

Claudine HAMMELRATH  
Orthophoniste  
42 rue d'Amiens  
62000 ARRAS

Institut d'orthophonie  
Gabriel Deceox  
Pôle formation  
59045 LILLE CEDEX

**RÉSUMÉ :**

*Cet article présente les résultats d'une analyse informatisée de corpus de langage spontané de 258 enfants. Les premiers enregistrements ont été recueillis à l'occasion d'un test de dépistage en moyenne section de maternelle. Un deuxième corpus a été enregistré auprès des mêmes enfants pendant les mois de mai-juin de la grande section de maternelle en utilisant le même support imagé. Certains enfants avaient bénéficié de séances de stimulation du langage à l'école. Le CLAN\* de Computerized Language Analysis : Brian Mac Whinney et coll\* a permis l'analyse quantitative des données : longueur moyenne d'énoncé, indice de diversité lexicale, nombre total de mots et de mots différents. Il autorise aussi une analyse qualitative : il est aisé de dresser le paradigme de chaque catégorie lexicale. Il ne peut être présenté ici les résultats qualitatifs de tous les corpus mais je propose l'analyse des catégories morphosyntaxiques des corpus de six enfants. L'étude d'un grand nombre de fichiers témoigne de la variabilité des résultats, souligne aussi des faits de langue comme par exemple le redoublement de sujet 63 % des enfants entre 3 ans et 5 ans et 84 % des enfants entre 5 ans 4 mois et 6 ans 4 mois.*

**MOTS-CLÉS :**

Linguistique - Informatique - Morphosyntaxe - Enfants.

# ANALYSE INFORMATISÉE DE CORPUS DE LANGAGE SPONTANÉ DE 258 ENFANTS DE MATERNELLE

par Claudine HAMMELRATH

## SUMMARY : *Analyse computerized corpus of spontaneous language of 258 children of nursery school*

*This article has the results of a computerized analysis of corpus of spontaneous language of 258 children. The first recordings were collected at the time of a test of tracking on average section of nursery school. A second corpus was recorded near the same children during May June of the great section of nursery school by using the same picturesque support. Some children had profited from meetings of stimulation of the language at the school. The CLAN\* : Computerized Language Analysis of Brian Mac Whinney and al\* the quantitative analysis of the data allowed: average length of statement, index of lexical diversity, numbers total different words and words. It authorizes also a qualitative analysis : it is easy to draw up the paradigm of each lexical category. It cannot be had here the qualitative results of all the corpora but we propose the analysis of the morphosyntax categories of the corpora of six children. The study of a great number of files testifies to the variability of the results, underlines also facts of language such as for example the redoubling of subject 63 % of the children between 3 years and 5 years and 84 % of the children between 5 years 4 and 6 years 4.*

**KEY-WORDS :**

Linguistic - Data processing - Morphosyntaxe - Children.

## INTRODUCTION

Une recherche action menée à la faculté d'orthophonie de Lille\* en 2001/2002/2003 a permis de recueillir des corpus de langage de 258 enfants en moyenne section, puis en grande section de maternelle. Ceux-ci ont été enregistrés et analysés par le CLAN (Computerized Language Analysis) système d'analyse de corpus de langage informatisé élaboré par Brian Mac Whinney et coll\*.

Nous pouvons ainsi proposer une étude quantitative et qualitative d'un grand nombre de données concernant la tranche d'âge 3 ans 6 mois à 6 ans 9 mois.

Le Normand\* fournit des données qualitatives et quantitatives concernant des enfants de 2 à 4 ans. Les enfants étaient enregistrés alors qu'ils étaient placés dans une situation de jeu libre pendant 20 minutes.

Le test PER 2000 de Ferrand\* nous a permis de recueillir un échantillon de langage spontané dans une situation standardisée puisque les 258 enfants ont été soumis au même matériel dans la même situation de parole. Après le récit libre, les enfants répondaient à un questionnaire. Ces réponses ont aussi été enregistrées et analysées par le CLAN.

\* Crunelle, 2003

\* MacWhinney, 1991

\* Le Normand, 1996

\* Ferrand, 1982

## MÉTHODOLOGIE

4 écoles du Nord et 4 du Pas de Calais classées en zone sensible ont été choisies par leurs deux inspecteurs d'académie et les inspecteurs de l'éducation nationale dans le cadre d'un projet régional de santé. « Elles ne sont pas concernées directement par d'autres procédures de repérage, dépistage ou suivi d'enfants en difficulté de langage »\*. 258 enfants de moyenne section de maternelle scolarisés dans ces écoles ont été évalués par des orthophonistes à partir d'un test de dépistage le PER 2000\*.

La population se compose de 125 filles et de 133 garçons de 3 ans 6 mois à 4 ans 10 mois.

116 des 258 enfants présentaient un retard discret ou avéré de l'acquisition du langage révélé par l'application du test PER 2000 de Ferrand. Certains enfants avaient aussi été signalés par les enseignants parce qu'ils suspectaient des difficultés de communication.

Ils ont tous bénéficié de séances de stimulation à l'école, une par quinzaine. Celles-ci ont été poursuivies en grande section de maternelle. Les enfants sont installés en petits groupes auprès d'adultes enseignants, orthophonistes, membres de réseaux d'aide. Ils sont invités à s'exprimer à propos d'albums. Les histoires sont lues page par page et chacun est questionné, stimulé, encouragé à s'exprimer. Tout sera exploité : vocabulaire, phonologie, morphosyntaxe, mémoire, capacité de récit ... Les mêmes séances, avec les mêmes groupes d'enfants et les mêmes albums sont répétées par les membres des réseaux d'aide pendant la semaine intermédiaire. En outre, les enfants « riches de savoir » transmettent leur connaissance au groupe classe.

En mai/juin 2003, le même test PER 2000 a été appliqué. 51 sujets ont disparu, ils ont changé d'école ou déménagé. 8 filles et 13 garçons nouveaux arrivants dans les écoles sont intégrés à la population. Celle-ci se compose alors de 107 filles et 121 garçons de 5 ans 3 mois à 6 ans 9 mois.

Parmi les différentes épreuves du PER 2000, les enfants, isolés avec une orthophoniste, devaient s'exprimer spontanément à propos de 3 images, *les cerises*, puis répondre à un questionnaire. Les productions étaient enregistrées puis retranscrites en écriture phonétique dès qu'elles étaient altérées sur le plan articulatoire. Tous les corpus obtenus par l'application des deux tests ont été transcrits selon les conventions du CHILDES\* et analysés par le programme CLAN.

\* Crunelle, 2003

\* Ferrand, 1982

\* Mac Whinney

## RÉSULTATS

### RÉSULTATS QUANTITATIFS AU PREMIER TEST : ENFANTS DE 3 ANS 6 MOIS À 4 ANS 10 MOIS GROUPE A

La longueur moyenne d'énoncé, rapport entre nombre d'énoncés et nombre de mots, est de 4,29 et l'écart type de 1,65.

Le nombre de mots varie de 0 à 431. La valeur moyenne est de 126 avec un écart type de 74.

Le nombre de mots différents varie de 0 à 204. La valeur moyenne est de 61 avec un écart-type de 29.

Lors du premier test, 8 enfants très inhibés ne décrivent pas spontanément les images, ne répondent pas au questionnaire. Le calcul de l'indice de diversité lexicale, rapport du nombre de mots différents au nombre de mots produits indépendamment de la taille de l'échantillon, a donc été impossible pour ces 8 enfants.

La valeur moyenne de l'indice de diversité lexicale est de 39 avec un écart-type de 18.

La population a ensuite été scindée en 3 groupes d'âge afin de visualiser les évolutions des scores.

		3 ans 6 - 4 ans (44 enfants)	4 ans - 4 ans 6 (120 enfants)	4 ans 6 - 5 ans (86 enfants)
Longueur moyenne énoncé	moyenne	4,12	4,22	4,60
	écart-type	1,43	1,85	1,47
Nombre de mots	moyenne	122 valeur 10 à 228	135 valeur de 4 à 395	132 valeur de 81 à 535
	écart-type	58	79	66
Nombre de mots différents	moyenne	60 valeur de 1 à 101	63 valeur de 4 à 126	63 valeur de 6 à 139
	écart-type	22	26	28
indice de diversité lexicale	moyenne	39,5	38,26	39,87
	écart-type	16,43	20,38	17,64

\*1981

La longueur moyenne d'énoncés varie de 4,12 à 4,60. Selon Miller et Chapman\*, la longueur moyenne de production verbale d'enfants entre 3 ans 9 mois et 4 ans 8 mois varie entre 4,09 et 5,32.

Le nombre de mots et de mots différents augmente proportionnellement à l'âge des enfants mais il existe une grande variabilité du nombre de mots pour chaque tranche d'âge.

### RÉSULTATS QUALITATIFS AU PREMIER TEST :

Etant donné l'importance du matériel linguistique, il est impossible de présenter une analyse exhaustive des structures morphosyntaxiques. J'ai choisi d'exposer trois corpus d'enfants de 4 ans 1 mois à 4 ans 9 mois dont les scores longueur moyenne d'énoncé (LME), indice de diversité lexicale, nombre de mots et nombre de mots différents se situent dans la moyenne. Leurs corpus seraient donc représentatifs des corpus d'autres enfants du même âge. Si l'on étudie la **première partie des corpus** c'est à dire l'expression libre sur les images, avant le questionnement de l'adulte, on remarque l'hétérogénéité des modes d'expression.

Alexandre dénomme sans générer aucune phrase ce qui favorise le score élevé de son indice de diversité lexicale.

Akim utilise une structure syntaxique redondante en utilisant le présentatif « y a ».

Alicia qui est la plus jeune des 3 enfants produit des énoncés phrases déjà plus élaborés : sujet + verbe + complément d'objet direct.

Alexandre 4 ans 7 :

- \*CHI : un chat.
- \*CHI: une fille.
- \*CHI: un papa.
- \*CHI: un garçon.
- \*CHI: pis après un arbre.
- \*CHI: d'échelle.
- \*CHI: des feuilles.
- \*CHI: des branches.
- \*CHI: un chat.
- \*CHI: des paniers.
- \*CHI: une voiture.

\*ADU: que vois tu encore sur cette image?

- \*CHI: un bonhomme.
- \*CHI: d' échelle.
- \*CHI: un arbre.
- \*ADU: qu' est ce qu' ils font?
- \*CHI: ils ramassent.
- \*ADU: combien sont ils?
- \*CHI: un deux trois quatre cinq six sept.
- \*ADU: où sont ils?
- %act: pas de réponse.
- \*ADU: quand est ce que ça se passe?
- \*CHI: Opro:subj y a une maison d' abord.
- \*ADU: ça se passe le jour?
- \*CHI: on est dehors.
- \*ADU: pourquoi sont ils là?
- \*CHI: parce que.
- \*ADU: qu' est ce qui se passe sur ces images?
- \*CHI: Opro:subj y a même un oiseau.

Akim 4 ans 9

- \*CHI : Opro:subj y a une voiture.
- \*CHI: un chat.
- \*CHI: un panier.
- \*CHI: Opro:subj y a les oiseaux.
- \*CHI: Opro:subj y a les arbres.
- \*CHI: Opro:subj y a le soleil.
- \*CHI: Opro:subj y a le ciel.
- \*CHI: Opro:subj y a les nuages.
- \*CHI: Opro:subj y a des maisons.
- \*CHI: Opro:subj y a des feuilles.
- \*CHI: Opro:subj y a un chat.
- \*CHI: Opro:subj y a un panier.
- \*CHI: Opro:subj y a des oiseaux.
- \*CHI: Opro:subj y a un chat.
- \*CHI: Opro:subj y a des céréales.

\*ADU: que vois tu encore sur cette image?

- \*CHI: et pis un chat.
- \*CHI: une voiture.
- \*CHI: et pis des oiseaux.
- \*CHI: un arbre.
- \*ADU: qu' est ce qu' ils font?
- \*CHI: il s' en va.
- \*ADU: combien sont ils?
- \*CHI: par deux.
- \*ADU: où sont ils?

Alicia 4 ans 1 :

- \*CHI: il s' en va.
- \*CHI: et pis le monsieur il enlève les fraises [\*].
- %err : fraises=cerises
- \*CHI: et pis le monsieur i [: il] pot [: porte] le panier.
- \*CHI: il prend des feuilles.
- \*CHI: et pis la madame elle touche l' ab [: arbre].
- %pho : ab
- \*CHI: ils mangent.
- \*CHI: le (n) enfant il lève son doigt.
- \*CHI: i [: il] regarde le monsieur.
- \*CHI: i [: il] va Opro:refl en aller.

\*ADU : que vois tu encore sur ces images ?

- \*ADU: qui vois tu ?
- %act : pas de réponse
- \*ADU: il n'y a pas quelqu'un sur cette image ?
- %act : pas de réponse
- \*ADU: où sont ils?

\*ADU: que font ils sur cette image?

- \*CHI: ils montent sur le [\*] branche.
- %err: le=la
- \*CHI: et Opro:subj y a même un chat monté sur une branche.
- \*ADU: comment sont ils montés dans l'arbre?
- \*CHI: ils ont monté dans [\*] l' échelle.
- %err: dans=sur
- \*CHI: ils ont sauté pis après i [: ils] ont monté ici.
- \*ADU: ils ne cueillent pas quelque chose?
- \*CHI: ils cassent une branche ici.
- \*ADU: qu' est ce qu' ils cueillent?
- \*CHI: des feuilles.
- \*ADU: pourquoi ils cueillent des cerises?
- \*CHI: parce que.
- \*ADU: qu' est ce qui se passe sur ces images?
- \*CHI: Opro:subj y a un oeuf cassé.
- \*ADU: Que fait la maman?
- \*CHI: elle fait.
- \*ADU: où sont ils?
- \*CHI: Opro:subj avait un papa dehors.
- \*ADU: comment fait elle le gâteau maman?
- \*CHI: elle rou [: roule].
- \*ADU: où est le chat?
- \*CHI: en dessous de la table.
- \*CHI: moi aussi j' [: je] sais faire un noend.

\*CHI: dehors.

- \*ADU: quand est ce que ça se passe?
- \*ADU: ça se passe le jour?
- \*CHI: oui.
- \*ADU: pourquoi sont ils là?
- \*CHI: parce que.
- \*ADU: qu' est ce qui se passe sur ces images?
- \*CHI: Opro:subj y a un chat.
- \*CHI: et pis Opro:subj y a des oiseaux.
- \*CHI: et pis Opro:subj y a une branche d' arbre.
- \*ADU: que font ils sur cette image?
- \*CHI: ils regardent sur les arbres.
- \*ADU: comment sont ils montés dans l'arbre?
- \*CHI: ils ont sauté pis ils ont attrapé une branche.
- \*ADU: ils ne cueillent pas quelque chose?
- \*CHI: des raisins.
- \*ADU: pourquoi ils cueillent des cerises?
- \*CHI: parce que ils vont les manger.
- \*ADU: que fait la maman?
- \*CHI: elle prépare à manger.
- \*ADU: où sont ils?
- \*CHI: dans la maison.
- \*ADU: comment fait elle le gâteau maman?
- \*CHI: à céréales.
- \*ADU: où est le chat?
- \*CHI: par terre il mange des cerises.
- @End

\*CHI: au magasin.

- \*ADU : quand est ce que ça se passe ?
- %act : pas de réponse
- \*ADU: ça se passe le jour ?
- \*ADU: qu' est ce qui se passe sur ces images?
- \*CHI: opro :subj y a un sa [: chat] et pis un oiseau.
- %pho : sa
- \*CHI: et pis le monsieur il va xx.
- \*ADU: comment sont ils montés dans l'arbre?
- \*CHI: il tient le bois.
- \*CHI: pis après ils montent.
- \*ADU : que se passe t il sur cette image ?
- \*CHI: ils mangent des cerises.
- \*ADU: Que fait la maman?
- \*CHI: parce qu' elle fait un gâteau.
- \*CHI: c' est l' heure de manger.
- \*CHI: ils mangent des cerises.
- \*ADU: où sont ils?
- \*CHI: dans la maison.
- @End

ADU = Adulte  
 CHI = Children  
 Opro : subj = pas de pronom sujet  
 % err = erreur lexicale  
 % pho = erreur phonologique  
 % act = erreur pragmatique

La longueur moyenne des énoncés est un bon indicateur de la complexité morphosyntaxique. Ainsi, Ophélie 4 ans 9 mois, produit des structures sujet + verbe + préposition, sujet + verbe + complément d'objet direct, elle utilise plus de pronoms. Le corpus d'Ophélie ne peut être comparé à celui des 3 enfants précédemment présentés car la longueur moyenne de ses énoncés est supérieure : 6,24.

Ophélie 4 ans 9 :

\*CHI: ils s'amusaient dans la cour.  
 \*CHI: là elle partait à la montagne et pis en vacances.  
 \*CHI: elle elle partait avec elle.  
 \*CHI: et lui il balançait le panier.  
 \*CHI: le (n) enfant i [: il] grimpe.  
 \*CHI: le sa [: chat] i [: il] xx de xxx les (z) oiseaux.  
 \*CHI: paque [: parce que] la madame i [: il] [\*] xxx.  
 \*CHI: et i [: il] revenait ici.  
 \*CHI: paque [: parce que] c' était celle+là.  
 \*CHI: là elle est partie serser [: chercher] son xx.  
 \*CHI: ce sa [: chat] i [: il] boit la soupe.  
 \*CHI: et pis i [: il] manze [: mange] tous les raisins de papa.  
 %pho: ma~z

\*ADU: qui vois tu?  
 \*CHI: la voiture i [: il] regarde la voiture.  
 \*ADU: que vois tu encore sur cette image?  
 \*CHI: j' [: je] vois l' enfant.  
 \*CHI: ben l' [: le] panier.  
 \*ADU: il n' y a pas quelque' un sur cette image?  
 \*CHI: si.  
 %act: l' enfant montre.  
 \*ADU: qu' est ce qu' ils font?  
 \*CHI: elle s' appelle queque [: quelque] soz [: chose].  
 %pho: kekoesoz  
 \*CHI: i [: en] train d' [: de] s' amuser d' [: de] regarder le soleil.  
 \*CHI: li [: lui] i [: il] balan [: balance] l' [: le] panier.  
 %pho: li bala~

\*ADU: combien sont ils?  
 \*CHI: quatre un deux trois quatre.  
 \*ADU: où sont ils?  
 \*CHI: à la mairie.  
 \*ADU: quand est ce que ça se passe?  
 \*CHI: ça se passe à la maison paque [: parce que] ça c' est la maison à lui  
 %pho: pakoe  
 \*ADU: ça se passe le jour?  
 \*CHI: la nuit.  
 \*ADU: pourquoi sont ils là?  
 \*CHI: paque [: parce que] i [: ils] s' amusent.  
 \*CHI: et pis elle s' appelle Miranda elle.  
 \*ADU: qu' est ce qui se passe sur ces images?  
 \*CHI: i [: il] voit son gason [: garçon] avec le grand (n) arb [: arbre].  
 %pho: gaso~ narb  
 \*CHI: paque [: parce que] c' est ça i [: il] est embêté par une xx.  
 %pho: pakoe  
 \*ADU: ils ne cueillent pas quelque chose?  
 \*CHI: si i [: ils] cueillent des sorises [: cerises] et du poison.  
 %pho: soriz  
 \*ADU: qu' est ce qu' ils cueillent?  
 \*CHI: des cerises et du poison.  
 \*ADU: pourquoi ils cueillent des cerises?  
 \*CHI: paque [: parce que] c' est pour les manger.  
 %pho: pakoe  
 \*ADU: qu' est ce qui se passe sur ces images?  
 \*CHI: i [: il] a fait tomber les cerises encore.  
 \*CHI: et les ti [: petits] bébés de elle sont xx.  
 \*CHI: et les enfants et le papa i [: il] est là.  
 \*CHI: i [: il] veut pas manzer [: manger].  
 %pho: ma~zer  
 \*CHI: i [: il] manze [: mange] par la fenêtre.  
 @End

L'analyse par le CLAN permet de lister les occurrences propres à chaque catégorie morphosyntaxique comme Parisse et coll \*. En outre, les données longueur moyenne d'énoncé, indice de diversité lexicale et nombre de mots produits témoignent du niveau de l'enfant.

\*Normand, 1998

	Longueur moyenne énoncé	Indice de diversité lexicale	Nombre total de mots	Nombre de mots différents
Alexandre 4 ans 7	3,39	42,93	129	68
Akim 4 ans 9	3,75	19,78	135	70
Alicia 4 ans 1	5,10	34,76	97	55
Ophélie 4 ans 9	6,24	50,34	219	101

	Alexandre		Akim		Alicia		Ophélie	
	OD	OC	OD	OC	OD	OC	OD	OC
Noms	37	55	33	54	22	27	41	59
Déterminants	7	32	7	29	7	19	9	33
Verbes	14	21	11	27	16	21	32	39
Adjectifs	7	7	1	1	0	0	5	6
Prépositions	4	5	5	5	3	3	9	18
Pronoms	6	15	6	25	7	20	10	45
Adverbes	9	15	4	9	3	8	9	18
Conjonctions	3	4	2	6	2	6	3	17

Alicia produit moins de mots mais elle est la plus jeune. L'indice de diversité lexicale de Akim est le plus faible. L'enfant a d'ailleurs bénéficié des séances de stimulation du langage à l'école. Au terme de la première année, son indice de diversité lexicale est supérieur au score moyen du groupe des enfants de 5 ans 3 mois à 6 ans 9 mois.

Pour Ophélie, le nombre plus élevé de conjonctions et de prépositions témoigne de la complexité morphosyntaxique. La présence de ces marqueurs va de pair avec la longueur moyenne des énoncés dont le score est plus élevé.

### RÉSULTATS QUANTITATIFS AU SECOND TEST : ENFANTS DE 5 ANS 3 MOIS À 6 ANS 9 MOIS GROUPE B

La longueur moyenne d'énoncés est de 6,32 mois et l'écart type de 1,48 mois.

Le nombre de mots varie de 661 à 57. La valeur moyenne est de 233 avec un écart-type de 109.

Le nombre de mots différents varie de 42 à 270. La valeur moyenne est de 87 avec un écart-type de 27.

La valeur moyenne de l'indice de diversité lexicale est de 41,73 avec un écart-type de 11,57.

La population a été scindée en 3 groupes d'âge afin de visualiser les évolutions des scores.

### RÉSULTATS QUANTITATIFS DES ENFANTS DE 5 ANS 6 MOIS À 6 ANS 6 MOIS :

		5 ans 5 ans 6 (8filles/8garçons)	5 ans 6 - 6 ans (54 filles/56garçons)	6 ans - 6 ans 6 (44 filles/57garçons)
Longueur moyenne énoncé	moyenne	5,91	6,34	6,38
	écart-type	1,52	1,41	1,55
Nombre de mots	moyenne	273 77 à 601	227 valeur de 81 à 535	237 valeur 72 à 661
	écart-type	131	101	119
Nombre de mots différents	moyenne	95 51 à 152	83 valeur de 36 à 153	88 valeur 41 à 186
	écart-type	26,43	22	27
Indice de diversité lexicale	moyenne	41,51	41,13	44
	écart-type	11,51	8,83	11,89

Comme le dit Le Normand\* le score LME n'a plus la même validité « ce que dit l'enfant est davantage fonction du contexte dans lequel est recueilli le langage que de ses connaissances linguistiques ». Les valeurs de l'indice de diversité lexicale ont augmenté d'un point. Par contre, les enfants disposent de plus de mots comme en témoignent les scores : nombre de mots et nombre de mots différents.

\*Le Normand, 1996

### RÉSULTATS QUALITATIFS DES ENFANTS DE 5 ANS 6 MOIS À 6 ANS 6 MOIS :

Voici deux corpus dont les scores longueur moyenne d'énoncé sont similaires. Les indice de diversité lexicale, nombre de mots et nombre de mots différents se situent dans la moyenne.

**Steven 5 ans 11 :**

\*ADU: qu'est ce qui se passe sur ces images?  
 \*CHI: la voiture i [: il] [\*] va conduire [\*] avec le ti  
 [: petit]  
 (n) enfant et la maman là-bas.  
 %err: il=elle va conduire=va les conduire  
 %pho: ti  
 \*CHI: eux i [: ils] sont en krain [: train] de cueillir des  
 cerises.  
 %pho: kri~  
 \*CHI: après le sa [: chat] i [: il] mange des cerises les  
 souris aussi.  
 %pho: sa  
 \*CHI: les (z) autres i [: ils] font une tarte aux cerises.  
 \*ADU: qu'est ce qui se passe sur ces images?  
 \*CHI: on voit un monsieur qui cueille des cerises et un  
 petit  
 (n) enfant.

\*ADU: qui vois tu?  
 \*CHI: un monsieur.  
 \*CHI: la voiture la maman la sœur la grande soeur le  
 petit frère le papa.  
 \*ADU: que vois tu encore sur cette image?  
 \*CHI: ça une église.  
 \*CHI: i [: ils] vont aller à l' église.  
 \*ADU: qu' est ce qu' ils font?

**Nadia 6ans 2 :**

\*ADU: qu'est ce qui se passe sur ces images?  
 \*CHI: Opro :subj y a une dame avec un garçon.  
 \*CHI: Opro :subj y a de l' herbe et des arbres.  
 \*CHI: Opro :subj y a quelqu' un avec une échelle.  
 \*CHI: Opro :subj y a un soleil et quatre oiseaux.  
 \*CHI: la dame est avec un panier et un chat.  
 \*CHI: la fille est avec un panier aussi.  
 \*CHI: le garçon lui donne la main.  
 \*CHI: il a un panier sur la tête.

\*ADU: qu'est ce qui se passe sur cette image?  
 \*CHI: ils se promènent en voiture.  
 \*ADU: qu' est ce qu' ils font?  
 \*CHI: ils vont couper de les [\*] arbres pour Oadv:neg pas  
 qu' il fait [\*] froid.  
 %err: de les=des fait=fasse  
 \*ADU: combien sont ils?  
 \*CHI: cinq.  
 \*ADU: où sont ils?  
 \*CHI: ils sont dans une forêt.  
 \*ADU: quand est ce que ça se passe?  
 \*CHI: le jour.  
 \*ADU: pourquoi sont ils là?  
 \*CHI: pour se promener dans la forêt.  
 \*ADU: qu' est ce qui se passe sur ces images?  
 \*CHI: Opro :subj y a quelqu' un qui grimpe aux arbres.  
 \*CHI: Opro :subj y a des oiseaux parce que [\*] le chat  
 veut

\*CHI: i [: ils] vont cueillir des cerises avec le monsieur.  
 \*ADU: où sont ils?  
 \*CHI: dehors  
 \*ADU: quand est ce que ça se passe?  
 \*CHI: le zour [: jour].  
 \*ADU: pourquoi sont ils là?  
 \*CHI: pour cueillir des cerises.  
 \*ADU: qu'est ce qui se passe sur ces images?  
 \*CHI: ils cueillir [\*] des cerises.  
 %err: cueillir=cueillent  
 \*ADU: comment sont ils montés dans l'arbre?  
 \*CHI: i [: ils] (z) ont pris une éselle [: échelle].  
 %pho: esel  
 \*ADU: qu'est ce qui se passe sur ces images?  
 \*CHI: i [: ils] font une tarte aux cerises.  
 \*ADU: que fait la maman?  
 \*CHI: elle va faire le tour de la tarte.  
 \*ADU: où sont ils?  
 \*CHI: à l' intérieur.  
 \*ADU: comment fait elle le gâteau maman?  
 \*CHI: elle a pris des (z) oeufs après elle a pris de la farine.  
 et après elle le roule.  
 \*ADU: où est le chat?  
 \*CHI: en dessous de la table.  
 @End

manger.  
 %err: parce que=que  
 \*CHI: Opro :subj y a un tronc avec des feuilles.  
 \*CHI: la maman dit +»/.  
 \*CHI: +» fais attention de ne pas tomber!  
 \*ADU: comment sont ils montés dans l'arbre?  
 \*CHI: sur les échelles.  
 \*CHI: une dame.  
 \*ADU: ils ne cueillent pas quelque chose?  
 \*CHI: des feuilles des cerises avec un panier.  
 \*ADU: pourquoi ils cueillent des cerises?  
 \*CHI: pour manger.  
 \*CHI: c' est bon.  
 \*ADU: qu'est ce qui se passe sur ces images?  
 \*CHI: ils sont dans la cuisine.  
 \*CHI: le monsieur les regarde.  
 \*ADU: que fait la maman?  
 \*CHI: la maman fait des gâteaux.  
 \*CHI: la fille est en train de manger des cerises.  
 \*CHI: c' est une maison avec plein de fenêtres.  
 \*ADU: comment fait elle le gâteau maman?  
 \*CHI: elle roule avec un rouleau.  
 \*ADU: où est le chat?  
 \*CHI: il est par terre en train de manger des cerises.  
 \*CHI: Opro :subj y a des oeufs par terre.  
 \*CHI: Opro :subj y a un frigo.  
 \*CHI: Opro :subj y a une chaise avec des chats et des  
 souris.  
 @End.

	Longueur moyenne d'énoncé	Indice de diversité lexicale	Nombre total de mots	Nombre de mots différents
Steven 5 ans 11 mois	6,9	40,68	145	66
Nadia 6 ans 2 mois	6,38	42,59	217	92

	Steven		Nadia	
	OD	OC	OD	OC
Noms	18	36	33	54
Déterminants	8	46	8	46
Verbes	18	30	19	38
Adjectifs	2	4	2	2
Prépositions	5	9	8	26
Pronoms	8	18	10	27
Adverbes	3	3	6	7
Conjonctions	1	3	3	7

OD : nombre d'occurrences différentes OC : total d'occurrences

Les énoncés sont plus longs, les structures morphosyntaxiques plus complexes, avec des coordonnants, des prépositions, des relatifs.

Seuls les corpus d'Ophélie (1<sup>er</sup> groupe) et de Nadia (2<sup>ème</sup> groupe) sont comparables : La longueur moyenne d'énoncé est similaire 6,24/6,38, le nombre de mots est sensiblement le même 219/217, le nombre de mots différents ne varie pas beaucoup 101/92. L'indice de diversité lexicale plus élevé pour Ophélie va de pair avec le paradigme de marqueurs morphosyntaxiques plus varié.

## DE L'INTÉRÊT DE L'ÉTUDE DES FRÉQUENCES DE MOTS

L'analyse par le CLAN permet aussi de lister les mots les plus fréquents. J'ai relevé les mots lexicaux les plus utilisés par les enfants et me suis arrêtée à la fréquence de 20 comme l'avaient fait Gougenheim et coll\* dans leurs travaux sur le Français fondamental. Il est ainsi possible de déterminer la liste des mots attendus à l'épreuve les cerises du PER2000\*. Par exemple pour les enfants de 3 ans 6 mois à 4 ans, j'ai retenu les mots :

Chat 86, arbre 81, maison 52, monte 51, manger 48, maman 41, monsieur 39, soleil 29, garçon 37, cerises 35, vont 35, deux 34, voiture 32, échelle 28, oiseaux 27, fille 26, papa 25, table 23.

Il est aussi intéressant de lister le paradigme des mots grammaticaux utilisés par tranche d'âge. Voici par exemple le paradigme des prépositions et des conjonctions des 6 corpus présentés par ordre chronologique d'âge.

\*1958

### Alicia :

1 prep:artlau	5 conjlet
1 prep:artde	1 conjparce
1 prepdans	1 conjlque
-----	-----
3 Total nombre de mots différents utilisés	3 Total nombre de mots différents utilisés
3 Total nombre d'occurrences de mots	7 Total nombre d'occurrences de mots

### Alexandre :

2 preplsur	2 conjlque
1 prep:artde	1 conjlaussi
1 prep:artlen	1 conjlet
1 prepdans	-----
4 Total nombre de mots différents utilisés	3 Total nombre de mots différents utilisés
5 Total nombre d'occurrences de mots	4 Total nombre d'occurrences de mots

### Akim :

1 prep:artpar	4 conjlet
1 prepdans	2 conjlque
1 preplpar	-----
1 preplsur	2 Total nombre de mots différents utilisés
4 Total nombre de mots différents utilisés	6 Total nombre d'occurrences de mots
4 Total nombre d'occurrences de mots	

### Ophélie :

2 prep:artdu	10 conjlet
2 prep:artlen	6 conjlque
2 preplavec	2 conjparce
2 preplde	1 conjjsi
1 prep:artde	-----
1 prepdans	4 Total nombre de mots différents utilisés
1 preplpar	19 Total nombre d'occurrences de mots
1 preplpour	
8 Total nombre de mots différents utilisés	
12 Total nombre d'occurrences de mots	

**Steven :**

3 preplde  
 2 prep:artlen  
 2 preplavec  
 1 prep:artlde  
 1 preplpour

3 conjlet

1 Total nombre de mots différents utilisés

3 Total nombre d'occurrences de mots

5 Total nombre de mots différents utilisés

9 Total nombre d'occurrences de mots

**Nadia :**

8 preplavec  
 4 preplde  
 3 prep:artlen  
 3 prepldans  
 3 preplpour  
 2 prep:artipar  
 2 preplsur  
 1 prep:artlde

4 conjlet

2 conjlque

1 conjlavec

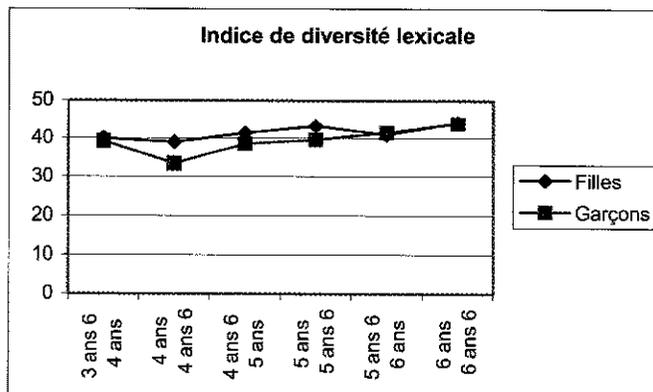
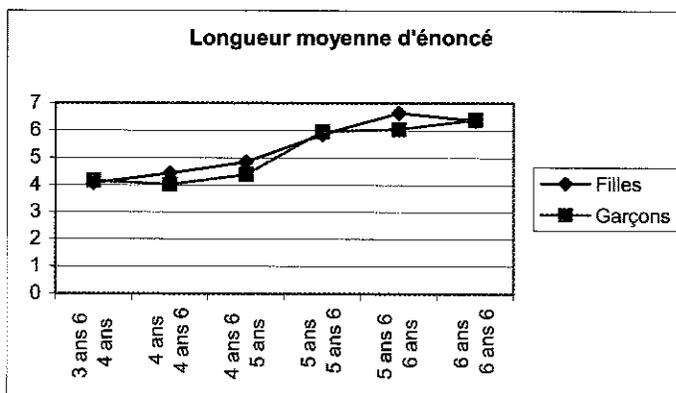
3 Total nombre de mots différents utilisés

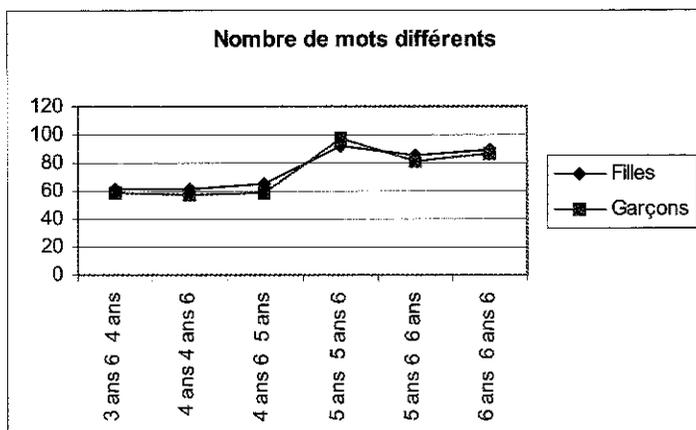
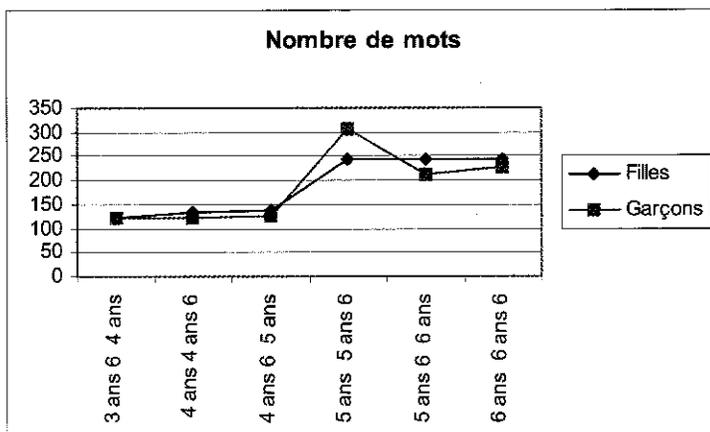
7 Total nombre d'occurrences de mots

8 Total nombre de mots différents utilisés

26 Total nombre d'occurrences de mots

**Résultats quantitatifs comparatifs garçons/filles**





Les scores des garçons sont globalement toujours légèrement inférieurs à ceux des filles mais la différence est moins marquée pour cette population issue d'écoles classées en zone sensible. Les scores longueur moyenne d'énoncés sont inférieurs à ceux de l'étude de Le Normand\*. Les scores en nombre de mots différents sont inférieurs à ceux de la population défavorisée de Le Normand\*.

\*1996

\*1996

## DISCUSSION

Il est indispensable de comparer des corpus dont les longueurs moyennes d'énoncé sont identiques. Même lorsque la situation est standardisée il existe des « petits » parleurs et des « grands » parleurs. Les scores témoignent d'une grande variabilité du nombre de mots produits par les enfants. Sans doute plusieurs facteurs interviennent : démarrage du langage à des âges différents, retards d'acquisition, inhibition, milieux plus ou moins stimulants, différences filles/garçons, enfants uniques, scolarisation tardive etc.... Nous aurions dû en ce qui concerne la partie spontanée limiter ou imposer un temps de parole. La deuxième partie du test permet d'obtenir parfois des réponses plus élaborées. On peut imaginer que nous sommes alors dans une situation conversationnelle, mais on peut aussi penser que la situation est un peu artificielle et incite l'enfant à s'exprimer...

Je pense comme Rondal\* que pour étudier le langage spontané, il faudrait disposer de plusieurs échantillons de langage recueillis dans diverses situations représentatives de l'environnement habituel du sujet.

Le nombre de mots a beaucoup augmenté entre 5 ans et 5 ans 6. Il a été démontré que l'action de suivi des enfants ciblés en difficulté en moyenne section a entraîné une progression importante de leurs capacités linguistiques. (Rapport d'étape recherche action Lille à paraître.

\*1999

S'il est aisé de repérer les enfants dont les capacités articulatoires ne sont pas encore arrivées à maturation l'analyse par le CLAN permet de cibler plusieurs profils déficitaires par une analyse quantitative des énoncés : longueur moyenne d'énoncés, indice de diversité lexicale, nombre de mots sont inférieurs à la moyenne obtenus par les enfants du même âge (cf. tableau annexe) et par une analyse qualitative en observant le paradigme morphosyntaxique des énoncés.

Rageot\* dans son mémoire a recherché les marqueurs communs des déficits concernant les enfants les plus en difficulté de cette étude : au niveau quantitatif, « de nombreuses catégories grammaticales sont absentes ou sous représentées. C'est le cas des adjectifs, des adverbes de négation et de lieu, des conjonctions, des pronoms démonstratifs ainsi que toutes les sous catégories de verbes. Quantitativement, toutes les catégories morphosyntaxiques sont moins représentées... »

Au niveau qualitatif, pour les enfants dont les scores se situent à moins un écart type de la moyenne « Tous ces enfants ont des difficultés à élaborer des phrases. La plupart des énoncés sont du type déterminant nom. Leur corpus est énumératif. Ils ne décrivent pas ce qui se passe. Les enfants répondent aux questions sans construire de phrases complètes : les phrases simples, bien que le plus souvent correctement construites, sont peu nombreuses et apparaissent à la fin du corpus. Il y a peu, voire pas, de propositions coordonnées ou subordonnées. Elles viennent en réponse à la question posée sous l'effet de la stimulation de l'interlocuteur. En général, seul le présent de l'indicatif est utilisé. Les phrases négatives sont absentes. En ce qui concerne les enfants dont les scores se situent à moins deux écarts type de la moyenne, « Le principal marqueur de difficulté pour ces enfants est la quasi absence de langage au premier test à 3 ans » .

En outre, à l'enregistrement des corpus des faits de langue sont apparus fréquemment :

Le pronom il dans la locution il y a est très souvent omis. L'adverbe de négation : ne disparaît dans le Français parlé : « il mange pas ». Le redoublement de sujets est un phénomène très fréquent : 63 % des enfants entre 3 ans et 5 ans et 84 % des enfants entre 5 ans 4 et 6 ans 4.

Le listage des mots les plus fréquents produits par les sujets permet aussi de dresser la liste des mots attendus dans le test. Ce travail pourrait faire partie de l'étape de validation de tests...

## CONCLUSION

Notre pratique clinique de terrain nous donne l'occasion d'enregistrer quantités de corpus de langage. Le programme CLAN en permet une analyse fine tant sur le plan quantitatif que qualitatif. L'enregistrement systématique de données linguistiques d'un même sujet permet d'étudier les faits de langue de manière synchronique et diachronique.

L'absence de langage à trois ans. Dès quatre ans, des scores inférieurs à la moyenne pour la longueur moyenne d'énoncés, diversité lexicale, nombre de mots (en un temps déterminé au préalable), l'absence d'adjectifs, d'adverbes de négation et de lieu, de conjonction, de pronoms démonstratifs et de toutes les sous catégories de verbes constituent des signes d'alerte majeurs de troubles de communication et doivent nécessiter une intervention la plus précoce possible afin de ne pas engager l'enfant dans la spirale de l'échec scolaire.

## BIBLIOGRAPHIE

- CRUNELLE, D. et COLL. (2003), « Dépistage et suivi d'enfants à risques de difficultés scolaires à l'école maternelle », Rapport d'étape d'une recherche action, « <http://www.bienlire.education.fr/o4-media/la-depistage.asp> » <http://www.bienlire.education.fr/o4-media/la-depistage.asp>
- FERRAND, P. A. (1982). PER2000. Isbergues: Ortho édition.
- GOUGENHEIM, G., MICHEA, R., RIVENC, P., SAUVAGEOT, A. (1958). *L'élaboration du Français Fondamental*. Paris : Didier.

- LE NORMAND, M-T. (1996). « *Modèles psycholinguistiques de développement du langage* », *Le langage de l'enfant, aspects normaux et pathologiques*. Paris, Masson.
- MAC WHINNEY, B. (1991) *the CHILDES project (tools for analysing talk)*. Hillsdale, NJ : Erlbaum.
- MILLER, J.F., CHAPMAN, R.S. (1981). «The relation between age and mean length of utterance in morphemes». *Journal of Speech and Hearing Research*, 6, 139-149.
- PARISSÉ, C., LE NORMAND, M-T. (1998). Traitement automatique de la morphosyntaxe chez l'enfant. *Glossa*, 61, 22-29.
- RAGEOT, J. (2006). *A l'écoute de leur langage. Evolution du langage des enfants suivis en classes maternelles pendant deux ans. Existait-il des marqueurs spécifiques pour les enfants les plus en difficulté ?* Mémoire d'orthophonie, Lille.
- RONDAL, J-A. (1999). *Troubles du langage*. Bruxelles : Mardaga.

## ANNEXES

\* Pour des raisons de place nous n'insérons dans l'article que la première des quatre pages de données. L'intégralité de cette annexe peut être obtenue auprès de l'auteur : «mail:chammelrath@wanadoo.fr»  
chammelrath@wanadoo.fr

sujet 1er test	sujet 2ème test	âge 1er test	âge 2ème test	sexe	mlu 1	mlu 2	vocd 1	vocd 2	nombre de mots	nombre de mots 2	nombre de mots différents	nombre de mots différents 2
adoc801	adoc801	4,6	5,11	F	3,68	6,69	52,18	38,11	70	198	46	81
akoc101	akoc101	4,9	6,4	G	3,75	5,45	19,78	44,23	135	170	53	76
aloc106	aloc106	4,9	6,3	G	5,26	7,65	43,60	36,61	119	297	65	95
aloc116	aloc116	4,4	5,11	F	6,10	6,69	22,05	44,47	241	317	69	108
aloc126	aloc126	4,1	5,9	F	5,10	5,95	34,76	31,63	97	133	55	58
aloc208	aloc208	4,4		G	2,16		37,95	X	54		37	
aloc222	aloc222	4,3	5,11	G	5,77	6,74	28,37	30,50	127	182	56	77
aloc35	aloc35	4,7	6,1	G	3,39	7,65	42,93	30,07	129	306	68	89
aloc378	aloc378	3,11	5,7	G	6	5,67	40,14	31,38	127	193	61	66
aloc380	aloc380	4,7	6,1	G	3,62	6,6	27,35	55,97	144	235	63	97
aloc395	aloc395	4,10	6,4	G	6,24	9,25	48,00	32,09	206	324	85	99
aloc501	aloc501	4,8	6,4	F	6,90		46,28	30,44	143	72	70	42
aloc519	aloc519	3,10	5,6	G	4,79		32,90	54,43	157	225	65	93
aloc707	aloc707	4,7	6,1	G	5,68	12,13	43,99	50,82	88	661	204	186
aloc814	aloc814	4,7	6,4	F	7,43	7	55,22	36,90	224	381	100	113
aloc817	aloc817	4,8	X	F	6,12		63,27	X	152		79	
aloc937	aloc937	4,7	6,2	G	3,41	7,51	35,58	36,34	79	279	45	92
amno130	amno130	3,11		F	2,79		62,47	X	67		47	
amno224	amno224	4,1	5,7	F	5,51	8,14	49,62	38,95	161	228	84	85
amoc804	amoc804	3,10	5,5	F	4,62	5,55	33,14	37,23	148	225	65	83
amoc917	amoc917	4,6	6	F	4,62	7,60	29,66	69,21	133	327	58	120
amoc944	amoc944	4,5	6	F	4,41	7,06	30,51	44,10	99	353	53	109
anoc15	anoc15	4,5	6	F	5	7,14	52,80	74,60	174	254	82	112
anoc27	anoc27	4	5,7	F	4,04	6,42	48,70	26,99	97	168	58	65
anoc3	anoc3	3,10	X	F	4,58		44,91	X	196		83	
anoc4	anoc4	3,10		G	3,6	5,81	44,27	X	109		60	
anoc409	anoc409	4,1	5,8	F	6,41	5,81	37,43	33,49	225	194	92	69
anoc613	anoc613	4,4	5,11	F	3,10	5,34	44,84	52,21	115	155	62	77
anoc812	anoc812	4,8	6,3	G	6,46	7,5	41,06	54,85	181	225	80	97
anoc819	anoc819	3,10	5,4	F	2	4,78	0,00	29,69	23	89	17	270
anoc916	anoc916	4,0	5,6	G	5	5,15	49,11	48,93	120	171	65	81
anoc922	anoc922	4,2	5,6	G	4,55	6,13	34,05	50,32	132	227	62	92
aroc518	aroc518	4,2	5,9	G	4,51	4,78	34,15	45,27	156	153	70	69
auoc34	auoc34	4,10		G	4,54		18,04	X	159		60	
auoc909	auoc909	3,6	5,11	F	2,89	5,87	27,44	42,76	81	282	44	90
axoc9	axoc9	3,9	5,5	G	2,63	4,23	38,28	40,78	59	110	38	59
baoc388	baoc388	4,9	6,4	G	6,5	4,23	68,30	45,94	236	183	116	83
beoc407	beoc407	4,6	6,2	G	5,75	5,71	36,80	39,82	192	96	46	53
beoc507	beoc507	4,7	6,2	G	3,23	3,71	33,96	25,83	55	78	35	41
broc111	broc111	4,8	6,4	G	4,85	6,47	38,58	50,49	131	220	60	90
broc33	broc33	4,9		G	0		0,00	X	0		0	
broc805	broc805	3,11		G	3,56		37,81	X	89		50	
broc818	broc818	4,7	6	F	4,29	6,07	31,17	27,22	103	264	49	91
broc908	broc908	4,0		G	4,17		32,87	X	185		79	
broc911	broc911	3,7	5,11	G	5,80	6,31	39,12	38,33	185	274	79	91
broc934	broc934	3,9	5,4	G	5,13	5,51	28,85	16,51	116	321	54	92
caoc214	caoc214	3,11	X	F	4,84		64,66	X	126		73	
caoc701	caoc701	4,0	5,7	F	2,83	9,67	0,00	51,64	34	391	25	125
caoc802	caoc802	4,0	5,6	F	4,17	7,34	30,77	34,19	142	426	63	116
ceoc926	ceoc926	4,0	5,7	F	3,45	5,2	35,40	32,94	83	182	46	70
choc406	choc406	4,0	5,7	F	3,94	4,31	68,46	26,68	76	83	55	42
choc502	choc502	3,9	X	F	3,13		34,82	X	72		43	