthodique du français*. Presses Universitaires de France.
- Rivest, A. (2013). La régression de Poisson multiniveau généralisée au sein d'un devis longitudinal : un exemple de modélisation du nombre d'arrestations de membres de gangs de rue à Montréal entre 2005 et 2007.
- Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, 83(5), 1762-1774.
- Ségéral, P. (2008). *La lexicologie et la lexicographie du français contemporain*. Presses Universitaires de France.
- Snow, C. E. (1995). Issues in the study of input: Fine-tuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 180-193).
- Snow, C. E. (1999). Social perspectives on the emergence of language. *The emergence of language*, 257-276.
- Snow, C. E. (2010). *Academic language and the challenge of reading for understanding*. *American Educator*, 34(2), 8-17.
- Stern, D., Spina, D., & Gelfand, M. (1985). *The emotional development of the infant*. Cambridge University Press.
- Syssau, A., & Font, N. (2005). Évaluations des caractéristiques émotionnelles d'un corpus de 604 mots. *Bulletin de psychologie*, (3), 361-367.
- Syssau, A., & Monnier, C. (2009). Children's emotional norms for 600 French words. *Behavior Research Methods*, 41(1), 213-219.
- Tardif, T. (1996). Nouns are learned before verbs: Evidence from Mandarin-speaking children. *Developmental Psychology*, 32(3), 449-454.
- Theakston, A., Lieven, E., Pine, J., & Rowland, C. (2001). The role of performance limitations in the acquisition of verb-argument structure. *Journal of Child Language*, 28, 127-152
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, 4(12), 197-211.

- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child development*, 1454-1463.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tomasello, M. (2009). The usage-based theory of language acquisition. In *The Cambridge handbook of child language* (pp. 69-87). Cambridge Univ. Press.
- Tomasello, M. (2009). *The cultural origins of human cognition*. Harvard university press.
- Tournier, P. (1998). *La personnification des objets inanimés dans le discours figuré*. *Linguistique et Cognition*, 12(3), 45-58.
- Vellard, J. (2007). *Sémantique et syntaxe des noms animés et inanimés*. *Langue et Grammaire*, 25(1), 89-102.
- Veneziano, E., Sinclair, H., & Berthoud, I. (1990). From one word to two words: repetition patterns on the way to structured speech. *Journal of child language*, 17(3), 633-650.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Winter, B., & Bürkner, P. C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and Linguistics Compass*, 15(11), e12439.
- Yamaguchi, Naomi (2012). *Parcours d'acquisition des sons du langage chez deux enfants francophones*, Ph.D. Dissertation, Université Sorbonne Nouvelle Paris 3.