



HAL
open science

Etudier la complexité syntaxique chez l'enfant de moins de six ans dans une perspective interactionnelle : choix d'une méthodologie qualitative

Emmanuelle Canut, Martine Vertalier

► To cite this version:

Emmanuelle Canut, Martine Vertalier. Etudier la complexité syntaxique chez l'enfant de moins de six ans dans une perspective interactionnelle : choix d'une méthodologie qualitative. Josie Bernicot, Edy Veneziano, Michel Musiol, .. [et al.]. Interactions verbales et acquisition du langage, L'Harmattan, pp.239-260, 2010. hal-00523554

HAL Id: hal-00523554

<https://hal.science/hal-00523554>

Submitted on 9 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etudier la complexité syntaxique chez l'enfant de moins de six ans dans une perspective interactionnelle : choix d'une méthodologie qualitative

Emmanuelle Canut

ATILF-CNRS, Nancy Université (France)

Martine Vertalier

Université de la Sorbonne Nouvelle – Paris 3 (France)

1. Des visées théoriques aux principes méthodologiques des recherches sur l'acquisition de la syntaxe

1.1. La psycholinguistique développementale

Les recherches actuelles en acquisition du langage se caractérisent par leur grande hétérogénéité. La majorité d'entre elles s'inscrivent dans le champ de la psycholinguistique développementale dont les orientations théoriques oscillent entre des conceptions strictement cognitivistes, dans la continuité de la grammaire générative (Radford, 1995 pour une synthèse), et des conceptions connectionnistes, plus fonctionnalistes (*functional-cognitive approach*), avec la prise en compte des contraintes de la communication (Bates et MacWhinney, 1989 ; Tomasello, 2000, 2003)¹. Conformément aux visées épistémologiques des sciences cognitives, les psycholinguistes cherchent à décrire des mécanismes généraux de traitement de l'information et à déterminer des étapes de développement (ou stades d'acquisition c'est-à-dire ce qui est acquis à un moment donné) en fonction des capacités cognitives de l'enfant ou du type d'*input* (contraintes linguistiques). Leur travail s'appuie principalement sur les méthodologies expérimentales de la psychologie cognitive :

- Réalisations de tâches variées selon le domaine d'étude et les âges ciblés : rappel de syllabes, de phrases, de textes (en production orale ou écrite), mimes d'actions ou descriptions d'images, tests...

- Analyses automatisées ou semi-automatisées des productions : expériences de succion non nutritive en fonction de la stimulation

¹ Voir Lust (2006) pour une synthèse.

auditive, utilisation de l'orientation préférentielle ou de la fixation du regard (compatibilité entre durée du regard sur l'image et compréhension de la phrase test), procédure de rotation ou de mouvement préférentiel de la tête, etc.

Néanmoins, de plus en plus de chercheurs s'appuient sur des corpus de données « authentiques », « spontanées », pouvant être combinées à des rapports parentaux (comme le CDI ; *MacArthur Communicative Development Inventories* ou le DLPF : *Développement du Langage de Production en Français*). Quelle que soit la procédure choisie, les psycholinguistes procèdent majoritairement à des études transversales (comparaison de grands nombres d'enfants d'âges différents), lesquelles peuvent être toutefois comparées à des études de corpus longitudinaux.

Les études expérimentales mettent en évidence que les enfants ont de grandes difficultés à comprendre les phrases complexes. Or les études faites à partir de la production de langage « spontané » montrent qu'ils en produisent dès l'âge de deux ans. Malgré des disparités dans les résultats, les études sur l'émergence grammaticale chez les enfants de moins de trois ans ont dégagé un ordre général d'apparition : pour l'anglais, les structures coordonnées et subordonnées émergent entre deux et quatre ans, sans que certaines formes ne soient acquises avant l'âge de quatre ou cinq ans. Dans un premier temps, apparaissent les constructions infinitives avec *to* et les coordinations avec *and*. Dans un deuxième temps, entre 27 et 36 mois, apparaissent les constructions relatives, causales et temporelles (*when, while, because*) les complétives avec *that* (Bloom, 1991 ; Diessel, 2004, Peterson & McCabe, 1987). L'ordre d'apparition est cependant très variable selon les enfants, de même que le degré de maîtrise de ces complexités. Pour le français, Clark (2003) indique des similitudes avec l'anglais : apparition des constructions infinitives avec leurs différentes combinaisons centrées autour des verbes (verbe suivi de l'infinitif, préposition *à* suivie de l'infinitif), puis vers deux ans et demi, les enfants commencent à employer *que* pour introduire les relatives et les complétives, le temporel *quand*, ou encore le conditionnel *si*, et *parce que* pour marquer les relations causales.

Chez l'enfant de plus de trois ans, l'étude des constructions syntaxiques complexes (grosso modo la subordination) est essentiellement expérimentale : il existe une longue tradition de

recherche sur la compréhension et la production des phrases passives et le traitement de la compréhension des « phrases » relatives (pour le français voir Amy, 1983 ; Kail, 2000 pour une revue). Ces études transversales portent généralement sur un seul phénomène. Plusieurs auteurs ont souligné la rareté des études portant sur la production des constructions complexes dans le langage spontané : l'étude de Diessel (2004) sur les complexités syntaxiques du langage d'enfants de langue anglaise est considérée comme la plus importante avec 12000 énoncés référencés. Pour l'espagnol le constat est le même : Aparici et collègues (2001) indiquent le peu d'études à disposition sur la production « spontanée » des complexités syntaxiques au-delà de trois ans, en raison, entre autres, du manque de corpus disponibles.

Bien que les procédures méthodologiques quantitatives soient en adéquation avec les objectifs de recherche (décrire le langage de l'enfant « moyen »), la psycholinguistique développementale fait l'objet de critiques. D'une part sur le plan théorique : seul le processus de maturation d'un fonctionnement interne semble expliquer le passage d'un stade de développement à un autre ; d'autre part sur le plan méthodologique : « les données expérimentales sont-elles, comme on l'écrit parfois, fragiles et systématiquement sur-interprétées ? » (Lécuyer, 2002, p. 21). En effet, la sélection préalable des données écarte le langage adressé à l'enfant (il peut être recueilli mais il n'est pas pris en compte dans l'explication des processus d'acquisition du langage) et il paraît alors difficile de généraliser à l'usage naturel du langage des résultats obtenus en laboratoire (hors de toute fonction communicative) et ne prenant pas en compte la dimension interactionnelle et donc sociale du langage.

1.2. Les approches socio-interactionnistes

D'autres études s'inscrivent dans le champ d'une psycholinguistique que l'on pourrait qualifier de socio-interactionniste. Elles s'intéressent davantage au développement de l'enfant d'un point de vue socio-historico-culturel : « toute activité mentale est culturellement située, c'est-à-dire qu'elle n'est jamais isolée ou menée sans assistance » (Bruner, 1991). La langue n'est pas un système homogène et stable, une entité qui préexisterait indépendamment des individus ; la langue se construit au sein des

pratiques langagières. Dans cette approche, les interventions de l'entourage jouent un rôle décisif dans le développement des capacités cognitivo-langagières de l'enfant : l'évolution du langage de l'enfant est liée à la fréquence de l'*input* (c'est-à-dire des données que l'enfant reçoit des adultes qui lui parlent). On entrevoit toutefois deux façons d'aborder le problème : soit en procédant à des études quantitatives de corrélation entre les verbalisations de l'enfant et celles de l'adulte, soit en essayant de déterminer qualitativement les modalités d'interaction favorables à un apprentissage du langage.

Un état des lieux des études quantitatives sur l'*input*, montre que la quasi-totalité des recherches porte sur des enfants de moins de 3 ans (Gallaway et Richards, 1994 ; Snow, 1995). Par ailleurs, bien que le langage adressé à l'enfant prenne sens dans le contexte conversationnel, ces travaux n'en tiennent pas compte, ou plus souvent procèdent à une analyse disjointe du langage de l'adulte et de celui de l'enfant.

Dans une perspective pragmatique, en s'appuyant notamment sur la « psychologie culturelle » fondée par Bruner (1983), d'autres études ont tenté de dépasser ces limites en analysant le déroulement des dialogues et leurs liens discursifs, y compris chez des enfants plus âgés. Ces études ont eu pour objectif de montrer que l'insertion de l'enfant dans le dialogue est une condition *sine qua non* du développement langagier (François *et al.*, 1984 ; Hudelot, 1987 pour le français). Cette implication dans le dialogue est aussi le moyen pour l'enfant de s'approprier les constructions de la langue : l'évolution, l'organisation et l'appropriation des ressources langagières s'établissent dans une relation « d'étayage » qui rend possible la transformation d'un niveau actuel par rapport à un niveau potentiel². C'est ce que cherche par exemple à montrer Veneziano (2000), en combinant la dimension pragmatique à une perspective socio-constructiviste. Elle repère dans des dialogues entre des mères et leur enfant de moins de 3 ans des séquences de plus d'un tour de parole, avec des mouvements discursifs de reprises interprétatives de la mère, qui comportent des reformulations dans des énoncés grammaticalement bien formés,

² On verra ici une référence à la notion de *Zone Proximale de Développement* selon Vygotsky (1934).

associées aux reprises imitatives de l'enfant. Les travaux de Ervin-Tripp (2002 pour une synthèse) sur l'impact du cadre interactionnel (pragmatique) montrent également chez des enfants d'âges variables (jusqu'à 5 ans) des différences d'utilisation des formes, et en particulier des constructions syntaxiques, selon les contextes linguistiques et les situations dialogiques. Les travaux en linguistique de l'acquisition de Lentin et de ses collaborateurs se sont plus spécifiquement intéressés aux modalités interactionnelles du premier apprentissage en focalisant sur l'acquisition de la syntaxe entre trois et six ans (pour un historique, voir Canut, 2009).

2. Etude de la complexité syntaxique en linguistique de l'acquisition

Dès les années 70, Lentin (1973) s'appuie sur l'analyse de corpus longitudinaux d'interactions entre adulte et enfant de deux à six ans pour montrer le rôle crucial du fonctionnement syntaxique dans le développement cognitivo-langagier et l'impact déterminant des caractéristiques linguistiques du langage adressé à l'enfant. De ces premières observations naissent une conception théorique et méthodologique de recueil et d'analyse des données, qui s'affinera et se consolidera au fil des années (Lentin, 2009).

2.1. Objectifs poursuivis

Les chercheurs en linguistique de l'acquisition s'inscrivant dans la lignée des travaux de Lentin s'intéressent à la dynamique de l'apprentissage. Ils n'ont pas pour objectif de rechercher des stades ou de comparer les enfants entre eux pour dégager une norme de développement. Ils cherchent, d'une part, à repérer une évolution dans le langage de chaque enfant, des formes émergentes vers des formes stabilisées, maîtrisées. D'autre part, ils cherchent à évaluer les modalités des interactions favorables à cette évolution, c'est-à-dire les interactions se situant dans une zone potentielle de développement langagier et ayant un pouvoir « catalyseur » pour l'intégration des éléments linguistiques au fonctionnement cognitivo-langagier de chaque enfant). Plusieurs études ont déjà montré qu'une absence ou une inadéquation du langage adressé à l'enfant (formulations en deçà ou au-delà de ce que l'enfant peut

capter) peut sévèrement retarder l'acquisition³. Trois axes sont plus particulièrement explorés :

- les reprises et reformulations d'un locuteur à l'autre, considérées comme le facteur clé de l'appropriation des constructions de la langue ;
- les « essais », c'est-à-dire les tentatives de l'enfant d'employer des complexités syntaxiques, considérés comme un fonctionnement cognitivo-langagier essentiel pour un accès progressif à la maîtrise de la langue et pour la structuration de la pensée, en relation avec les reformulations offertes par l'adulte ;
- les « schèmes sémantico-syntaxiques créateurs », dont la mise en œuvre dépend des phénomènes étudiés dans les deux axes précédents, et qui renvoient au processus d'apprentissage du langage dans ses aspects dynamiques et progressifs.

Les schèmes sémantico-syntaxiques créateurs ne fonctionnent pas comme des modèles à mémoriser et à reproduire. Ce sont « des stimulations mentales de mise en relation d'éléments verbalisables. Ce fonctionnement mental permet à l'apprenant, dans d'autres circonstances, de s'approprier pour son propre système langagier un fonctionnement dont on lui a donné l'expérience » (Lentin, 1988, p. 20). Cette conception trouve certains échos dans des travaux récents s'inscrivant dans l'approche *usage-based grammar* (les « grammaires de construction ») avec la notion d'*abstract constructional schemas* (« schémas constructionnels abstraits ») proposée par Diessel (2004) pour rendre compte de l'évolution du langage de l'enfant, allant de *prefabricated chunks* (blocs figés et infinitifs) vers des complexités croissantes (constructions complétives puis circonstanciées).

Assez récemment, le besoin s'est fait jour de généraliser les résultats accumulés au fil des années, plutôt que de continuer à étudier des phénomènes isolés dans des corpus dédiés à une recherche particulière, et de combiner analyse qualitative et quantitative pour fournir une explication générale du processus interactionnel d'acquisition, en particulier pour les constructions syntaxiques complexes et la mise en fonctionnement de leur

³ Voir par exemple les travaux précurseurs de Wyatt (1969), ceux de Bonnel et Lentin (1985) ou encore ceux de Nelson (1984).

combinatoire. Autrement dit, même si les résultats du travail préliminaire ont été complétés et affinés (par exemple Lentin *et al.*, 1988 ; Karnoouh-Vertalier, 1998 ; Canut, 2001 ; Canut et Vertalier, 2009) et se trouvent concorder en partie avec des résultats récents d'études psycholinguistiques (par exemple Clark, 2003), il s'agit d'éprouver l'hypothèse des schèmes sémantico-syntaxiques créateurs en travaillant sur de « grands corpus ».

2.2. Constitution d'un « grand corpus » d'interactions adulte-enfant

Depuis les années 70, et jusqu'à une époque récente (2004), les chercheurs en linguistique de l'acquisition ont enregistré un très grand nombre de données sur des magnétophones à cassettes. Il s'agit principalement de corpus longitudinaux d'interactions individuelles entre adultes et enfants de 2 à 7 ans, constitués de dialogues libres (scènes de vie quotidienne) et le plus souvent de conversations ou de narrations à partir de livres illustrés (recueil spontané ou semi-expérimental). Cette base de données, comme la majorité des corpus recueillis par les chercheurs dans le domaine de l'acquisition en français, n'a pas été pensée pour être rendue immédiatement accessible à la communauté scientifique mais pour répondre à des questionnements de recherche spécifiques. En France, ce n'est que très récemment que les chercheurs ont pris conscience de la nécessité de mutualiser les acquis et de favoriser les échanges, et qu'ils constituent des bases de données informatisées en accès libre⁴. C'est le cas du projet *Léonard (acquisition du langage et grammaticalisation)*⁵ qui devrait comporter à terme de nombreux enregistrements d'enfants monolingues et bilingues (français/anglais) entre 12 mois et 3 ans, filmés dans leur famille. C'est également le cas du projet TCOF (« Traitement de Corpus Oraux en Français »), rattaché à l'Unité

⁴ Les pays anglo-saxons ont été plus précoces : pour l'acquisition, le CHILDES (*CHILd Language Data Exchange System*), fondé par MacWhinney et Snow en 1984, comporte un nombre croissant de corpus transcrits de dialogues adulte-enfant (essentiellement anglophones même s'il existe un élargissement à d'autres langues) et des programmes informatiques pour l'analyse automatique des données. <http://childes.psy.cmu.edu>

⁵ Projet financé par l'Agence Nationale de la Recherche (ANR), coordonné par Morgenstern et actuellement rattaché à la base CLAPI (Corpus de Langue Parlée en Interaction) de l'Université de Lyon 2.

Mixte de Recherche CNRS - Nancy Université : ATILF⁶, qui rassemble des corpus de français parlé et des corpus de dialogues entre adulte et enfant de moins de 7 ans.

L'objectif du projet TCOF est de pouvoir travailler sur de nombreux corpus (équivalant à plusieurs millions de mots), autrement dit sur un nombre de données statistiquement pertinent. Sa vocation première et immédiate est d'archiver les données existantes dans un format moderne (sous forme numérique avec un alignement texte-son) pour ensuite les mettre à disposition de la communauté scientifique. Ainsi, depuis 4 ans, l'équipe a informatisé, stocké et sauvegardé sur une machine dédiée près de 400 heures de données orales dont 175 heures d'enregistrements de dialogues adulte-enfant (voir tableau 1)⁷.

Tableau 1 - Répertoire des données archivées dans le cadre du projet TCOF

Corpus Numérisés	Nombre	Durée totale	Etat des transcriptions en 2009	Disponibilité
<i>Uniques</i> Ages : 2 ans 3 mois à 5 ans 10 mois	140	36 h	Alignement texte-son avec <i>Transcriber</i> (phase de vérification et d'anonymisation)	Une centaine de corpus uniques disponibles en accès libre en 2011
<i>Longitudinaux</i> Ages : 2 ans 10 mois à 7 ans 6 mois	60 corpus	139 h	Alignement texte-son en cours avec <i>Transcriber</i> (phase de transcription)	5 corpus longitudinaux disponibles en accès libre en 2010
TOTAL	1040	175 h		

⁶ Ce projet géré par André, Benzitoun, Canut, Debaisieux, Gaiffe et Jacquey (2008 et à paraître ; Canut, 2008 ; Debaisieux, 2005).

⁷La constitution d'une telle base de données est un investissement important car elle implique la préparation et la vérification des données et toute une réflexion sur leur archivage. Comme le soulignent Cappeau (2005) ou encore Habert (2000 et 2005), un corpus exploitable suppose une taille suffisante (c'est-à-dire des échantillons représentatifs) et une certaine diversité (c'est-à-dire des situations de production variées).

Il a été décidé de numériser toutes les données orales sous le même format WAV (avec le logiciel *Wavelab*) et de faire une transcription alignée avec le son en utilisant le logiciel *Transcriber*⁸. Ces transcriptions sont actuellement en cours de réalisation⁹. Le principe adopté est de transcrire intégralement les énoncés, en orthographe standard, sans trucage orthographique. Au fil de la réflexion, nous avons élaboré plusieurs versions des conventions de transcription. Nous avons adopté le principe d'utilisation des balises internes au logiciel *Transcriber*. Nous reproduisons ci-dessous les symboles utilisés dans les transcriptions (figure 1). Ce passage à la transcription assistée par ordinateur a eu un impact important sur la fiabilité et le temps de transcription.

Figure 1 - Récapitulatif des symboles de transcription

{...}	Commentaires (balise <i>Transcriber</i>)
[...]	Prononciations notées avec l'alphabet phonétique SAMPA (balise <i>Transcriber</i>)
(...)	Variantes graphiques indécidables
+	Pauses
///	Pauses très longues
=	Liaison non standard remarquable
/..., .../	Hésitations entre transcription
...-	Amorces
*	Syllabe incompréhensible
***	Suite de syllabes incompréhensibles
###	Passage enregistré non transcrit
\$\$\$	Coupure de l'enregistrement

L'objectif à plus long terme du projet TCOF est d'utiliser des outils informatiques qui permettraient d'alléger une analyse « à la main », et d'enrichir les analyses syntaxiques existantes. L'avantage sera non seulement la possibilité de confronter entre elles les analyses réalisées sur un grand nombre de corpus de dialogues

⁸ <http://trans.sourceforge.net/>. *Transcriber* étant un logiciel libre, il ne fait pas obstacle à la diffusion de nos corpus.

⁹Le passage d'un fichier au format *Transcriber* accompagné de sa fiche documentaire vers un fichier au format de transcription TEI (*Text Encoding Initiative*) est également en cours d'automatisation.

adulte-enfant, mais également de comparer les caractéristiques linguistiques des verbalisations d'adultes s'adressant à des enfants à celle d'adultes dans des situations de communication ordinaire (récits, conversation, explication, etc.). Dans l'immédiat, tant que toutes les données ne sont pas transcrites, nous ne pouvons pas réaliser cette étude de grande envergure. Nous pouvons néanmoins tester sur des données plus modestes la validité d'outils informatiques comme les concordanciers pour en comparer les résultats à ceux des grilles d'analyse utilisées jusqu'à présent.

2.3. Une méthodologie d'analyse qualitative de l'appropriation de constructions syntaxiques

La linguistique de l'acquisition considérant l'apprentissage du langage dans sa dimension qualitative et prenant le cadre interactionnel comme lieu d'observation du processus d'appropriation, les recherches se donnent pour objectif de :

- repérer une évolution dans la production de constructions syntaxiques, des premières tentatives jusqu'à leur verbalisation maîtrisée, en identifiant les critères linguistiques qui la caractérisent ;
- tenter une description du processus d'appropriation repéré en s'appuyant sur les données du langage adressé à l'enfant ;
- déterminer les processus généraux et les variabilités individuelles de cette appropriation¹⁰.

Pour cela, l'analyse s'appuie sur le repérage, dans les énoncés de l'enfant et de l'adulte, des phénomènes suivants :

- catégories syntaxiques (éléments isolés, association de deux ou trois éléments, juxtaposition et coordination de constructions simples et complexes, profondeur syntaxique des énoncés et combinatoire des constructions complexes) et leur évolution dans le temps ;
- reprises immédiates ou différées par l'enfant de fonctionnements syntaxiques présents dans les verbalisations de l'adulte ;

¹⁰ L'âge n'est pas considéré comme un repère d'évolution. Il ne s'agit pas de savoir à quel âge est acquise telle ou telle construction : cette question n'aurait pas de sens sauf à considérer l'acquisition d'un point de vue normatif, en envisageant un passage de l'absence à la présence d'une forme qui serait considérée comme acquise puisque produite, selon des critères de productivité fixés par les auteurs.

- tentatives d'appropriation de constructions complexes (« essais ») ;
- reprises ou reformulations par l'adulte de configurations présentes dans les énoncés de l'enfant
- verbalisations autonomes de l'enfant.

En ce qui concerne l'appropriation de constructions complexes, ce repérage est réalisé au moyen d'une grille d'analyse en catégories syntaxiques d'énoncés. Nous appliquerons, pour exemple, cette méthodologie à deux corpus de dialogues réalisés à 15 jours d'écart. Ahlem, l'enfant interagissant avec l'adulte, est âgée de 6 ans 3 mois et 18 jours lors du premier enregistrement et de 6 ans 4 mois et 2 jours lors du deuxième enregistrement. Il s'agit d'une situation semi-expérimentale : l'adulte a lu intégralement à l'enfant le texte d'un livre illustré *Le nouveau dentifrice de Mélanie*¹¹ juste avant de proposer à l'enfant de raconter l'histoire en s'appuyant sur les illustrations. L'adulte avait pour objectif d'offrir des reformulations des énoncés de l'enfant se situant dans une zone proche de son fonctionnement langagier et l'excédant légèrement.

Dans la grille, nous classons (manuellement) les énoncés de l'adulte et de l'enfant dans des catégories syntaxiques « larges » : énoncés comportant maximale une construction « simple », plusieurs constructions « simples » juxtaposées ou coordonnées, une ou plusieurs constructions « complexes », essais de constructions complexes. Les constructions simples sont identifiées comme telles, par défaut, en l'absence d'introducteurs de complexité. Le repérage des constructions complexes s'effectue en fonction de la liste d'« introducteurs de complexité » (IC) établie par l'équipe de Lentin (1971)¹². Certains IC, dits sélectionnés, ont été repérés comme présentant une croissance régulière dans un corpus longitudinal, en liaison avec l'articulation du raisonnement

¹¹ Lentin, Lauer, Editions Fleurus, 1988.

¹² Les Introducteurs de Complexité (IC) ont été déterminés expérimentalement au cours d'une enquête préliminaire à la recherche de Lentin, en 1969. La liste a été établie à partir d'un corpus de dialogues entre adultes et enfants âgés de 3 à 7 ans. Depuis la publication de la liste alphabétique originale en 1971, la liste des introducteurs de complexité a connu quelques ajustements mais n'a pas été fondamentalement remise en cause. Au fil des années, la liste des introducteurs de complexité présents dans des énoncés d'enfants entre 3 et 7 ans a été confirmée.

(numéros soulignés dans la figure 3). La présence d'« introducteurs de complexité » syntaxique n'est pas pour autant considérée comme un témoignage d'opérations cognitives nécessairement plus complexes. L'objectif est de repérer l'étendue maximale du répertoire linguistique de l'enfant.

On trouvera dans la figure 2 un extrait de la grille d'analyse en catégories syntaxiques du corpus d'Ahlem¹³. Dans cette grille sont indiqués les numéros (arbitraires) des introducteurs de complexité. Chaque énoncé n'est répertorié qu'une seule fois, dans la catégorie où la construction est maximale.

Figure 2 - Extrait d'une analyse en catégories syntaxiques d'énoncés

Prénom et âge de l'enfant : Alhem 6;3;18 n° corpus : 1

Numéro de l'énoncé E= Ahlem A= Adulte	Enoncés comportant une construction simple		Enoncés comportant des constructions simples multiples		Enoncés comportant des constructions complexes (cf liste IC) + : juxtaposition / : emboîtement		Essais (cf liste IC)
	A	E	A	E	A	E	
(...)							
A20					<u>20/12/13</u> + 10/18 + <u>20/12/13</u>		
E20						<u>20</u>	
A21					<u>13</u>		
(...)							
E23						19	
A24					<u>13</u>		
E24							<u>13</u>
A25					<u>13</u>		
Total	16	16	1	2	8	4	2

Une fois les grilles remplies pour chacun des corpus, nous calculons le pourcentage de chaque catégorie et dans chaque corpus afin de repérer s'il existe une évolution dans le temps. Les résultats des analyses des corpus 1 et 2 d'Ahlem reportés dans le tableau 2 montrent clairement dans le deuxième corpus une

¹³ Faute de place, les corpus transcrits intégralement et les analyses complètes ne peuvent être insérées ici.

diminution chez l'enfant d'énoncés comportant des constructions simples uniques au profit d'énoncés plus longs (comportant plusieurs constructions simples) ou syntaxiquement plus complexes (avec notamment une augmentation des essais et l'apparition de constructions comportant plus d'un introducteur de complexité). Nous pouvons également observer que, dans ce cas, le langage de l'adulte suit la même évolution (augmentation du pourcentage de constructions simples multiples et de constructions complexes).

Tableau 2 - Pourcentage des catégories syntaxiques d'énoncés

N° du corpus	1		2	
	A (/25)	E (/24)	A (/30)	E (/29)
Adulte (A) Enfant (E= Ahlem)				
Constructions simples (%)	64	66,66	43,33	37,93
Constructions simples multiples (%)	4	8,33	10	20,69
Constructions complexes (%)	32	16,66	46,66	20,69
Essais (%) ¹⁴	-	8,33	-	20,69

Ces résultats ne sont pas isolés : les analyses de plusieurs autres corpus longitudinaux (étudiés sur une période plus longue) indiquent des évolutions similaires. Néanmoins, une analyse qualitative est nécessaire pour déterminer ce qui a exactement évolué et pour évaluer l'impact du discours de l'adulte. De ce point de vue, l'emploi du discours indirect est un phénomène intéressant à observer dans le corpus d'Ahlem. Sans prétendre en faire ici une analyse exhaustive, on peut voir les tâtonnements de l'enfant pour passer du discours direct non introduit à du discours direct avec introduction du locuteur, en passant par des essais de discours indirect (construction avec *dire que/dire de*). L'énoncé charnière se situe à l'énoncé E15 du corpus 1 quand l'enfant passe brusquement d'une narration à la troisième personne à du discours direct (après une pause et un grand soupir) :

¹⁴La catégorie des essais ne concerne que les productions de l'enfant.

E 13 et quand elle a fini de broser ses dents quand elle a fini de broser ses dents elle donne à sa soeur elle donne à sa soeur euh
A 14 le dentifrice elle prête le dentifrice à sa soeur
E 14 elle prête le dentifrice à sa soeur
A 15 voilà quand elle a fini de se broser les dents elle prête le dentifrice à sa soeur
E 15 mm + {soupir} euh ouvre la bouche

Suit une séquence où l'adulte fait préciser à l'enfant le locuteur des paroles rapportées :

A 16 qui est-ce qui dit ça ?
E 16 sa grand-mère
A 17 sa maman
E 17 sa maman
A 18 d'accord sa maman

L'enfant poursuit son récit en produisant de nouveau des paroles rapportées non introduites par un locuteur (E18 et E19), ce qui amène l'adulte (A20) à proposer une reformulation de l'énoncé avec, parmi d'autres constructions complexes, un discours indirect¹⁵ :

E 18 mmm ça sent bon
A 19 alors
E 19 ça sent bon ton dentifrice à la fraise
A 20 oui la maman en regardant si les dents de Mélanie sont bien propres elle dit que le dentifrice sent bon la framboise tu veux essayer d(e) tout me redire ? en regardant si les dents de Mélanie sont bien propres maman dit que le dentifrice sent bon la framboise

Sur la proposition de l'adulte, Ahlem tente une première fois de reprendre cette construction de discours indirect mais réutilise le discours direct sans introduire le locuteur (E20). L'énoncé est une nouvelle fois reformulé en discours indirect par l'adulte (A21) :

E 20 en en regardant les dents de Mélanie de Mélanie ça sent euh ça sent bon à la framboise
A 21 elle dit qu(e) ça sent bon la framboise ensuite qu'est-ce qu'elle fait Mélanie ?

¹⁵ Il s'agit ici d'une reprise quasi identique d'une phrase du texte du livre support.

Deux énoncés plus loin, l'enfant produit encore un discours non introduit (E23), immédiatement reformulé par l'adulte en A24 en discours indirect (sous la forme *dire de* cette fois). En E24, l'enfant reprend la forme introductrice du discours indirect *dire que* mais y insère la construction du discours direct (*donne*). En A25, l'adulte reformule cet essai en énonçant la forme du discours indirect qui convient dans ce contexte :

*E 23 et donne la main à ta grande soeur pour y aller jusqu'à l'école jusqu'à l'école ****

A 24 oui maman dit à Mélanie de donner la main à sa grande soeur jusqu'à l'école maternelle

E 24 mais sa maman lui dit que donne la main à ta grande soeur jusqu'à la maternelle

A 25 d'accord elle dit de donner la main à sa grande soeur jusqu'à l'école maternelle c'est bien Ahlem

Dans ce premier corpus, on remarque que l'enfant emploie d'autres complexités (des constructions temporelles : *quand* en E13, le gérondif en E20). Les reformulations de l'adulte sont adaptées car elle reprend les complexités employées par l'enfant et lui en propose d'autres, ce qui contribue à faire émerger des constructions non encore maîtrisées. Ahlem est potentiellement capable de produire la construction du discours indirect mais elle a encore besoin d'expérimenter ce fonctionnement syntaxique à partir des propositions faites par l'adulte¹⁶. Ces tentatives indiquent toutefois qu'Ahlem n'en est qu'au début de son apprentissage de cette construction. Nous avons d'ailleurs observé que dans le deuxième enregistrement, malgré les propositions de l'adulte, l'enfant n'a pas fait d'autres essais du discours indirect mais qu'une autre évolution se fait jour : Ahlem introduit systématiquement les locuteurs des paroles rapportées (par exemple en E24 et E29) :

E 24 sa maman euh sa maman lui dit oh tu sens bon la framboise

E 29 sa maman lui dit sa maman lui dit donne ta main à ta grande soeur pour arriver jusqu'à l'école maternelle

Avant de pouvoir acquérir le fonctionnement du discours indirect l'enfant a probablement besoin d'éprouver le

¹⁶Dans le corpus d'Ahlem, on pourrait faire le même type d'analyse avec les subordonnées temporelles, en particulier celles introduites par *pendant que*.

fonctionnement du discours direct et des modes d'explicitation des instances énonciatives. Le chemin vers la maîtrise du discours indirect sera donc encore probablement long et son appropriation sera tributaire des offres et des ajustements des adultes qui interagiront avec elle.

3. Pour conclure : les outils informatique peuvent-ils nous aider dans l'analyse ?

Une analyse « manuelle » sur grilles en catégories syntaxiques d'énoncés est coûteuse en temps. On peut difficilement travailler rapidement sur un grand nombre de corpus. Par ailleurs, une analyse manuelle n'est pas à l'abri de l'erreur humaine : oublis d'occurrences, vérification aléatoire. De ce point de vue, un certain nombre d'outils informatiques sont en mesure de résoudre facilement ces deux problèmes : possibilité de prendre appui sur une base de données importante, gain de temps et fiabilité. Dans notre démarche, la phase de repérage des phénomènes peut être facilitée par des outils informatiques, comme avec un concordancier. Un concordancier permet, entre autres, de répertorier dans un ou plusieurs corpus un même élément (ou une chaîne de caractères) et d'afficher les occurrences en contexte. Nous avons opté pour l'utilisation du logiciel (libre et gratuit) *Antconc*¹⁷, qui est relativement simple d'utilisation. Pour repérer dans les corpus les complexités syntaxiques, nous pouvons demander au logiciel de rechercher les introducteurs de complexité, les uns après les autres. Cette façon de faire nous permet de noter la présence ou l'absence d'un élément produit de façon canonique et sa fréquence chez l'enfant et l'adulte. La copie d'écran en figure 3 montre le résultat obtenu dans les corpus 1 et 2 d'Ahlem pour le temporel *pendant que*.

¹⁷http://www.antlab.sci.waseda.ac.jp/antconc_index.html. Les fichiers insérés dans *Antconc* doivent être au format texte (.txt). Les fonctionnalités de ce concordancier sont notamment : affichage de toutes les occurrences du mot avec son contexte droit et gauche (*File View*), recherche de plusieurs mots ou d'une chaîne de caractère (*Regex*), liste des mots des plus fréquents aux moins fréquents (*Word List*), repérage de co-occurrences directes (*Clusters*), visualisation d'un diagramme de répartition des occurrences dans le corpus (*Concordance Plot*).

Figure 3 - Copie d'écran du concordancier *Antconc* : résultats pour *pendant (que)*

Concordance	Concordance Plot	File View	Clusters	Collocates	Word List	Keyword List
KWIC						File
lors elle ouvre son paquet pendant qu'Emile et Florence mangent						Corpus3Ahlem.t:
son paquet avant que + A 5 pendant qu'Emile et Florence mangent						Corpus3Ahlem.t:
Florence mangent E 5 pendant Emile et Florence mangent A 4						Corpus3Ahlem.t:
nie ouvre son petit paquet pendant que son frère et sa soeur Em:						Corpus3Ahlem.t:
e ouvre son petit paquet + pendant que euh + euh + <*** A 7 tu :						Corpus3Ahlem.t:
nie ouvre son petit paquet pendant qu'Emile et Florence mangent						Corpus3Ahlem.t:
nie ouvre son petit paquet pendant que euh sa soeur et s/ sa soe						Corpus3Ahlem.t:
re euh {E souffle} *** A 8 pendant qu'Emile et Florence mangent						Corpus3Ahlem.t:
t leurs céréales E 8 pendant Emile et Florence mangent le:						Corpus3Ahlem.t:

Cependant, ce type d'outil informatique a également ses limites. En effet, s'il permet un repérage automatique quasi instantané des complexités, et donc de travailler sur un nombre de données plus important, il ne répertorie pas d'emblée les formes non canoniques, approximatives, les essais de l'enfant (il faudrait les connaître par avance ou les découvrir par hasard). Il ne donne pas non plus une vision globale des phénomènes de reprises et de reformulations dans le dialogue. Le retour à une analyse « manuelle » des corpus est indispensable pour détecter des processus d'appropriation comme les « essais » et pour étudier les interactions qui favorisent l'émergence de ces essais et la stabilisation de constructions en cours d'acquisition. Autrement dit, les procédures informatisées ne remplaceront jamais l'intuition et l'expérience du linguiste mais elles sont un complément non négligeable de l'étude qualitative du langage enfantin pour permettre une généralisation des résultats.

Références

Amy, G. (1983). Étude génétique de la compréhension des phrases relatives. In J.-P. Bronckart, M. Kail et G. Noizet (Eds.),

- Psycholinguistique de l'enfant : recherches sur l'acquisition du langage* (pp. 135-151). Neuchâtel : Delachaux et Niestlé.
- André, V., Benzitoun, C., Canut, E., Debaisieux, J.-M., Gaiffe, B., Jacquy, E. (à paraître). Traitement informatique de données orales : quels outils pour quelles analyses ?. Actes du Colloque International Logiciels pour l'analyse qualitative ; innovations techniques et sociales, 21 et 22 octobre 2008, Luxembourg, *Revue de l'Association pour la Recherche Qualitative (ARQ)*. Québec.
- André, V., Benzitoun, C., Canut, E., Debaisieux, J.-M., Gaiffe, B. et Jacquy, E. (2008). Fiche documentaire pour les données orales : description et traitement informatique. *Colloque International Catalogage et Codage des Corpus Oraux*, Université d'Orléans, 4-5 décembre 2008.
- Aparici, M., Serrat, E., Capdevila, M. & Serra, M. (2001). Acquisition of complex sentences in Spanish and Catalan speaking children. In K. E. Nelson, A. Aksu-Koç & C. E. Johnson (Eds.), *Children's language*, Vol. 11 (1-25). Hillsdale, NJ : Lawrence Erlbaum Associates.
- Bates, E. & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3-73). Cambridge : Cambridge University Press.
- Bloom, L. (1991). *Language development from two to three*. Cambridge : Cambridge University Press.
- Bonnel, B. et Lentin, L. (1985). *Apprendre à parler : le rôle de l'école maternelle*. Paris : PUF / CTNERHI.
- Bruner, J. S. (1983). *Le développement de l'enfant : savoir faire, savoir dire*. Paris : P.U.F.
- Bruner, J. S. (1991). *...car la culture donne forme à l'esprit. De la révolution cognitive à la psychologie culturelle*. Paris : Eshel.
- Canut, E. (2008). Présentation et avancée du projet Traitement de corpus oraux en français. Description et comparaison de productions langagières (interactions entre adulte, entre adulte et enfant). *L'Acquisition du Langage Oral et Ecrit*, 60-61, AsFoReL, 35-44.
- Canut, E. (2009). La syntaxe : un socle indispensable à l'apprentissage du langage. Les conceptions interactionnistes de la linguistique de l'acquisition à la lumière des approches

- contemporaines. L'exemple de parce que. In E. Canut et M. Vertalier (Eds.). *L'apprentissage du langage : une approche interactionnelle. Réflexions théoriques et pratiques de terrain* (pp. 71-128). Paris : L'Harmattan.
- Canut, E et Vertalier, M. (Eds.) (2009). *L'apprentissage du langage : une approche interactionnelle. Réflexions théoriques et pratiques de terrain. Mélanges offerts par ses collègues, ses élèves et ses amis en hommage à Laurence Lentin*. Paris : L'Harmattan.
- Cappeau, P. (2005). Ce que l'oral nous a appris sur la syntaxe du français. *Lidil*, 31, 157-177
- Clark, E. (2003). *First language acquisition*. Cambridge : Cambridge University Press.
- Debaisieux, J.-M. (2005). Les corpus oraux : Situation, exploitation linguistique, bilan et perspectives. *Varia, Scolia*, 19, 9-40.
- Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge : Cambridge University Press.
- Ervin-Tripp, S. (2002). Development pragmatics : state of art and perspectives. In J. Bernicot, A. Trognon, M. Guidetti, et M. Musiol (Eds.), *Pragmatique et psychologie* (pp. 179-193). Nancy : Presses Universitaires de Nancy.
- François, F., Hudelot, C. et Sabeau-Jouannet, É. (1984). *Conduites linguistiques chez le jeune enfant*. Paris : P.U.F.
- Gallaway, C. & Richards, B. (1994). *Input and Interaction in Language Acquisition*. Cambridge : Cambridge University Press.
- Habert, B. (2000). Des corpus représentatifs : de quoi, pour quoi, comment ? *Cahiers de l'Université de Perpignan*, 31, 11-58.
- Habert, B. (2005). Face à la disette dans la profusion. *Scolia*, 19, 41-61.
- Hudelot, C. (1987). Organismes discursifs du dialogue adulte-enfant : une esquisse tracée à partir de quelques dialogues (enfants de 5 à 6 ans). *Modèles linguistiques*, IX:1, 33-51.
- Kail, M. (2000). Acquisition syntaxique et diversité linguistique. In M. Kail et M. Fayol (Eds.), *L'acquisition du langage. Le langage en développement au-delà de trois ans* (pp. 9-44). Paris : P.U.F.

- Karnouh-Vertalier, M. (1998). Évolution du fonctionnement syntaxique et variantes énonciatives. Observation d'interactions langagières entre adulte et enfant au cours d'activités de narration. *Langue Française*, 118, 84-103.
- Lécuyer, R. (2002). Introduction : Inné fable ? *Intellectica*, 34, 11-28.
- Lentin, L. (1971). Recherches sur l'acquisition des structures syntaxiques chez l'enfant entre 3 et 7 ans : une méthode, premiers résultats. *Études de Linguistique Appliquée*, 4, 7-52.
- Lentin, L. (1973). Interaction adultes-enfants au cours de l'acquisition du langage. Un exemple : Étude contrastive du langage de deux enfants entre 3 et 4 ans et de celui de leurs parents (milieux socio-culturels différents). *Études de Linguistique Appliquée*, 9, 9-50.
- Lentin, L. (2009). *Apprendre à penser, parler, lire, écrire. Acquisition du langage oral et écrit*. Paris : E.S.F.
- Lentin, L. et al. (1988). *Recherches sur l'acquisition du langage*. Vol. 2. Paris : Presses de la Sorbonne Nouvelle - Paris 3.
- Lust, B. (2006). *Child Language. Acquisition and Growth*. Cambridge : Cambridge University Press.
- Nelson, K. E., Kaplan, B. J., Denninger, M. S., Bonvillian, J. D. & Baker, N. D. (1984). Maternal Input Adjustments and Non-Adjustments as Related to Children's Linguistic Advances and to Language Acquisition Theories. In A. Pelligrini & T. Yawkey (Eds.), *The Development of Oral and Written Language in Social Contexts*, Vol XIII (pp. 31-56). Norwood - New Jersey : Ablex.
- Peterson, C. & MacCabe, A. (1987). The connective 'and': Do older children use it less as they learn other connectives ? *Journal of Child Language*, 14, 375-381.
- Radford, A. (1995). Phrase structure and functional categories. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 483-507), Oxford : Blackwell.
- Snow, C. E. (1995). Issues in the Study of Input: Finetuning, Universality, Individual and Developmental Differences, and Necessary Causes. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 180-193), Oxford : Blackwell.
- Tomasello, M. (2000). Do young children have adult syntactic competence ? *Cognition*, 74, 209-253.

- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge : Harvard University Press.
- Veneziano, E., (2000). Interaction, conversation et acquisition du langage dans les trois premières années. In M. Kail et M. Fayol (Eds.), *L'acquisition du langage. Le langage en émergence de la naissance à trois ans* (pp. 231-265). Paris : P.U.F.
- Vygotsky, L. S. (1934/1986). *Thought and Language*. Cambridge : MIT Press.
- Wyatt, G. (1969/1973). *La relation mère-enfant et l'acquisition du langage*. Bruxelles : Mardaga.