

Christophe PARISSÉ
INSERM
MoDyco-Paris-X Nanterre-CNRS
Marie-Thérèse LE NORMAND
INSERM
Hôpital Robert Debré, Paris
parisse@vjf.cers.fr

RÉSUMÉ:

Le contexte de la maison « Fisher-Price » est un support familier de production pour recueillir et analyser du langage spontané chez le jeune enfant. Il s'agit d'une situation quasi-naturelle pouvant être utilisée avec des enfants ayant un âge de développement linguistique entre 2 et 4 ans. Des données qualitatives et quantitatives portant sur l'analyse du lexique et de la syntaxe sont présentées. Toutes ces données issues d'un large échantillon de 316 corpus compilés par les deux auteurs sont analysés avec les outils du CHILDES et plus particulièrement avec le programme POST. L'utilisation de cette méthode est particulièrement recommandée pour l'analyse du langage des enfants très jeunes ou repérés comme pouvant avoir des perturbations, des déficits ou simplement des retards transitoires de langage.*

MOTS-CLÉS:

Langage oral - Langage spontané - Morphosyntaxe - Evaluation.

UNE MÉTHODE POUR ÉVALUER LA PRODUCTION DU LANGAGE SPONTANÉ CHEZ L'ENFANT DE 2 À 4 ANS

par Christophe PARISSÉ et Marie-Thérèse LE NORMAND

SUMMARY : A method for assessing spontaneous language production in children aged two to four.

The "Fisher-Price" play family house is a familiar context to collect and to analyze child spontaneous language production. This is a quasi-natural situation which could be used for children with a language development between age two to four. This paper presents qualitative and quantitative data from 316 language samples on lexicon and syntax. All data used CHILDES tools, especially POST analyzer. Such method is particularly recommended for very young children or for children with suspected delays or deficits in language development.

KEY-WORDS :

Spoken language - Spontaneous language - Morphosyntax - Assessment.

Face à l'absence de dépistage précoce systématique du langage chez l'enfant entre 2 et 4 ans et confrontés quotidiennement aux problèmes de retards du développement du langage, les cliniciens et les thérapeutes ont besoin de disposer pour leur pratique de méthodes d'évaluation du langage adaptées pour le jeune enfant. Ce constat nous a conduits tout d'abord à rechercher des procédures de recueil et une démarche d'évaluation psycholinguistique en utilisant comme support la maison dite « Maison Fisher-Price »*. Ensuite, nous avons présenté les premières données de variation normale de développement qui permettent de différencier clairement retards et troubles spécifiques du langage*. Enfin, nous avons surtout développé de nouveaux logiciels dans le cadre de CHILDES**. Les outils de CHILDES dédiés au lexique et à la morphosyntaxe*** permettent un grain d'analyse du langage si fin que le clinicien peut aujourd'hui poser un diagnostic précoce et orienter l'enfant vers des prises en charge ciblées et spécifiques.

La tâche de la « Maison Fisher-Price » porte essentiellement sur le versant expressif du langage. La difficulté technique d'une telle épreuve est que certains enfants parlant peu, ou parfois même pas tout, il est difficile de les impliquer verbalement dans une procédure d'élicitation verbale induite et dirigée. En effet, si l'enfant parle et comprend peu, comment savoir s'il échoue parce qu'il ne comprend pas les consignes ou parce que son niveau de langage est faible. Il est possible de lever en partie ce type de difficultés en recourant à des méthodes quasi-naturelles utilisant des objets familiers et des jouets attrayants. Cette méthode peut avoir des limites car son aspect non contraignant impose parfois plusieurs séances pour amener à une bonne participation de l'enfant. Toutefois, elle reste intéressante dans le cas de suspicion précoce de trouble ou de retard très sévère. Elle est applicable pour des niveaux de développement linguistique entre 2 et 4 ans ou plus. Elle ne peut être considérée comme faisant partie des outils de dépistage à proprement parler et doit être utilisée en complément d'une évaluation plus large de type questionnaire DPLF* ou de type psycholinguistique comme la BEPL** qui mettent en jeu sur le plan réceptif les stratégies de perception phonétique, de compréhension du lexique, de la morphosyntaxe et sur le plan expressif les stratégies de répétition, de dénomination et de production de la morphosyntaxe.

TRAVAUX SUR L'ÉVALUATION DE LA PRODUCTION DE LANGAGE CHEZ LE JEUNE ENFANT: L'UTILISATION DE LA LONGUEUR MOYENNE D'ÉNONCÉ (LME)

Il existe peu d'études normatives utilisant l'évaluation de production de langage spontané chez le jeune enfant*. La principale étude de ce type est celle de Miller** qui a utilisé la longueur moyenne d'énoncé (LME) pour mesurer le développement du langage de l'enfant et a démontré le lien fort qu'il existe entre LME et âge*. La LME est fortement corrélée avec l'âge, et cette corrélation diminue dans les dernières années de maternelle*. Les normes développées par Miller sont utilisées dans l'instrument d'évaluation du langage SALT**.

La LME est une mesure très fréquemment utilisée dans la littérature pour situer le niveau de développement langagier des enfants. Elle a été utilisée par Brown* qui a introduit l'idée de mesurer la longueur d'énoncé, soit en nombre de morphèmes, soit en nombre de mots. Depuis de nombreux travaux ont utilisé cette mesure de développement du langage des enfants*, même si certains auteurs ont fort justement critiqué cette mesure qui peut donner des résultats erronés dans certains cas**. Les différences entre les calculs en mots et en morphèmes peuvent être importantes, surtout dans les langues qui possèdent beaucoup de morphologie. Ainsi, en français, la plupart des verbes ne forment qu'un seul mot mais contiennent deux morphèmes : la racine et la terminaison (par exemple, « dormons » vs. « dorm + ons »). La même chose s'applique aux pluriels irréguliers, ainsi qu'à certains féminins (par exemple : « ancienne » vs. « ancien- + ne »). En dépit des différences importantes entre les deux types de calcul, il apparaît que les résultats

*Le Normand, 1986 ; Le Normand, 1991

*Le Normand, 1997 ; 1999 ; 2004
 **Child Language Data Exchange System – système d'échanges des données du langage chez l'enfant – voir MacWhinney et Snow, 1985; MacWhinney, 2000
 ***Parisse et Le Normand, 1998; Parisse et Le Normand, 2000a

*Bassano, 2005
 **Chevrie-Muller, Simon, Le Normand, Fournier, 1997

*Grégoire, Rondal et Pérée, 1984; pour une revue, voir Rondal, 1997
 **1981

*Miller et Chapman, 1981

*Miller et Chapman, 1981; Rondal, Ghiotto, Bredart et Bachelet, 1987
 **Miller et Chapman, 1982

*1973

*voir les nombreuses références citées par Parker et Brorson, 2005
 **Meline et Meline, 1981; Klee et Fitzgerald, 1985; Rollins, Snow, Willett, 1996

qu'on obtient sont fortement corrélés. Ceci est vrai pour l'anglais qui est une langue possédant peu de marques morphologiques : Parker et Brorson* ont montré qu'il y a une corrélation de 0.998 entre un calcul réalisé sur les mots et un calcul réalisé sur les morphèmes. Par ailleurs, Parker et Brorson confirment l'existence d'une corrélation entre l'âge et la LME de 0.69 (la même valeur de corrélation est obtenue quelque soit le type de calcul de la LME). Pour vérifier si l'absence de différence entre calcul de la LME en mots ou en morphèmes s'applique au français qui présente beaucoup plus de marques morphologiques que l'anglais, l'étude de Parker et Brorson a été reproduite sur le corpus de cet article. La différence entre les deux méthodes de calcul (LME en mots, LME en morphèmes) est présentée dans l'exemple suivant :

Énoncé : elle fait rentrer son bébé à sa maison

Calcul en mots : pro : subj/elle v : mdl/faire v:inf/rentrer det:poss/son n/bébé
prep/à det : poss/sa n/maison

Calcul en morphèmes : pro : subj/elle-FEM v:mdl/faire-3SV v/rentrer-INF det:poss/son-
MASC-SING n/bébé prep/à det : poss/sa-FEM-SING n/maison

Dans le calcul en morphèmes, on compte les mots et tous les éléments qui suivent un tiret comme -FEM, -INF, -MASC-SING (qui compte pour deux), etc. Dans le calcul en mots, on ne compte que les mots. Pour l'ensemble des enfants de 2 à 4 ans, on obtient une LME moyenne de 3,06 en mots et de 4,83 en morphèmes. En dépit de la large différence entre les deux valeurs, la corrélation entre ces deux calculs est de 0,991. Il a donc été choisi de retenir le calcul en mots qui est plus simple à réaliser et peu sujet à erreur. On conseille de calculer la LME sur au moins 50 mots (LME 50), ou sur tout l'enregistrement (LME).

LE CONTEXTE DE LA MAISON « FISHER-PRICE » POUR ÉVALUER LA PRODUCTION DU LANGAGE

Pour évaluer la production du langage du petit enfant, la meilleure technique reste encore celle de le laisser parler spontanément. En effet, une évaluation plus directive pose des problèmes, comme par exemple de passer du temps pour que l'enfant se familiarise avec l'examineur, ou de vérifier que l'enfant comprenne les consignes qui lui sont adressées. Pour faire parler facilement l'enfant, on essaie, autant que faire se peut, de créer un espace d'échange avec un partenaire familier, souvent l'un de ses parents. Les consignes d'accompagnement sont données verbalement à l'enfant et à l'adulte. Le rôle de l'adulte est d'inciter l'enfant à s'exprimer, à décrire des états, des actions et des événements de la vie quotidienne. Sa tâche est de le suivre en laissant parler l'enfant autant que possible sans intrusion et de l'inciter à parler s'il reste silencieux. sans lui poser de questions mais en formulant des stimulations verbales en utilisant au plus près le contexte de jeu*.

Matériel

Pour comparer les enfants entre eux, il est nécessaire de les impliquer dans une situation standard comme celle de la maison « Fisher-Price ». Il s'agit du modèle de la maison de famille qui comprend un ensemble d'accessoires et de figurines identiques pour tous les enfants :

- une maison de famille de grande taille (4 pièces – salle à manger, cuisine, 2 chambres – et un garage séparé).
- cinq personnages (deux adultes, un enfant, un bébé, un chien).
- seize objets miniatures (un cheval à bascule, une poussette, trois lits, deux tables, deux fauteuils et cinq chaises, deux voitures).

*Le Normand, 1986 ; Le Normand, 1991

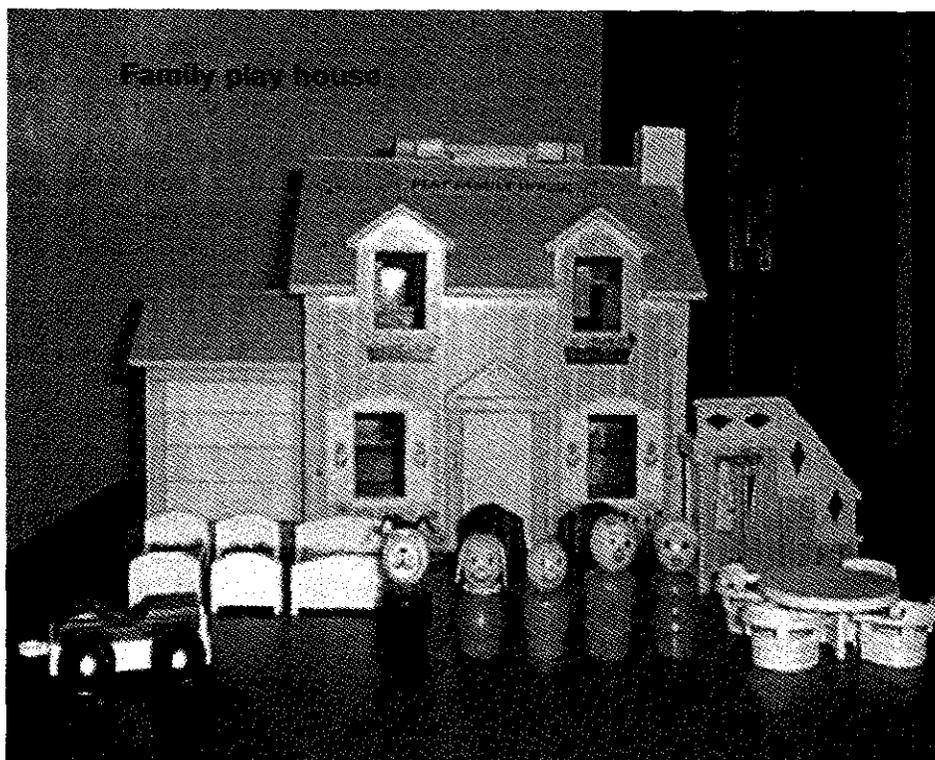


Figure 1 : La maison de famille « Fisher-Price »

Ainsi, le vocabulaire utilisé par les enfants sera relativement uniforme. Par contre, comme on le verra ci-dessous la complexité des énoncés produits et du vocabulaire peut varier de manière importante. La description ci-dessus correspond au modèle 952 de la maison "Fisher-Price". Pour les nouvelles évaluations, il est possible d'utiliser le nouveau modèle en le complétant avec d'autres panoplies de jouets. Certaines précautions doivent être prises cependant. Certains partenaires familiaux parlent beaucoup, souvent dans le but d'amener l'enfant à « obtenir de bons résultats ». Cette attitude peut être en fait contre-productive car l'expérience montre que les enfants parlent encore moins. Il faut donc surtout laisser l'enfant silencieux quelque temps, ce qui peut l'amener à prendre de l'assurance et une fois détendu par la manipulation des jouets à se mettre à parler spontanément lorsqu'il se sent à l'aise. Ce constat est très fréquent chez l'enfant entre 3 et 4 ans. Par ailleurs, pour certains enfants inhibés dont l'entourage atteste qu'ils parlent chez eux « plus que lors de l'examen », il peut être nécessaire de réaliser une seconde séance qui peut amener des résultats très différents. En effet, même si les différences entre séances amènent rarement une grande différence de complexité du langage produit, le nombre d'énoncés obtenus peut varier énormément. Comme pour des besoins statistiques, un minimum de 50 énoncés est exigé pour obtenir une bonne évaluation, dans les cas de faible production une seconde séance est à prescrire.

Transcription

Pour pouvoir noter la prestation et l'état des connaissances lexico-grammaticales de l'enfant, un enregistrement vidéo (éventuellement audio) est nécessaire. En effet, il n'est pas possible de transcrire les productions de l'enfant en temps réel, sauf pour les enfants qui parlent très peu. Toutefois, une transcription en temps réel, même partielle, ou une prise de notes ne doit pas être négligée lorsque c'est possible. Elle peut se révéler une aide précieuse ou même indispensable lors de la transcription de l'enregistrement. Dans la plupart des cas, il faut revoir ou réécouter l'enregistrement pour réaliser une transcription de qualité. Lorsque les enfants sont inintelligibles ou parlent à voix chuchotée, la vidéo peut être nécessaire pour réaliser un décodage correct. En effet, faute de comprendre ce que l'enfant a voulu dire le décodage est souvent impossible.

Les transcriptions peuvent être réalisées à la main ou plus facilement en utilisant des outils informatiques comme par exemple CLAN (<http://childes.psy.cmu.edu>) de CHILDES*, qui autorisent à transcrire en phonèmes comme en graphèmes et permettent de lier le son ou la vidéo aux transcriptions. Il existe plusieurs formats de transcription. On distingue deux grandes familles de transcriptions, les transcriptions phonétiques ou phonologiques dans lesquelles on fait une transcription en phonèmes et les transcriptions graphémiques dans lesquelles on utilise les conventions de la langue écrite. Par ailleurs dans chaque famille on pourra inclure ou non des indicateurs supplémentaires (durée, intonation, éléments pragmatiques, chevauchement entre locuteurs).

Il peut être nécessaire de réaliser une transcription phonétique lorsque les mots sont très déformés. Par exemple, dans des cas de dyspraxie on peut trouver des productions comme par exemple /a a õ a œ œ / produit pour « xx xx maison xx monsieur » ou « xx » est un mot non identifiable. La transcription phonétique permet de contrôler lors de l'évaluation le niveau effectif de langage de l'enfant. Par contre les phonèmes ne permettent pas de savoir ce que l'enfant a voulu dire et ne permettent donc pas de faire de statistiques sur son développement lexical. C'est pourquoi on est amené à compléter la transcription phonétique par une indication graphémique qui correspond au mot attendu. Cette forme qu'on appelle forme « redressée » permet de réaliser des statistiques lexicales ou syntaxiques. Toutes les données présentées ici à propos de la maison « Fisher-Price » ont été lexicalement redressées afin de pouvoir être analysées à partir de leur forme graphémique. Pour une analyse phonologique fine de langage spontané de jeunes enfants, on pourra se référer à Parisse et Maillart** ou à l'outil Phon d'Yvan Rose en cours de développement***.

Lorsque l'enfant a une prononciation de bonne qualité, on réalise directement une transcription graphémique sans intermédiaire phonologique. Au besoin, on redressera les mots de l'enfant. Il faut toutefois être très rigoureux lors de cette opération car elle conditionne la qualité de l'évaluation de langage subséquente. Il faut évidemment faire attention à ne pas « améliorer » le langage de l'enfant lors de la transcription.

Découpage en énoncés de la parole recueillie de l'enfant

Le découpage en énoncés et en mots de la parole recueillie de l'enfant doit suivre des règles les plus strictes possible pour que les mesures de productivité, les calculs de longueur moyenne d'énoncés (LME) et les mesures lexicales soient fiables (voir ci-dessous). Malheureusement il n'existe pas de consignes parfaites car la bonne segmentation en énoncés dépend souvent de l'enfant, de son âge et de la situation. On utilise trois critères à appliquer au mieux et en privilégiant le critère le moins « ambigu » pour une transcription donnée :

- un énoncé doit respecter une logique syntaxique et être la plus courte construction syntaxique indépendante du contexte (d'un point de vue syntaxique),
- un énoncé correspond à une et une seule courbe intonative (montante, descendante, alternée),
- un énoncé est limité (avant ou après) par un silence (par définition d'au moins 400 millisecondes) ou un tour de parole (c'est à dire l'intervention d'un autre locuteur).

Aucun critère n'est absolu et, selon les circonstances, ils peuvent se contredire entre eux (par exemple pour un enfant avec une diction très lente, le critère 3 primera sur le critère 2). Les énumérations ou les suites de constructions sans verbe doivent être divisées selon les critères intonatifs (courbe intonative ou reprise de souffle). Les interjections accolées à un énoncé comme « oh », « ben », sont incluses dans l'énoncé si elles forment un ensemble intonatif cohérent. S'il y a un silence, même court, entre une exclamation et un énoncé, on découpera l'ensemble en deux énoncés.

Découpage en mots

Le découpage en mots est optionnel dans le cas d'une transcription phonologique, obligatoire dans le cas d'une transcription graphémique. Il se fait sur la base de la langue

écrite. Il existe des exceptions qui concernent essentiellement les formes composées comme « pomme de terre » et les formes toutes faites comme « d'accord », « tout à fait », qui sont considérées comme ne formant qu'un seul mot. Il en est de même pour les formes « compactées » dans leur prononciation comme « il y a » souvent prononcé /ja/. Inversement, on séparera le pronom figurant à la fin de constructions verbales comme dans « donne-le ». La liste complète de ces formes composées peut être trouvée sur le site de CHILDES (<http://childes.psy.cmu.edu>) en téléchargeant la version française de l'outil MOR de CLAN.

EXPLOITATION DES TRANSCRIPTIONS

Analyses générales : productivité, LME et diversité lexicale

Plusieurs évaluations très générales des caractéristiques de l'enregistrement sont couramment pratiquées. Elles sont faciles à réaliser, surtout si l'on a pris le soin de transcrire les données avec un outil tel que CLAN de CHILDES. Trois types de mesures sont couramment utilisées.

Une première mesure est la productivité qui est une mesure de fluence calculée en comptant *le nombre total d'énoncés* ou de mots. Le nombre total de mots produits, c'est-à-dire en comptant les mots autant de fois qu'ils sont répétés, sera appelé ci-dessous *nombre d'occurrences*. Le nombre de mots différents produits, c'est-à-dire en ne comptant qu'une seule fois les mots répétés, sera appelé ci-dessous *nombre de types*. Les valeurs de productivité peuvent être normalisées en fonction de la durée effective de l'enregistrement lorsque l'on n'a pas respecté la durée théorique de 20 minutes.

Une seconde mesure est la *longueur moyenne d'énoncés* (LME). Cette mesure a été présentée dans l'introduction ci-dessus et est un des moyens les plus efficaces de comparer des travaux d'origines différentes. C'est aussi un bon moyen de comparer les enfants entre eux.

Une troisième mesure couramment utilisée est le *rapport entre nombre de types et nombre d'occurrences* (rapport types/occurrences : RTO). Cette mesure fournit une image de la richesse lexicale de l'enfant. Le rapport du nombre de types sur le nombre d'occurrences permet de calculer le rapport « types/occurrences » qui est parfois utilisé pour identifier les enfants ayant des troubles de langage. Il s'agit d'une mesure de variabilité. Lorsque cette valeur est très basse (faible variabilité), c'est un indice d'une grande pauvreté de vocabulaire et de retard de langage. Comme la valeur du RTO diminue en fonction du nombre de mots pour laquelle elle est calculée, il faut la calculer pour un nombre de mots constant. A cet effet, deux valeurs ont été calculées, la RTO 50 calculée pour 50 mots et la RTO 200 calculée pour 200 mots.

Analyse morphosyntaxique

L'analyse morphosyntaxique des corpus permet de dégager des statistiques d'usage lexical par catégorie syntaxique, ce qui permet de déduire le développement de constructions syntaxiques liées à des catégories précises (ex : déterminant, pronom personnel). Cette analyse consiste à déterminer quelle est la catégorie syntaxique d'un mot, par une analyse en constituants immédiats. La liste des catégories utilisées est présentée dans le tableau A de l'annexe. Cette liste n'inclut que les catégories de base. D'autres analyses plus fines pourraient être réalisées en incluant notamment le temps des verbes, les indications de personnes, de nombre ou de genre.

L'analyse morphosyntaxique peut être longue à réaliser car en français beaucoup de mots peuvent avoir plusieurs catégories hors contexte. Par exemple, « le » peut être un article (« le manteau ») ou un pronom personnel objet (« il le donne »). L'analyse morphosyntaxique fastidieuse de transcriptions peut être largement facilitée par des outils comme MOR et POST de CHILDES. L'analyse automatique réalisée en quelques secondes par ces deux outils donne des résultats proches de ceux d'une personne travaillant à la main (compte tenu de la fatigue et des éventuelles incertitudes de notation). Une vérification manuelle est toujours possible pour améliorer la qualité de l'analyse.

Les tableaux B1 et B2 de l'annexe contiennent, pour chaque catégorie, les mots attestés chez 25 % et chez 50 % au moins des enfants et l'âge auquel ils apparaissent. Ces mots figurent en forme complète, y compris les éventuels accords qui correspondent aux situations d'énonciation. Ainsi, pour la catégorie nom (n), le mot « voiture » apparaît dès 24 mois, tandis que « voitures (pluriel) » apparaît à 45 mois. Pour cet exemple de mot très courant, on ne constate pas de différence entre la production chez 25 % et chez 50 % des enfants. Pour un autre exemple, celui des mots de la catégorie verbe modal lexical (v : *mdllex*), « fait » apparaît à 27 mois et « fais » à 42 mois pour 25 % des enfants. Par contre, on trouve des valeurs différentes pour 50 % des enfants : « fait » apparaît à 30 mois et « fais » n'apparaît pas. Ces tableaux ont pour but de mieux visualiser quels mots composent effectivement chaque catégorie, de servir de support à la discussion des résultats et d'aider les cliniciens lors des évaluations de langage.

Statistiques réalisées à partir de l'analyse morphosyntaxique

Une fois que les mots sont classés en fonction de leur catégorie morphosyntaxique, deux catégories de calculs statistiques peuvent être réalisées pour chaque catégorie : un calcul du nombre d'occurrences total (token) et un calcul du nombre de mots différents (types). Ces catégories de valeurs peuvent être calculées de deux manières : soit en valeur absolue (nombres bruts), soit corrigées en pourcentages par rapport au nombre total d'occurrences ou de types. Il y a donc quatre manières principales de réaliser des statistiques à partir de l'analyse morphosyntaxique : sur les occurrences en valeur absolue, sur les occurrences en pourcentages, sur les types en valeur absolue, sur les types en pourcentages. Sur ces quatre manières, les trois premières sont couramment utilisées dans la recherche et pour l'évaluation. Elles ne sont pas redondantes car elles mesurent des réalités différentes et présentent des intérêts complémentaires. Le dernier type de calcul est rarement utilisé et ne sera pas présenté ci-dessous pour limiter la taille de l'article.

Le calcul des occurrences permet d'évaluer la fréquence d'usage d'une catégorie (ou d'une structure syntaxique si elle requiert l'usage d'une catégorie spécifique). Il donne des informations de productivité (sur 20 minutes) et d'usage sur toutes les catégories syntaxiques, y compris les catégories ayant peu de représentants. En cela, il s'oppose aux autres types de calcul. Ainsi, les valeurs calculées en pourcentage ne donnent pas d'indication de la productivité puisque par définition on normalise les valeurs par la fluence globale de l'enfant. Les valeurs calculées en types ne donnent pas non plus d'information de productivité. En effet, une catégorie ne contenant que peu de représentants peut n'avoir que peu de types alors que le nombre d'occurrences peut être grand. De plus, l'usage des valeurs absolues est préféré par certains statisticiens car les variations ne sont pas atténuées par le calcul de pourcentages.

L'utilisation des pourcentages permet de comparer aisément des enregistrements de longueurs différentes. Comme on divise le nombre de mots comptabilisés dans chaque catégorie par le nombre total de mots produits, on ne tient pas compte de la fluence de l'enfant. Ce calcul permet d'atténuer l'effet de catégories très fréquentes comme le nom ou le verbe et renforce les caractéristiques des catégories rares. Il permet aussi d'entrevoir des phénomènes de réduction d'usage de certains éléments, car ces réductions n'apparaissent qu'en proportion des autres catégories, et non pas en valeur absolue. Enfin, l'usage des pourcentages permet de mesurer l'usage systématique ou non d'une catégorie. Par exemple, sachant que la plupart des noms sont précédés d'un déterminant, le pourcentage de production de déterminant doit être proche de celui des noms. Une différence importante indique un usage pas encore systématique. Cette valeur est usuellement calculée en pourcentage car ainsi elle ne dépend pas de la taille des corpus.

Les valeurs calculées en types ont un usage plus qualitatif. Elles permettent de voir si un enfant a une production variée et une maîtrise approfondie d'une catégorie. Par exemple, un nombre réduit de pronoms personnels montre un usage relativement figé du verbe, alors qu'un plus grand nombre de pronoms suggère un accès plus analytique et plus maîtrisé. Ces valeurs sont en général calculées en valeur absolue non corrigée, car elles sont directement influencées par la production lexicale de l'enfant.

Représentativité des résultats par catégorie syntaxique

Comme les résultats présentés ci-dessous sont nombreux, il est nécessaire de chercher à savoir quels résultats sont les plus intéressants à exploiter. Plusieurs méthodes peuvent être utilisées pour choisir les catégories les plus intéressantes pour un type de mesure. On peut d'abord vérifier si l'évolution d'une catégorie suit celle de l'âge des enfants. Si c'est le cas, on peut penser que la catégorie est un bon indicateur de développement. Une meilleure méthode pour vérifier cette évolution avec l'âge est de calculer les corrélations avec l'âge des enfants. Plus forte est la corrélation, plus la catégorie ainsi corrélée sera un bon indicateur.

Une autre manière est de repérer les résultats qui sont les plus représentatifs. Pour cela les valeurs d'écart-type sont fournies avec tous les résultats ci-dessous. Non seulement, ces valeurs permettent d'apprécier l'écart entre un enfant que l'on désire tester et la population de référence, mais elles permettent aussi de calculer le coefficient de variabilité pour chaque catégorie (écart-type divisé par la moyenne). Si ce coefficient est faible, alors cela indique que l'usage d'une catégorie est relativement stable pour un âge ou une LME donné. Dans ce cas, on peut penser que les valeurs statistiques sont assez fiables. Inversement, un coefficient élevé suggère qu'une catégorie a un usage variable, différent d'un enfant à l'autre, et que, soit cette catégorie est au début de son apprentissage, soit il faut être prudent en utilisant les valeurs statistiques obtenues.

Pour un maximum de clarté, les résultats d'évolution, de corrélation et de variabilité ne seront pas donnés dans les tableaux de résultats, mais séparément dans le tableau C de l'Annexe. Cette présentation a aussi l'avantage de permettre de visualiser les différences de nature entre types de calcul : en occurrences brutes, en pourcentages d'occurrences, en types.

RÉSULTATS STATISTIQUES

Productivité, LME et RTO

Tableau 1 : Caractéristiques générales par âge du corpus, longueur moyenne d'énoncé (LME) et rapport types-occurrences (RTO)

AGE (mois)	24	27	30	33	36	39	42	45	48
Nb enfants	40	31	36	36	40	33	34	34	31
Nb énoncés	66(41)	(35)	88(38)	109(40)	113(45)	129(63)	116(68)	115(59)	108(54)
Nb occur.	108(84)	185(139)	225(132)	364(169)	397(210)	457(233)	464(340)	475(306)	454(270)
Nb types	43(25)	66(39)	86(37)	120(42)	130(47)	141(47)	147(66)	150(57)	156(65)
LME	1,5(0,4)	2,1(0,8)	2,5(0,7)	3,2(0,7)	3,4(0,9)	3,5(0,5)	3,7(1,1)	4,0(0,7)	4,0(1,0)
LME 50	1,5(0,4)	2,1(0,7)	2,4(0,6)	3,2(0,7)	3,3(0,9)	3,4(0,5)	3,6(1,1)	3,9(0,7)	4,0(1,0)
RTO 50	0,59(0,1)	0,53(0,1)	0,60(0,1)	0,64(0,1)	0,65(0,1)	0,66(0,1)	0,68(0,1)	0,66(0,1)	0,69(0,1)
RTO 200	0,50(0,2)	0,41(0,1)	0,44(0,1)	0,43(0,1)	0,45(0,1)	0,45(0,1)	0,48(0,1)	0,46(0,1)	0,49(0,1)

Note : les écarts-types sont présentés entre parenthèses.

Le tableau 1 contient toutes les valeurs de productivité, de LME et de RTO. On constate que pour une durée de 20 minutes, les performances moyennes en productivité des enfants évoluent de manière importante de l'âge de 2 à 3 ans, et qu'elles stagnent de 3 à 4 ans. Plus précisément, il semble que le nombre d'énoncés stagne à partir de l'âge de 2 ans 9 mois, le nombre d'occurrences et de types à partir de l'âge de 3 ans 3 mois. Ces résultats sont confirmés par les calculs de corrélation. Les valeurs de corrélation sont données dans le tableau 2.

Les enfants de 3 ans ne sont pas inclus dans le calcul spécifique des enfants de moins de 3 ans et celui des enfants de plus de 3 ans pour égaliser le nombre d'enfants dans les deux groupes et améliorer la séparation entre les deux groupes. Les valeurs de r inférieures à 0,50 sont à prendre avec prudence, même lorsqu'elles sont significatives en raison du grand nombre de sujets (316 en tout, 143 de moins de 3 ans, 133 de plus de 3 ans).

On observe dans le tableau 2 que les corrélations de nombre d'énoncés, d'occurrences, de types et de LME sont significatives de l'âge de 2 ans à l'âge de 3 ans, mais ne le sont pas de l'âge de 3 à 4 ans. Seul le RTO ne donne pas de résultat significatif (sauf RTO 50 sur l'ensemble des enfants).

Tableau 2 : Corrélations entre l'âge des enfants et les indicateurs de productivité, LME et RTO

Corrélations sur l'intervalle allant...	de 2 à 4 ans	de 2 à 2 ; 9 ans	de 3 ; 3 à 4 ans
Enoncés	$r = 0,28 (z = 5,18)****$	$r = 0,37 (z = 4,64)****$	$r = -0,12 (z = -1,35)$
Types	$r = 0,60 (z = 12,26)****$	$r = 0,63 (z = 8,78)****$	$r = 0,09 (z = 1,04)$
Occurrences	$r = 0,48 (z = 9,30)****$	$r = 0,57 (z = 7,67)****$	$r = 0,00 (z < 1)$
LME	$r = 0,71 (z = 15,79)****$	$r = 0,70 (z = 10,42)****$	$r = 0,23 (z = 2,70)**$
LME 50	$r = 0,71 (z = 15,60)****$	$r = 0,71 (z = 10,61)****$	$r = 0,28 (z = 3,25)**$
RTO 50	$r = 0,37 (z = 6,87)****$	$r = 0,21 (z = 2,40)$	$r = 0,09 (z = 1,06)$
RTO 200	$r = 0,08 (z = 1,47)$	$r = -0,17 (z = -2,14)$	$r = 0,17 (z = 1,97)$

Note : ** signifie que la corrélation est significative avec $p < 0.01$ et $p < 0.001$, **** signifie que la corrélation est significative avec $p < 0.0001$. Les autres corrélations ne sont pas significatives.

Un autre calcul de corrélation réalisé pour les mêmes âges entre les deux valeurs de LME et de RTO confirme que ces valeurs peuvent être calculées sur le plus petit nombre d'éléments. En effet, on trouve une corrélation de $r = 0,97 (z = 38,27)$ entre le LME et le LME 50, une corrélation de $r = 0,63 (z = 13,14)$ entre le RTO 50 et le RTO 200. Les corrélations obtenues pour les enfants avant 3 ans (corrélation LME et LME 50, $r = 0,98, z = 27,20$; corrélation RTO 50 et RTO 200, $r = 0,69, z = 9,96$) ou pour les enfants après 3 ans (corrélation LME et LME 50, $r = 0,93, z = 19,37$; corrélation RTO 50 et RTO 200, $r = 0,54, z = 6,89$) confirment ces résultats.

Tableau 3 : Moyenne et écart-type par âge du nombre total de mots produits par catégorie syntaxique (occurrences)
(*ci-contre*)

AGE (mois)	24	27	30	33	36	39	42	45	48
adj	3 (3)	5 (5)	6 (5)	9 (7)	8 (7)	7 (5)	8 (8)	8 (6)	10 (8)
adv	3 (3)	3 (4)	5 (5)	8 (7)	11 (10)	11 (10)	15 (16)	16 (13)	16 (14)
adv: int	1 (2)	1 (1)	2 (5)	3 (5)	2 (3)	2 (3)	2 (3)	2 (3)	2 (3)
adv: neg	3 (3)	7 (10)	6 (5)	11 (8)	11 (8)	13 (11)	11 (12)	12 (11)	11 (8)
adv: place	11 (10)	16 (10)	17 (12)	20 (13)	21 (14)	27 (22)	20 (18)	25 (20)	17 (13)
adv: yn	3 (5)	5 (5)	6 (5)	7 (6)	9 (10)	14 (17)	10 (11)	10 (11)	10 (15)
co	7 (7)	6 (7)	5 (3)	6 (7)	6 (6)	14 (16)	12 (16)	12 (14)	7 (9)
co: act	8 (10)	8 (7)	6 (6)	9 (6)	9 (10)	13 (14)	12 (18)	8 (8)	8 (8)
conj	1 (2)	2 (5)	3 (3)	6 (6)	10 (13)	14 (17)	17 (20)	17 (16)	21 (23)
det*	6 (9)	18 (17)	23 (17)	43 (24)	45 (25)	48 (29)	51 (34)	55 (39)	50 (30)
det: dem	0 (0)	0 (0)	0 (0)	0 (0)	0 (1)	0 (1)	1 (1)	1 (1)	1 (1)
det: poss	0 (0)	0 (1)	2 (4)	4 (5)	4 (4)	4 (4)	6 (8)	4 (3)	5 (4)
n*	22 (18)	33 (20)	39 (22)	60 (29)	59 (29)	61 (32)	65 (41)	68 (44)	65 (38)
n:prop	4 (6)	5 (6)	4 (4)	3 (5)	5 (6)	3 (3)	4 (4)	7 (8)	6 (8)
num	0 (0)	0 (0)	0 (1)	0 (1)	1 (2)	1 (1)	0 (1)	0 (1)	1 (2)
prep*	0 (1)	5 (7)	7 (8)	14 (8)	14 (10)	15 (10)	18 (15)	19 (12)	18 (11)
prep:art	1 (1)	1 (2)	2 (2)	3 (4)	3 (3)	4 (4)	5 (4)	3 (3)	4 (4)
pro	1 (2)	3 (3)	4 (4)	6 (6)	7 (5)	8 (5)	10 (9)	9 (8)	9 (7)
pro: dat	0 (0)	0 (0)	0 (1)	1 (1)	2 (2)	1 (1)	2 (3)	2 (2)	2 (2)
pro: dem	4 (4)	5 (5)	7 (6)	11 (8)	10 (7)	10 (11)	10 (8)	8 (8)	10 (8)
pro: int	1 (1)	2 (3)	2 (4)	5 (7)	5 (5)	4 (3)	4 (5)	5 (6)	4 (4)
pro: obj	0 (1)	1 (1)	1 (1)	2 (2)	3 (5)	4 (5)	2 (2)	4 (4)	3 (3)
pro: refl	0 (1)	1 (3)	1 (2)	3 (3)	4 (4)	5 (5)	6 (6)	4 (4)	7 (6)
pro: rel	0 (0)	0 (0)	0 (0)	2 (2)	2 (2)	1 (1)	2 (3)	2 (2)	2 (3)
pro: subj*	4 (8)	11 (18)	16 (14)	32 (19)	42 (27)	49 (29)	52 (42)	51 (30)	51 (30)
pro: y	0 (0)	0 (1)	0 (0)	0 (1)	1 (1)	1 (1)	2 (4)	1 (2)	1 (2)
v	3 (4)	9 (11)	12 (10)	16 (11)	18 (13)	23 (16)	23 (18)	21 (18)	21 (14)
v: aux	2 (3)	3 (5)	5 (5)	8 (6)	9 (7)	8 (6)	8 (6)	8 (6)	8 (6)
v: exist*	6 (7)	10 (11)	12 (9)	24 (13)	23 (17)	24 (13)	25 (23)	25 (16)	21 (18)
v: inf	4 (6)	8 (8)	11 (8)	15 (7)	16 (12)	23 (12)	18 (14)	21 (16)	19 (10)
v: mdl*	1 (2)	4 (7)	6 (7)	13 (9)	16 (12)	23 (15)	19 (16)	24 (18)	21 (15)
v: mdllex*	1 (1)	3 (4)	5 (4)	9 (8)	10 (6)	11 (8)	14 (12)	13 (10)	13 (8)
v: poss	1 (1)	1 (2)	2 (2)	3 (3)	3 (3)	4 (4)	4 (5)	4 (3)	4 (4)
v: pp	6 (7)	6 (6)	6 (5)	8 (6)	9 (8)	9 (6)	8 (6)	8 (5)	8 (5)

Note: Les catégories accompagnées d'une astérisque ont une corrélation avec l'âge (de 2 à 3 ans) supérieure à 0,50 et celles qui sont présentées en italique ont un coefficient de variabilité inférieur à 1. Les valeurs entre parenthèses représentent les écarts-types.

Le tableau 3 présente la moyenne du nombre d'occurrences pour l'ensemble des catégories syntaxiques (toutes les valeurs d'évolution, de corrélation et de variabilité sont présentées dans le tableau C de l'annexe). On constate pour la totalité des catégories une augmentation du nombre d'éléments produits avec l'âge. Toutefois, cette augmentation varie dans des proportions importantes selon la catégorie. On peut diviser ces variations en trois grandes tendances : faible (moins de 2 fois plus de mots), moyenne (de 2 à 4 fois plus de mots) et forte (plus de 4 fois plus de mots). Cette division est en partie artificielle mais permet de séparer des grandes tendances. En effet, les catégories qui augmentent faiblement sont probablement acquises tôt, celles qui augmentent fortement sont plus probablement acquises tardivement, et celles qui augmentent moyennement forment l'ensemble de référence par rapport auquel on peut comparer les catégories.

Les faibles évolutions couvrent les catégories *adj*, *adv: int*, *adv: place*, *adv: yn*, *co*, *co: act*, *n: prop*, *pro: dem*, et *v: pp*. Toutes ces catégories ont un indice de variabilité faible (de 0,7 pour *adv: place* à 1,4 pour *adv: int*), ce qui veut dire que les statistiques obtenues sont relativement fiables. La faible évolution des catégories indique qu'elles sont utili-

sées par les enfants dès leur plus jeune âge. A l'exception de la catégorie *v: pp*, aucune de ces catégories ne dispose d'un jeu de marques morphosyntaxiques et dans la plupart des cas, elles ne comprennent qu'un nombre relativement limité de mots. Même pour les deux catégories « ouvertes » (noms propres et communicateurs), le nombre de mots réellement utilisés par un même enfant est limité. Pour les noms propres (*n: prop*), il s'agit de *bébé, maman, papa* (lorsqu'il ne sont pas précédés d'un article), ainsi que du nom de l'enfant et de ses proches. Même si ces noms varient d'une famille à l'autre, pour un même enfant, ils forment un petit ensemble bien circonscrit. Il en est souvent de même pour les communicateurs (*co*) qui comprennent des formules toutes faites ainsi que des interjections. Même si le nombre potentiel d'interjections est infini, en réalité seul un ensemble relativement limité est utilisé dans un même cercle familial. Pour toutes les autres catégories, on est en présence de mots invariables, souvent utilisés de manière isolée, même par l'adulte, et maintes fois répétés car ils correspondent à des situations banales apparaissant dans toutes les interactions langagières des enfants. Ces mots sont parmi les plus faciles à apprendre et les plus précocement produits. En revanche, ils évoluent peu et restent à peu près les mêmes en fin de quatrième année de l'enfant. La présence du verbe au participe passé dans ces faibles augmentations ne s'explique pas bien. Il est possible que l'ambiguïté fréquente entre formes au participe passé et à l'infinitif ait conduit à surestimer le nombre de participes passés chez le jeune enfant et à constater une stagnation là où on est peut-être face à une augmentation moyenne comme les autres catégories de verbes (voir ci-dessous).

Les évolutions moyennes couvrent les catégories *adv: neg, det, n, prep: art, pro, pro: int, v, v: aux, v: exist, v: inf et v: poss*. On trouve dans cette liste les catégories ouvertes de base comme le nom (*n*), et le verbe (*v et v: inf*). Comme ces catégories sont les plus fréquentes et forment la « base » des énoncés, il est logique qu'elles soient statistiquement dans la moyenne de l'évolution de toutes les catégories. Presque toutes ces catégories « moyennes » ont une faible variabilité, ce qui indique que leur usage est généralisé et stable (seules les catégories *pro: int, prep: art* ont un usage moins stable). Les autres catégories sont, soient des marqueurs de base du nom (*det, prep: art*), soit les marqueurs les plus « internes » du verbe (*adv: neg, v: aux, v: exist, v: poss*). Les marqueurs internes du verbe s'opposent aux marqueurs « externes » comme les pronoms personnels (voir ci-dessous augmentations fortes) car ils sont, d'un point de vue distributionnel, plus proches du noyau verbal, et plus à même d'être phonologiquement intégrés dans ce noyau. On peut souligner que la même dichotomie interne vs. externe s'applique aux déterminants et prépositions-articules (*det et prep: art*) qui sont plus internes que les prépositions (*prep*) qui ne figurent pas dans la liste des évolutions moyennes. Les propriétés distributionnelles expliquent certainement que les éléments internes sont produits plus tôt par l'enfant et n'ont plus tendance à se développer à partir de sa quatrième année. Deux catégories ne rentrent pas dans ces analyses, *pro et pro: int* dont l'usage et les caractéristiques rappellent en fait ceux des catégories ayant une faible augmentation comme *adv: place, co: act, pro: dem et même n: prop*. La différence serait que ces catégories seraient plus utilisées chez l'enfant de trois ans et se développeraient plus tardivement et plus longtemps. Ces catégories présentent par ailleurs des propriétés identiques (absence de marques morphosyntaxiques, forte valeur sémantique mise en œuvre de manière quotidienne et systématique dans les interactions de l'enfant).

Les fortes évolutions couvrent les catégories *adv, conj, det: dem, det: poss, num, prep, pro: dat, pro: obj, pro: refl, pro: subj, pro: y, v: mdl et v: mdllex*. Encore une fois, on peut distinguer plusieurs grands types de catégories : les marqueurs externes du verbe (*pro: dat, pro: obj, pro: refl, pro: subj et pro: y*) et les marqueurs externes du nom (*prep*), et des catégories qui techniquement (d'un point de vue distributionnel) ressemblent, soit aux catégories ayant une augmentation faible (*adv, conj et pro: rel*), soit aux catégories ayant une augmentation moyenne (*det: dem, det: poss et num* pour les marqueurs du nom ; *v: mdl et v: mdllex* pour les marqueurs du verbe). Parmi toutes ces catégories, on en trouve peu qui ont une faible variabilité. Ce n'est le cas que pour *prep, pro: subj et v: mdllex*. Ces trois catégories seraient donc acquises plus tardivement que d'autres tout en

étant fréquentes (à partir de trois ans) et d'un usage stable chez l'ensemble des enfants. Par contre, on trouve dans les autres catégories celles qui ont la plus grande variabilité comme *det: dem, num, et pro:y* ou des catégories qui sont très variables chez l'enfant de deux à trois ans (*conj, det: poss, pro: dat, pro: obj, pro: refl et pro: rel*) tout en se stabilisant un peu après trois ans. On peut imaginer que toutes ces catégories dont l'augmentation d'usage est grande et qui se développent tardivement sont d'une certaine façon plus complexes à acquérir. Elles peuvent de ce fait être de meilleurs indicateurs de troubles du langage et amener à plus de travail de rééducation. L'explication de cette complexité serait la position distributionnelle pour les éléments en tête des groupes nominaux ou verbaux, la complexité de la valeur des marques grammaticales (tous les champs spatiaux ou temporels pour les prépositions ; toutes les valeurs modales pour *v:mdl et v:mdllex* ; toutes les relations personnelles et anaphoriques envisageables pour les pronoms préverbaux ; notions numériques ; liens entre énoncés pour les conjonctions), et pour les autres catégories, une fréquence beaucoup plus faible dans le langage adulte*.

*Parisse et Le Normand, 2000b

Les corrélations entre l'âge des enfants et le nombre d'occurrences pour chaque catégorie présentent des valeurs intéressantes proches de celles obtenues entre l'âge et la LME. Pour 24 des 34 catégories testées dans cet article, on obtient une corrélation significative entre l'âge de 2 à 3 ans et le nombre d'occurrences. Par contre, aucune corrélation n'est obtenue pour l'âge de 3 à 4 ans. Certaines catégories (*det, n, prep, pro: subj, v: exist, v: mdl et v: mdllex*) ont même une corrélation supérieure à 0,50. Les catégories qui ne présentent pas de corrélation du tout sont *adv: int, adv: place, co, co:act, det: dem, n: prop, num, pro: y, v: poss et v: pp*. Parmi ces catégories, certaines sont logiquement des catégories qui apparaissent chez les enfants jeunes (voir faibles augmentations ci-dessus). Les autres catégories sont probablement trop peu fréquentes pour amener à des valeurs statistiques significatives.

NOMBRE D'OCCURRENCES EN POURCENTAGES

Tableau 4 : Moyenne et écart-type par âge du pourcentage d'occurrences produits par catégorie syntaxique
(ci-après)

AGE (mois)	24	27	30	33	36	39	42	45	48
<i>adj</i>	2,3(2,3)	2,2(1,6)	2,7(1,9)	2,5(1,8)	2,0(1,7)	1,6(0,8)	1,5(1,0)	1,8(1,2)	2,4(1,6)
<i>adv</i>	2,6(3,0)	1,4(1,2)	2,0(1,9)	2,1(1,3)	2,6(1,6)	2,3(1,3)	2,8(1,8)	3,0(1,3)	3,2(1,9)
<i>adv: int</i>	1,2(1,8)	0,3(0,7)	0,6(1,1)	0,7(1,0)	0,4(0,6)	0,6(0,8)	0,6(0,5)	0,4(0,6)	0,4(0,4)
<i>adv: neg</i>	2,8(2,6)	2,7(2,3)	2,6(1,6)	3,0(1,7)	2,8(1,6)	2,7(1,3)	2,3(1,4)	2,4(1,4)	2,5(1,7)
<i>adv: placē</i>	12,6(9,4)	10,1(5,8)	8,3(5,7)	5,5(2,6)	5,7(3,4)	5,8(2,8)	4,9(3,3)	5,2(2,4)	3,7(1,7)
<i>adv: yn</i>	4,6(6,5)	2,9(1,9)	3,3(3,4)	2,3(2,3)	2,6(3,5)	2,6(1,9)	1,9(1,2)	1,7(1,6)	1,8(1,6)
<i>co⁺</i>	7,8(6,7)	5,1(6,4)	3,3(3,7)	1,9(2,4)	1,6(1,5)	2,7(2,3)	2,2(1,8)	2,0(1,7)	1,3(1,1)
<i>co:act⁺</i>	8,3(8,6)	4,8(4,1)	3,3(3,2)	3,0(3,1)	2,2(2,0)	2,6(1,7)	2,3(1,7)	1,8(1,8)	1,7(1,9)
<i>conj</i>	0,6(1,0)	0,8(2,2)	1,0(1,3)	1,4(1,2)	2,0(1,7)	2,8(2,5)	2,8(2,4)	3,2(1,8)	3,6(2,4)
<i>det*</i>	3,8(4,2)	9,8(5,8)	9,9(4,5)	11,8(3,3)	11,6(3,0)	10,6(3,0)	11,5(3,1)	11,5(2,4)	11,5(2,8)
<i>det:dem</i>	0,0(0,1)	0,0(0,1)	0,0(0,1)	0,0(0,1)	0,0(0,1)	0,1(0,2)	0,1(0,2)	0,1(0,2)	0,1(0,2)
<i>det: poss</i>	0,1(0,2)	0,2(0,4)	0,6(1,1)	0,8(0,9)	0,9(0,9)	0,7(0,6)	1,0(0,8)	1,0(0,7)	1,0(0,7)
<i>n</i>	19,1(9,6)	19,6(6,8)	18,2(4,0)	16,7(4,7)	15,7(4,3)	13,7(2,7)	15,5(4,6)	14,7(2,6)	15,0(2,9)
<i>n: prop⁺</i>	4,1(5,4)	3,4(4,6)	2,2(2,5)	0,9(1,4)	1,6(1,8)	0,7(0,7)	1,3(2,2)	1,7(1,9)	1,4(1,6)
<i>num</i>	0,0(0,0)	0,0(0,1)	0,1(0,2)	0,1(0,3)	0,1(0,5)	0,1(0,3)	0,1(0,3)	0,1(0,3)	0,2(0,3)
<i>prep*</i>	0,3(0,6)	1,9(2,1)	2,8(2,1)	3,7(1,5)	3,4(1,7)	3,3(1,3)	3,9(2,1)	4,2(1,3)	4,1(2,0)
<i>prep:art</i>	0,3(0,7)	0,4(0,5)	1,0(0,9)	1,0(1,2)	0,8(0,7)	0,9(0,9)	1,1(1,0)	0,7(0,6)	0,9(0,7)
<i>pro</i>	0,8(1,2)	1,3(1,5)	1,8(1,5)	1,6(1,4)	1,9(1,4)	1,8(0,9)	2,2(1,7)	1,8(1,2)	2,0(1,4)
<i>pro: dat</i>	0,0(0,0)	0,0(0,1)	0,1(0,3)	0,2(0,3)	0,3(0,4)	0,2(0,3)	0,3(0,4)	0,3(0,4)	0,3(0,4)
<i>pro: dem</i>	4,7(9,5)	3,2(3,3)	2,8(1,6)	3,1(2,0)	2,6(1,9)	2,5(2,2)	2,5(2,2)	1,9(1,3)	2,1(1,8)
<i>pro: int</i>	1,1(2,0)	1,3(2,1)	1,0(1,3)	1,2(1,4)	1,1(1,1)	1,0(1,0)	0,8(0,9)	1,1(1,3)	0,9(0,9)
<i>pro: obj</i>	0,1(0,3)	0,2(0,5)	0,2(0,3)	0,4(0,5)	0,6(0,6)	0,8(1,0)	0,4(0,4)	0,8(0,7)	0,5(0,5)
<i>pro: refl</i>	0,0(0,2)	0,5(1,0)	0,5(0,6)	0,6(0,6)	1,0(1,0)	1,0(1,0)	1,3(1,2)	1,0(0,8)	1,4(0,8)
<i>pro: rel</i>	0,0(0,2)	0,0(0,1)	0,1(0,2)	0,4(0,5)	0,4(0,4)	0,3(0,4)	0,3(0,3)	0,3(0,4)	0,4(0,4)
<i>pro: subj*</i>	2,2(3,1)	4,9(3,4)	6,7(3,3)	8,5(3,0)	9,9(3,8)	10,4(3,3)	10,3(3,7)	11,0(2,6)	11,0(2,3)
<i>pro: y</i>	0,0(0,1)	0,1(0,2)	0,0(0,1)	0,1(0,2)	0,2(0,4)	0,1(0,3)	0,4(1,0)	0,3(0,5)	0,3(0,4)
<i>v</i>	2,8(3,2)	4,8(2,7)	4,9(2,5)	4,5(2,3)	4,6(2,3)	4,8(2,3)	5,4(3,5)	4,0(1,6)	4,9(2,5)
<i>v: aux</i>	1,3(2,0)	1,8(2,3)	2,0(1,7)	2,3(1,7)	2,1(1,6)	2,0(1,4)	1,8(1,0)	1,8(1,3)	1,8(1,1)
<i>v: exist</i>	5,5(4,7)	4,7(4,1)	5,1(2,5)	6,6(1,9)	5,5(2,2)	5,4(2,0)	5,0(2,3)	5,4(1,8)	4,4(1,7)
<i>v: inf</i>	3,7(5,2)	4,3(4,2)	5,0(3,1)	4,2(1,8)	3,9(1,9)	5,1(2,0)	3,8(1,6)	4,5(2,0)	4,9(2,4)
<i>v: mdl</i>	1,0(1,7)	1,7(2,3)	2,5(2,1)	3,3(2,1)	3,9(2,0)	5,0(2,4)	3,9(2,0)	4,7(1,9)	4,7(1,9)
<i>v: mdlllex</i>	0,8(1,3)	1,3(1,6)	2,0(2,0)	2,6(2,3)	2,6(1,3)	2,5(0,9)	3,1(2,4)	2,8(1,5)	2,8(1,2)
<i>v: poss</i>	0,9(1,3)	0,5(0,8)	0,7(0,9)	0,6(0,6)	0,8(0,7)	0,8(0,8)	0,9(1,0)	1,0(0,7)	0,9(0,9)
<i>v: pp</i>	4,6(4,8)	3,7(3,6)	2,8(1,9)	2,4(2,0)	2,3(2,0)	2,3(1,8)	1,7(0,9)	1,7(1,2)	2,0(1,2)

Note : Les catégories accompagnées d'une * ont une corrélation avec l'âge (de 2 à 3 ans) supérieure à 0,50 et celles qui sont accompagnées d'un + ont une corrélation négative significative. Celles qui sont présentées en italique ont un coefficient de variabilité inférieur à 1. Les valeurs entre parenthèses représentent les écarts-types.

Le tableau 4 présente la moyenne du nombre d'occurrences pour l'ensemble des catégories syntaxiques (toutes les valeurs d'évolution, de corrélation et de variabilité sont présentées dans le tableau C de l'annexe). Le calcul en pourcentages amène des différences qualitatives importantes du point de vue de l'évolution et surtout de la corrélation des mesures avec l'âge. La variabilité des mesures reste par contre tout à fait proche de celle du calcul précédent (voir Annexe, tableau C). Les quelques valeurs qui sont modifiées le sont presque toujours dans le sens d'une plus faible variabilité.

Le calcul en pourcentages amène d'une façon générale une diminution de l'évolution des valeurs avec l'âge, ce qui est logique puisque les pourcentages tendent à écraser les valeurs. Par contre, on voit apparaître des évolutions négatives, ce qui n'existait pas avec les occurrences brutes. En effet, l'importance de certaines catégories diminue avec l'âge (voir les valeurs inférieures à 1 dans la colonne 3 du tableau C de l'Annexe), tout au moins en pourcentages. Cette tendance était suggérée par les faibles valeurs d'évolution relevées précédemment pour certaines catégories. Elle se confirme ici. Cette modification des valeurs entraîne aussi un changement important dans les valeurs de corrélation. En particulier des catégories comme *adv: place*, *co*, *co: act* et *n: prop* présentent des corrélations négatives significatives. Elles évoluent donc de manière inverse de l'âge. D'autres catégories comme le nom ou les catégories verbales de base (*v*, *v: aux*, *v: exist*, *v: inf* et *v: poss*) se retrouvent avec un évolution négative ou positive faible. Leur production en pourcentages est donc relativement stable avec l'âge.

L'intérêt majeur du calcul en pourcentages est qu'il fait ressortir de façon particulièrement nette l'importance de toutes les catégories qui correspondent à des marques morphosyntaxiques. Il s'agit essentiellement des catégories *det*, *det: poss*, *prep*, *prep: art*, *pro: refl*, *pro: rel*, *pro: subj*, *pro: y* auxquelles on peut ajouter les catégories *pro: dat*, *pro: obj*, et *pro: y* qui ont des caractéristiques proches ou une très forte valeur d'évolution. L'évaluation en pourcentages du nombre d'éléments dans ces catégories serait une bonne méthode pour apprécier le développement langagier de l'enfant. Deux autres catégories, *v: mdl* et *v: mdllex*, ont aussi une évolution nette du même type, avec une corrélation positive. On peut penser qu'il s'agit de marques verbales qui, bien que plus internes que les marques des pronoms préverbaux, présentent un développement important et tardif en raison de leur complexité.

NOMBRE DE TYPES

Tableau 5: Nombre moyen et écart-type par âge de mots différents produits par catégorie syntaxique (types)

AGE (mois)	24	27	30	33	36	39	42	45	48
<i>adj</i>	2 (2)	3 (3)	4 (3)	5 (3)	5 (4)	5 (3)	4 (3)	5 (3)	6 (4)
<i>adv</i>	1 (1)	2 (3)	3 (3)	5 (3)	6 (4)	6 (4)	8 (6)	8 (5)	9 (6)
<i>adv: int</i>	1 (1)	0 (1)	1 (1)	1 (1)	1 (2)	1 (1)	1 (1)	1 (1)	1 (1)
<i>adv: neg</i>	1 (1)	1 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)
<i>adv: place</i>	3 (1)	3 (1)	4 (2)	5 (3)	5 (2)	5 (2)	5 (2)	5 (2)	5 (3)
<i>adv: yn</i>	1 (1)	2 (1)	2 (1)	2 (1)	2 (1)	3 (1)	3 (2)	2 (1)	2 (1)
<i>co</i>	3 (3)	3 (3)	3 (2)	3 (2)	4 (3)	7 (7)	6 (6)	4 (4)	4 (3)
<i>co: act</i>	2 (1)	2 (1)	3 (1)	3 (1)	3 (2)	4 (2)	3 (2)	3 (2)	3 (2)
<i>conj</i>	0 (1)	1 (1)	1 (1)	3 (2)	4 (3)	4 (3)	5 (4)	5 (3)	6 (4)
<i>det*</i>	2 (2)	4 (2)	5 (2)	8 (2)	7 (1)	8 (2)	8 (2)	8 (2)	8 (2)
<i>det: dem</i>	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (1)	0 (1)	0 (1)	1 (1)
<i>det: poss</i>	0 (0)	0 (1)	1 (1)	2 (2)	2 (1)	2 (1)	2 (2)	2 (1)	2 (2)
<i>n*</i>	10 (7)	15 (9)	18 (8)	26 (10)	26 (10)	27 (10)	31 (15)	33 (14)	32 (15)
<i>n: prop</i>	1 (1)	2 (1)	2 (2)	2 (2)	3 (2)	2 (2)	3 (2)	3 (2)	3 (2)
<i>num</i>	0 (0)	0 (0)	0 (0)	0 (1)	0 (1)	1 (1)	0 (1)	0 (1)	1 (1)
<i>prep*</i>	0 (1)	2 (2)	3 (2)	4 (2)	5 (2)	5 (2)	5 (2)	5 (1)	6 (2)
<i>prep: art</i>	0 (1)	1 (1)	1 (1)	2 (1)	2 (1)	2 (1)	2 (2)	2 (1)	2 (2)
<i>pro</i>	1 (1)	1 (1)	2 (1)	3 (2)	3 (2)	4 (2)	4 (2)	4 (2)	4 (2)
<i>pro: dat</i>	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)	0 (1)	1 (1)	1 (1)	1 (1)
<i>pro: dem</i>	1 (1)	1 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)
<i>pro: int</i>	1 (1)	1 (1)	1 (1)	1 (1)	2 (1)	2 (1)	1 (1)	2 (1)	2 (1)
<i>pro: obj</i>	0 (0)	0 (1)	0 (1)	1 (1)	1 (1)	2 (1)	1 (1)	2 (1)	1 (1)
<i>pro: refl</i>	0 (0)	0 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	2 (1)	2 (1)
<i>pro: rel</i>	0 (0)	0 (0)	0 (0)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
<i>pro: subj*</i>	1 (1)	3 (2)	4 (2)	6 (2)	6 (2)	6 (2)	6 (2)	6 (2)	6 (2)
<i>pro: y</i>	0 (0)	0 (0)	0 (0)	0 (1)	0 (1)	0 (1)	1 (1)	1 (1)	1 (1)
<i>v*</i>	2 (2)	5 (3)	6 (4)	8 (5)	10 (4)	11 (6)	11 (7)	11 (6)	13 (7)
<i>v: aux</i>	1 (1)	1 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)
<i>v: exist*</i>	2 (1)	2 (1)	3 (1)	3 (1)	3 (1)	4 (1)	3 (1)	3 (1)	4 (1)
<i>v: inf*</i>	2 (3)	4 (4)	6 (4)	9 (4)	9 (4)	11 (5)	10 (6)	10 (6)	12 (6)
<i>v: mdl</i>	1 (1)	1 (2)	2 (2)	3 (2)	3 (1)	4 (2)	4 (2)	4 (2)	4 (2)
<i>v: mdllex*</i>	1 (1)	1 (2)	2 (2)	3 (2)	4 (2)	4 (2)	4 (2)	4 (2)	5 (2)
<i>v: poss</i>	0 (1)	0 (1)	1 (1)	1 (1)	1 (0)	1 (1)	1 (1)	1 (1)	1 (1)
<i>v: pp</i>	2 (2)	3 (2)	3 (3)	4 (3)	5 (3)	5 (3)	5 (3)	6 (4)	5 (3)

Note: Les catégories accompagnées d'une * ont une corrélation avec l'âge (de 2 à 3 ans) supérieure à 0,50 et celles qui sont présentées en italique ont un coefficient de variabilité inférieur à 1. Les valeurs entre parenthèses représentent les écarts-types.

Le tableau 5 présente la moyenne du nombre de types pour l'ensemble des catégories syntaxiques (toutes les valeurs d'évolution, de corrélation et de variabilité sont présentées dans le tableau C de l'annexe). Le calcul en types présente des valeurs d'évolution, de corrélation et de variabilité qui sont proches de celles du calcul brut en occurrences (on trouve un léger avantage aux types pour la variabilité). Cette absence de différence tend à indiquer qu'il n'y a pas à avoir de préférence nette pour l'usage des occurrences ou des types pour l'évaluation du langage. On préfère souvent les calculs en types parce qu'ils donnent plus facilement des indications qualitatives, à condition de regarder quels mots sont effectivement produits et de les comparer avec la liste des mots que produisent les enfants de référence (voir Annexe, tableaux B1 et B2). C'est particulièrement vrai pour de catégories comme *adv* ou *conj* qui se diversifient considérablement dans la quatrième année de l'enfant. Ce sera encore plus vrai si l'on s'intéresse à des marques syntaxiques plus rares comme les marques du pluriel, du féminin, du futur simple, de l'imparfait qui sont rares en occurrences brutes mais apparaissent dans le vocabulaire des enfants après l'âge de trois ans. On touche là aux limites de l'analyse quantitative brute. Par contre, il semble que le calcul en types ne puisse pas remplacer le calcul des occurrences en pourcentages, qui se révèle complémentaire.

APPLICATION CLINIQUE DE LA MÉTHODE

L'utilisation de recueil de langage spontané nécessite un travail important de la part des orthophonistes, travail qui n'est pas toujours compatible avec la durée normale d'une consultation. Toutefois dans les cas difficiles, notamment lorsque l'enfant a un retard qui ne permet pas l'utilisation de tests standardisés, cette méthode est très recommandée. C'est ainsi que Parker et Brorson* soulignent l'intérêt de telles analyses d'échantillons de langage spontané en clinique.

A travers l'analyse d'échantillon de langage (AEL), « des données spécifiques peuvent être obtenues en testant le comportement de l'enfant dans une conversation »*. Avec ces données spécifiques, les logopèdes peuvent plus facilement décrire les troubles de langage et les retards, sélectionner des buts et planifier les traitements de manière adaptée*. L'AEL tend à être plus utile que les tests standardisés pour déterminer des buts de rééducation spécifiques*. L'échantillon de langage est intéressant car il permet un discours plus naturel, alors que les tests standardisés demandent en général un usage artificiel dans lequel des réponses structurées sont attendues*.

Les résultats de telles analyses de langage présentées dans cet article ont déjà été appliqués pour des enfants francophones avec des objectifs de recherche. D'abord, on a pu constater des anomalies de production sur certaines catégories. Par exemple, certains dysphasiques produisent peu de verbes*, ou certains enfants avec une tendance autistique produisent moins de *co*, *co: act*, *pro: dem* et *adv: place***.

Ensuite, on a pu parfois constater une absence de diversité sur certaines catégories, c'est-à-dire un faible nombre de types avec un nombre d'occurrences normal. Enfin, on a constaté souvent chez les dysphasiques des faiblesses en pourcentages d'occurrences pour des catégories fonctionnelles liées au verbe ou au nom. Chez les dysphasiques plus âgés, on pourra s'intéresser directement à la production de certaines catégories comme par exemple le pronom personnel objet (*pro: obj*) qui est connu pour poser problème dans les langues romanes*.

De même, chez les enfants sourds profonds qui développent leur langage avec les implants cochléaires, l'évaluation a permis de mettre à jour leurs grandes difficultés dans la mise en place des catégories syntaxiques à faible saillance perceptive comme les *det*, les *pro: subj* et les *prep* que nous avons regroupées en mots grammaticaux*.

La méthode a donc été déjà utilisée en clinique. Un étalonnage précis avec regroupement de variables lexicales et grammaticales doit faire l'objet d'une monographie complète en langue française. Cette base de données du petit enfant doit être transmise à la base francophone Lexique*. Cela fournira au clinicien la possibilité de consulter directement la base et de tracer un profil d'évolution en différentes notes standard situant l'en-

*2005

*Owens, 1999, p. 120)

*Hughes, Fey et Long, 1992

*Paul, 2000

*Butler, 1992

*Le Normand et Chevie-Muller, 1989 ; Parisse et Le Normand, 2002 ; Le Normand, 2006

**Parisse, 1999

*Jakubowicz, Nash, Rigaut, Gérard, 1998 ; Jakubowicz, 2003 ; Paradis, Crago, Genesee, 2003

*Le Normand, 2004, 2005

*New, Pallier, Brysbaert, Ferrand, 2004

fant aux extrêmes des notes, en deçà ou au-delà de la zone moyenne ou médiane. Les tendances développementales de la diversité lexicale et grammaticale avec un contrôle minutieux de différents facteurs liés à la méthode seront également investiguées. C'est dans cette perspective qu'une étude contrôlant le facteur sexe et milieu socio-culturel de l'enfant a été réalisée*.

*Le Normand, Parisse et Cohen, à paraître

CONCLUSION

Les résultats présentés ci-dessus ont une valeur clinique et expérimentale. Ils permettent de donner une indication qualitative et quantitative suffisante du niveau de développement du langage d'un enfant par rapport à des enfants âgés de 2 à 4 ans. Après le recueil de corpus selon le contexte de la méthode de la maison « Fisher-Price », il faut transcrire le corpus puis les outils comme CLAN de CHILDES permettent d'évaluer rapidement le corpus obtenu.

Une première analyse est celle du nombre d'énoncés, de mots, et de la LME. Elle permet d'estimer un niveau global de langage et de mesurer un retard éventuel (voir tableau 1). Cette analyse peut être suivie d'une analyse morphosyntaxique du corpus qui fait ressortir les catégories syntaxiques des mots. On peut alors faire une analyse en occurrences brutes (voir tableau 3), en pourcentages d'occurrences (voir tableau 4) ou en types (voir tableau 5). Ces analyses présentent des caractéristiques variées et aucune n'est meilleure que les autres. Au contraire, elles se complètent les unes les autres. Il est donc intéressant de toutes les examiner lorsque l'on cherche à évaluer le langage d'un enfant, d'autant plus que les outils existants permettent de fournir les trois types de résultats assez simplement une fois que l'on dispose d'une transcription du langage d'un enfant.

Il convient de terminer la présentation de cette méthode d'évaluation en rappelant les limites théoriques. Les résultats présentés ci-dessus ne doivent pas être interprétés comme un modèle du développement du langage chez l'enfant. L'analyse réalisée doit être interprétée comme une simple description externe. Elle n'a que pour but de répondre à la question : comment comparer le langage produit par un enfant avec celui de ses pairs ? Attribuer une catégorie syntaxique aux mots produits par l'enfant ne prouve pas qu'il maîtrise cette catégorie. Ses connaissances peuvent être lexicales ou partielles*. Mais pour l'évaluation du langage, le procédé de description externe suffit.

*Tomasello, 2000

Ceci n'interdit pas par ailleurs d'utiliser les résultats obtenus pour développer des hypothèses théoriques, à condition de les interpréter avec précaution. Par exemple, l'analyse des occurrences brutes présentées ci-dessus a permis de proposer une typologie de base des catégories syntaxiques utilisées par l'enfant. On a séparé les catégories en plusieurs grands groupes :

- catégories invariables à vocabulaire limité et grande fréquence, de grande valeur sémantique dans les interactions, déjà bien acquises à l'âge de deux ans ;
- catégories lexicales de base, pouvant présenter des marques syntaxiques, avec un vocabulaire ouvert, fréquentes, ayant une valeur sémantique importante, déjà bien acquises à l'âge de deux ans ; catégories de marqueurs morphosyntaxiques « internes », du nom ou du verbe, développement soutenu durant la troisième année ;
- catégories de marqueurs morphosyntaxiques « externes », du nom ou du verbe, développement soutenu à partir de la troisième année et jusque durant la quatrième année ;
- catégories de marqueurs morphosyntaxiques complexes ou de mots-liens entre énoncés, plus rares et d'usage plus variable, dont l'acquisition ne commence vraiment qu'à partir de la quatrième année. Cette typologie reflète le développement des catégories syntaxiques présentées dans l'étude ci-dessus et prend en compte les propriétés segmentales, phonologiques, fréquentielles, sémantiques et morphosyntaxiques de la langue. Comme les groupes proposés ont des propriétés différentes, on peut faire des hypothèses sur les mécanismes sous-jacents et s'en servir pour mieux comprendre le développement du langage et de ses pathologies. A ce moment, on pourra peut-être pro-

poser d'autres caractérisations du langage de l'enfant et améliorer les tests de langage. En attendant ce moment, on peut se servir des techniques de description actuelles, même si elles ne reflètent pas *exactement* les processus langagiers, pour évaluer le langage des enfants.

BIBLIOGRAPHIE

- BASSANO, D. (2005). Le DPLF: un nouvel outil pour l'évaluation du développement du langage de production en Français. *Enfance*, 2005, 171-208.
- BROWN, R.W. (1973). *A first language: the early stages*. Cambridge, Mass. : Harvard University Press.
- BUTLER, K. (1992). From the Editor. *Topics in Language Disorder*, 12, IV-V.
- CHEVRIE-MULLER, C., SIMON, A.M., LE NORMAND, M.T., FOURNIER, S. (1997). *Batterie d'évaluation psycholinguistique pour enfants de 2 ans 9 mois à 4 ans 3 mois* (BEPL). Paris : Editions du Centre de Psychologie Appliquée.
- GRÉGOIRE, M.N., RONDAL, J.A., PÉRIÉ, F. (1984). Aspects morpho-syntaxiques et lexicaux du langage spontané d'enfants de 2 à 5 ans. *Enfance*, 1, 51-65.
- HUGUES, D.L., FEY, M.E., LONG, S.H. (1992). Developmental sentence scoring : still useful after all these years. *Topics in Language Disorder*, 12, 1-12.
- JAKUBOWICZ, C. (2003). Hypothèses psycholinguistiques sur la nature du déficit dysphasique. In GÉRARD, C.L., BRUN, V. (eds.), *Les dysphasies*. Paris : Masson.
- JAKUBOWICZ, C., NASH, L., RIGAUT, C., GÉRARD, C.L. (1998). Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition*, 7, 113-60.
- KLEE, T., FITZGERALD, M.D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, 12, 251-69.
- LE NORMAND, M.T. (1986). A developmental exploration of language used to accompany symbolic play in young, normal children (2 - 4 years old). *Child : Care, Health and Development*, 12, 121-34.
- LE NORMAND, M.T. (1991). La démarche de l'évaluation psycholinguistique chez l'enfant de moins de 3 ans. *Glossa*, 26, 14-21.
- LE NORMAND, M.T. (1997). Early morphological development in French children. In OLOFSSON, A.S., STRÖMQVIST, S. (eds.), *Learning disorders as a barrier to human development : Cross-linguistic studies of dyslexia and early language development*. Luxembourg : Office for Official Publication of the European Communities.
- LE NORMAND, M.T. (1999). Retards de langage et dysphasies. In RONDAL, J.A., SERON, X. (eds.), *Troubles du langage*. Sprimont : Mardaga.
- LE NORMAND, M.T. (2004). Le développement perturbé du langage : retard ou trouble spécifique ? *La nouvelle revue de l'AIS (Adaptation et Intégration scolaire)*, 27, 67-76.
- LE NORMAND, M.T. (2005). Production du lexique chez des enfants sourds profonds munis d'implants cochléaires sur un suivi de quatre ans. *Handicap, revue de sciences humaines et sociales*, 105-106, 21-31.
- LE NORMAND, M.T. (2006). *Premiers mots et émergences des catégories syntaxiques chez l'enfant*. Société de Neurophysiologie Clinique de Langue Française, Neurophysiologie du langage, 31-46. Paris : Elsevier, SAS.
- LE NORMAND, M.T., CHEVRIE-MULLER, C. (1989). Exploration du lexique chez six enfants dysphasiques. *Rééducation orthophonique*, 27, 345-61.
- LE NORMAND, M.T., CHEVRIE-MULLER, C. (1991). Individual differences in the production of word classes in eight specific language-impaired preschoolers. *Journal of Communication Disorders*, 24, 331-51.
- LE NORMAND, M.T., PARISSÉ, C., COHEN, H. (à paraître). Lexical diversity and productivity in French preschoolers : gender and biosocial aspects. *Journal of Multilingual Communication Disorders*.
- MAC WHINNEY, B. (2000). *The CHILDES project : Tools for analyzing talk* (3rd). Hillsdale, N.J : Lawrence Erlbaum.
- MAC WHINNEY, B., SNOW, C.E. (1985). The child language data exchange system. *Journal of Child Language*, 12, 271-96.
- MAILLART, C., PARISSÉ, C. (in press). Phonological deficits in French speaking children with SLI. *International Journal of Language and Communication Disorders*.
- MELINE, T.J., MELINE, N.C. (1981). Normal variation and prediction of mean length of utterance from chronological age. *Perceptual and Motor Skills*, 53, 376-8.
- MILLER, J.F. (1981). *Assessing language production in children : Experimental procedures*. London, UK : Edward Arnold.
- MILLER, J.F., CHAPMAN, R.S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24, 154-61.
- MILLER, J.F., CHAPMAN, R.S. (1982). *SALT : Semantic Analysis of Language Transcripts*. Wisconsin : Language Analysis Laboratory, Waisman Centre on Mental Retardation and Human Development, University of Madison.
- NEW, B., PALLIER, C., BRYLSBAERT, M., FERRAND, L. (2004). Lexique 2 : A New French Lexical Database. *Behavior Research Methods, Instruments, & Computers*, 36, 516-24.

- OWENS, R.E. (1999). *Language disorders : A functional approach to assessment and intervention* (3rd ed.). Needham Heights, MA : Allyn & Bacon.
- PARADIS, J., CRAGO, M., GENESEE, F. (2003). *Object clitics as clinical marker of SLI in French : evidence from French-English bilingual children*. Paper presented at the 27th Annual Boston University Conference on Language Development.
- PARISSÉ, C. (1999). Cognition and language acquisition in normal and autistic children. *Journal of Neurolinguistics*, 12, 247-69.
- PARISSÉ, C., LE NORMAND, M.T. (1998). Traitement automatique de la morphosyntaxe chez le petit enfant. *Glossa*, 61, 22-9.
- PARISSÉ, C., LE NORMAND, M.T. (2000a). Automatic disambiguation of morphosyntax in spoken language corpora. *Behavior Research Methods, Instruments, & Computers*, 32, 468-81.
- PARISSÉ, C., LE NORMAND, M.T. (2000b). How children build their morphosyntax : The case of French. *Journal of Child Language*, 27, 267-92.
- PARISSÉ, C., LE NORMAND, M.T. (2002). Production of lexical categories in French children with SLI and in normally-developing children matched for MLU. *Brain and Cognition*, 48, 490-4.
- PARISSÉ, C., MAILLART, C. (2004). Les déficits phonologiques des enfants francophones ayant des troubles spécifiques de développement du langage. *Glossa*, 89, 34-47.
- PARKER, M.D., BRORSON, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, 25, 365-76.
- PAUL, R. (2000). *Language disorders from infancy through adolescents* (2nd ed.). Saint Louis, MO : Mosby-Year Book.
- ROLLINS, P.R., SNOW, C.E., WILLETT, J.B. (1996). Predictors of MLU: Semantic and morphological developments. *First Language*, 16, 243-59.
- RONDAL, J.A. (1997). *L'évaluation du langage*. Hayen : Mardaga.
- RONDAL, J.A., GHIOTTO, M., BREDART, S., BACHELET, J.F. (1987). Age-relation, reliability, and grammatical validity of measures of utterance length. *Journal of Child Language*, 14, 433-46.
- ROSE, Y. (2003). *ChildPhon : A database solution for the study of child phonology*. *Proceedings of the 26th Annual Boston University Conference on Language Development*. Somerville, MA : Cascadilla Press.
- TOMASELLO, M. (2000). Do young children have adult syntactic competence ? *Cognition*, 74, 209-53.

ANNEXE

adj	adjectif
adv	adverbe
adv: int	adverbe interrogatif
adv: neg	adverbe de négation
adv: place	adverbe de lieu
adv: yn	adverbe oui/non
co	communicateur ou interjection
co: act	communicateur d'action (oh ; ah ; ça + là ; ...)
conj	conjonction
det	déterminant (articles définis et indéfinis)
det: dem	déterminant démonstratif
det: poss	déterminant possessif
n	nom
n: prop	nom propre
num	numéro ou nombre (supérieur à 1)
prep	préposition
prep: art	préposition-article
pro	pronom (général)
pro: dat	pronom personnel datif (objet indirect)
pro: dem	pronom démonstratif
pro: int	pronom interrogatif
pro: obj	pronom personnel objet direct
pro: refl	pronom réfléchi
pro: rel	pronom relatif
pro: subj	pronom personnel sujet
pro: y	pronoms y, en
v	verbe (conjugué)
v: aux	verbe auxiliaire
v: exist	verbe d'existence (être et il + y + a)
v: inf	verbe infinitif
v: mdl	verbe modal (je fais cuire un gâteau)
v: mdllex	verbe modal lexical (je fais un gâteau)
v: poss	verbe d'appartenance (verbe avoir lexical)
v: pp	verbe participe passé

Tableau A : Catégories morphosyntaxiques pouvant être analysées

adj	24-beau 27-autre 30-deux 30-petit 30-petite 33-grand 45-grande
adv	24-encore 30-aussi 30-comme+ça 33-bien 33-maintenant 36-alors 42-beaucoup 42-si 42-tout+seul 45-déjà 45-même 48-peut-être 48-trop
adv: int	24-où+il+est 39-pourquoi 42-comment
adv: neg	24-pas 24-plus 33-pu 36-n' 45-ne
adv: place	24-dedans 24-là 24-voilà 30-ici 30-à+côté 33-après 33-dehors
adv: yn	24-non 24-oui 39-ouais 45-ah+oui
co	24-au+revoir 24-boum 24-coucou 24-hein 30-bonjour 33-ben 33-hop 39-ben+non 39-ben+oui 48-et+après
co: act	24-ah 24-oh 24-ça+y+est 33-t'+as+vu 36-et+voilà 36-ça+va
conj	24-et 33-aussi 33-parce+que 36-et+puis 36-mais 36-parce+qu' 39-alors 39-si 42- qu' 42-quand 42-que 48-comme
det	24-l' 24-la 24-le 24-un 27-des 27-les 27-une 30-de 33-du 33-tous
det: dem	42-ce 45-cette
det: poss	33-ma 33-mon 33-sa 33-son 48-ta
n	24-bonhomme 24-bébé 24-chaise 24-cheval 24-chien 24-dodo 24-lit 24-poussette 24-table 24-voiture 27-dame 27-garage 27-maison 27-porte 33-maman 36-papa 39- place 42-escalier 45-voitures
n: prop	24-bébé 24-maman 24-papa
num	39-deux
prep	27-dans 27-pour 27-à 30-de 33-avec 33-sur 42-chez
prep: art	30-au 30-du 30-à 33-de 42-en
pro	24-autre 27-moi 33-lui 33-tout 36-toi 39-quelqu'+un 48-tout+le+monde
pro: dat	33-lui
pro: dem	24-ça 27-celui+là 33-celle+là
pro: int	27-où 30-quoi 33-qui
pro: obj	33-le 36-l' 36-la 36-les
pro: refl	30-se 33-s' 39-m' 42-me 48-te
pro: rel	33-qui
pro: subj	24-i 24-l' 27-elle 27-il 27-on 30-j' 30-je 30-tu 33-t' 36-ils
pro: y	36-en
v	24-regarde 27-attends 27-ferme 27-met 30-dort 30-mange 33-rentre 33-tombe 36- mets 36-monte 36-vois 39-range 48-descend 48-ouvre 48-prend
v: aux	24-est 27-a 30-c'+est 33-ai 33-sont 42-ont
v: exist	24-c'+est 24-est 24-y'+a 33-sont 33-y'+en+a
v: inf	24-manger 24-tomber 27-asseoir 27-fermer 27-mettre 27-monter 30-dormir 33- coucher 33-promener 33-rentre 36-chercher 36-prendre 36-voir 39-ouvrir 39- sortir 42-partir 42-ranger 45-jouer
v: mdl	24-va 30-peut 33-faut 33-vaïs 33-vont 36-sais 36-veut 39-aller 39-peux 39-veux 45-vas
v: mdllex	27-fait 27-va 30-aller 30-faire 36-allez 39-veut 42-fais 42-vont 45-sais 48-faut
v: poss	24-a 39-ai 42-ont
v: pp	24-assis 24-caché 24-fermé 24-parti 24-tombé 36-fini 36-mangé 39-mis 48-cassé 48-dit

Tableau B1 : Age de production (en mois) des mots attestés chez au moins 25 % des enfants

adj	33-deux 33-petit
adv	33-encore 36-aussi 39-comme+ça 42-maintenant
adv: neg	24-pas 39-plus
adv: place	24-là 24-voilà 39-après 39-dedans 39-ici
adv: yn	24-non 24-oui
co	39-hein
co: act	24-ah 24-oh 39-ça+y+est
conj	30-et 39-parce+que
det	27-l' 27-la 27-le 30-des 30-les 30-un 33-de 33-une
det: poss	36-sa
n	24-dodo 24-voiture 27-poussette 27-table 30-bonhomme 30-chaise 30-chien 30-maison 30-porte 33-garage 33-lit 33-maman 39-bébé 39-place 42-escalier 42-papa 45-voitures
n: prop	24-bébé 27-maman
prep	30-dans 30-pour 30-à 36-avec 36-de
prep: art	33-du 39-au
pro	33-moi 36-lui
pro: dat	36-lui
pro: dem	24-ça
pro: int	33-où 36-qui 39-quoi
pro: obj	39-le
pro: refl	33-se 36-s'
pro: rel	36-qui
pro: subj	27-i 27-il 30-elle 30-l' 33-je 33-on 36-j' 36-tu
pro: y	48-en
v	27-regarde 36-dort 36-met 39-attends 39-ferme
v: aux	27-est 30-a 30-c'+est 45-ai
v: exist	24-c'+est 24-est 30-y'+a 48-y'+en+a
v: inf	33-manger 33-mettre 36-dormir 39-fermer
v: mdl	30-va 36-peut 39-vais 42-veux 42-vont 48-faut
v: mdlex	30-fait 33-faire 33-va 36-aller
v: poss	30-a 45-ai
v: pp	30-tombé

Tableau B2 : Age de production (en mois) des mots attestés chez au moins 50 % des enfants

	Evolution			Corrélation			Variabilité		
	Occ.	Occ.%	Types	Occ.	Occ.%	Types	Occ.	Occ.%	Types
adj	2,0	0,8	1,8	0,39	0,05	0,44	0,9	0,7	0,8
adv	4,2	1,5	4,3	0,35	-0,06	0,48	1,0	0,7	0,8
Adv: int	1,7	0,6	2,2	0,17	-0,13	0,13	1,5	1,4	1,1
adv: neg	2,2	0,9	1,5	0,36	0,01	0,26	1,0	0,7	0,5
adv: place	1,4	0,4	1,5	0,27	-0,39	0,43	0,7	0,6	0,4
adv: yn	2,0	0,5	1,6	0,30	-0,17	0,31	1,1	1,0	0,6
co	1,7	0,3	1,4	-0,05	-0,40	-0,08	1,1	1,0	0,8
co: act	1,2	0,4	1,3	0,02	-0,34	0,25	1,0	0,9	0,6
conj	10,2	3,9	6,8	0,40	0,20	0,48	1,4	1,2	0,9
det	3,2	1,5	2,0	0,60	0,52	0,70	0,8	0,4	0,4
det: dem	11,0	5,9	9,9	0,07	0,19	0,07	4,0	4,2	4,0
det: poss	6,1	3,8	5,4	0,38	0,37	0,49	1,6	1,5	1,3
n	2,1	0,8	2,2	0,52	-0,13	0,57	0,6	0,3	0,5
n: prop	1,2	0,4	1,5	-0,08	-0,31	0,12	1,2	1,3	0,9
num	5,9	3,8	5,9	0,26	0,24	0,27	2,5	2,6	2,4
prep	4,5	2,5	3,3	0,61	0,61	0,69	1,0	0,7	0,7
prep: art	3,1	1,6	2,7	0,39	0,32	0,44	1,3	1,1	1,0
pro	3,5	1,5	3,0	0,42	0,22	0,48	1,0	0,9	0,7
pro: dat	12,5	7,2	9,6	0,30	0,29	0,34	1,9	1,9	1,7
pro: dem	1,8	0,6	1,3	0,41	-0,11	0,37	0,9	0,9	0,6
pro: int	2,2	0,8	2,2	0,30	-0,01	0,31	1,3	1,2	0,8
pro: obj	6,6	3,5	4,8	0,31	0,27	0,38	1,7	1,5	1,3
pro: refl	6,6	3,8	4,3	0,35	0,31	0,45	1,6	1,5	1,3
pro: rel	16,6	7,0	8,9	0,38	0,37	0,36	2,1	2,0	1,9
pro: subj	4,9	2,4	2,4	0,56	0,60	0,66	0,9	0,5	0,5
pro: y	11,8	7,9	7,8	0,19	0,19	0,22	2,6	2,7	2,3
v	2,7	1,1	2,8	0,47	0,22	0,53	0,9	0,6	0,6
v: aux	2,3	1,1	1,8	0,41	0,17	0,46	0,9	0,9	0,6
v: exist	2,6	1,0	1,7	0,51	0,10	0,53	0,8	0,5	0,4
v: inf	2,5	1,0	2,7	0,47	0,06	0,54	0,8	0,6	0,7
v: mdl	5,4	2,5	2,9	0,55	0,39	0,49	1,0	0,8	0,8
v: mdllex	4,7	2,1	3,2	0,54	0,36	0,58	0,9	0,8	0,7
v: poss	2,9	1,3	2,0	0,28	-0,08	0,23	1,1	1,1	0,7
v: pp	1,2	0,5	2,0	0,10	-0,26	0,33	0,8	0,8	0,7

Tableau C : Valeurs d'évolution, de corrélation et de variabilité avec l'âge pour chaque catégorie

Notes :

- (A) l'évolution est calculée comme le rapport entre le nombre d'éléments produits aux âges de 2 ; 0, 2 ; 3 et 2 ; 6 et le nombre d'éléments produits aux âges de 3 ; 6, 3 ; 9 et 4 ; 0. Les valeurs d'évolution moyennes sont en gras, les valeurs d'évolution fortes sont présentées sur fond grisé,
- (B) les valeurs de corrélation sont calculées pour les âges de 2 ; 0 à 2 ; 9 compris (pas de corrélation significative après l'âge de 3 ans). Les valeurs en gras sont positives et significatives. Les valeurs en gras et italiques sont positives, significatives et supérieures à 0,50. Les valeurs sur fond grisé sont négatives et significatives,
- (C) la variabilité est la moyenne des variabilités calculées séparément pour chacune des 9 tranches d'âge. Les variabilités faibles (inférieures à 1) sont en gras, les variabilités fortes (supérieures à 2) sont présentées sur fond grisé.