

Acquisition des constructions syntaxiques complexes chez l'enfant français entre 2 et 6 ans

Canut, Emmanuelle

Université de Lorraine – ATILF UMR CNRS 7118
Emmanuelle.Canut@univ-lorraine.fr

Dans le domaine de l'acquisition du français langue première, les constructions syntaxiques complexes ont essentiellement fait l'objet de recherches expérimentales ou d'analyses qualitatives sur un nombre de corpus limité, ce qui s'explique, entre autres, par le fait qu'il n'y a pas toujours de corpus disponibles diffusés et exploitables et qu'ils concernent pour beaucoup des enfants de moins de 3 ans. Notre objectif a été de faire une étude qui s'appuie sur un « grand » corpus de dialogues recueillis en situation de production réelle et de faire, pour les enfants de 2 à 6 ans :

- une exploration simultanée de plusieurs constructions syntaxiques complexes ;
- une comparaison entre deux types de discours : la conversation ordinaire et la narration dialoguée ;
- une étude développementale de la production des complexités syntaxiques en fonction de l'âge et du type de discours.
- Une comparaison de la production des complexités syntaxiques chez les adultes et les enfants en fonction de l'âge et du type de discours.

1 Les études sur le développement syntaxique en psycholinguistique

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, qui cherche à décrire des mécanismes généraux de traitement de l'information et à déterminer des étapes de développement en fonction des capacités cognitives de l'enfant et des contraintes linguistiques des langues considérées.

Du point de vue de l'analyse syntaxique des données, nombre d'études, majoritairement anglo-saxonnes, portent sur l'émergence et la compréhension précoce des relations grammaticales (avant trois ans), qu'elles s'inscrivent dans des approches cognitivistes ou fonctionnalistes (par exemple Hirsh-Pasek & Golinkoff, 1996). Pour la production, depuis le travail fondateur de Brown (1973), plusieurs autres études longitudinales s'appuyant sur du langage « spontané » se sont intéressées aux premiers stades d'acquisition des structures syntaxiques, du passage des énoncés à deux mots aux phrases coordonnées et subordonnées, entre 0 et 3 ans (notamment Bloom, 1991). Pour le français, on peut citer le travail de Clark (1998, 2003) sur l'acquisition des premières combinaisons de mots et l'intégration du lexique dans des structures de phrases de plus en plus complexes. Malgré des disparités dans les résultats, les recherches permettent de dégager 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 4 ou 5 ans (voir par exemple Eisenberg et Cairns, 1994 sur le développement des *object embedded clauses*). 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*), puis les complétives avec *that*. Les subordonnées temporelles autres que *when* (*while, since, after, before*) sont

rare avant 3 ans (Bloom et al, 1980, 1987 et 1989 ; Diessel, 2004, Peterson & McCabe, 1987). Cependant, l'ordre d'apparition et le degré de maîtrise de ces complexités sont variables selon les enfants.

Pour le français, on observe des similitudes avec l'anglais : d'abord l'émergence 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 *parce que* pour marquer les relations causales, le temporel *quand*, ou encore le conditionnel *si*, et un peu plus tard *que* pour introduire les relatives et les complétives (Clark, 2003, François, François, Sourdot et Sabeau-Jouannet, 1977, Le Normand, Moreno-Torres, Parisse et Dellatolas, 2013).

Chez l'enfant entre 3 et 6 ans - période durant laquelle la structuration du langage se complexifie et se diversifie- il existe une longue tradition de recherche expérimentale pour la compréhension et la production des subordonnées, en particulier le traitement de la compréhension des phrases passives et des relatives (Kail, 2000 pour une synthèse). Les résultats de ces études expérimentales mettent en évidence que plus les enfants sont jeunes plus ils auraient de difficultés à comprendre certaines phrases complexes. Or, les enregistrements de productions langagières authentiques, en conversation, montrent que les enfants en produisent dès l'âge de 2-3 ans.

Certains auteurs s'appuient sur des corpus (longitudinaux) de dialogues libres ou de narrations dialoguées d'enfants de plus de trois ans pour étudier la production des constructions syntaxiques complexes. Ainsi, dès les années 70, les travaux de Lentin et collaborateurs, se sont focalisés sur l'évolution de la syntaxe dans de nombreuses productions attestées d'enfant entre 3 et 6 ans dans diverses situations (Canut, 2001 et 2013a ; Lentin, 1971, 1975 ; Lentin et al, 1984, 1988). Au cours d'une étude préliminaire concernant une soixantaine de corpus d'enfants français de 3 à 7 ans, Lentin a répertorié les constructions syntaxiques dont le pourcentage présente une évolution globale. Elle a mis en évidence que certains éléments apparaissaient « comme les plus significatifs de la progression de la complexité syntaxique en liaison avec l'articulation du raisonnement dans le langage en voie d'acquisition » (Lentin, 1984, p. 23), dans la mesure où leur pourcentage dans les corpus présentait une augmentation (comme les relatifs, les conjonctions de subordinations) alors que les autres intervenaient en pourcentage à peu près constant (comme les constructions infinitives). Ces résultats trouvent des concordances avec des travaux plus récents s'inscrivant dans l'approche « usage-based grammar », considérant la grammaire comme un système dynamique en lien avec les mécanismes psychologiques de l'utilisation du langage (Diessel, 2004; Tomasello, 2003, Sekali, 2012). Diessel utilise la notion de « abstract constructional schemas » pour rendre compte de l'évolution du langage de l'enfant allant des « prefabricated chunks » (comme les infinitifs) vers des complexités croissantes (constructions complétives puis circonstancielles).

Cependant, comme les nombreuses données recueillies dans ce cadre jusque dans les années 2000 n'ont pas fait l'objet d'un archivage systématique, les analyses ne reposent pas sur un corpus de grande envergure et ne peuvent donc aboutir à des généralisations : le peu d'études portant sur la production des constructions complexes dans le langage spontané des enfants est dû, entre autres, à un manque de corpus disponibles. A titre d'exemple, l'étude de Diessel (2004) sur les complexités syntaxiques du langage d'enfants de langue anglaise est considérée comme la plus importante alors qu'elle ne s'appuie « que » sur 12 000 énoncés référencés (issus de la base CHILDES) et que seulement deux enfants sont étudiés au-delà de la période de 3 ans 5 mois (jusqu'à 5 ans 1 mois), permettant néanmoins de travailler sur un grand nombre d'énoncés pour chacun d'eux. Pourtant, ainsi que le soulignait déjà Moreau et Richelle au début des années 80, « une analyse exhaustive de corpus spontanés supposerait résolues un certain nombre de questions jusqu'ici sans réponse. » (1981, p. 28). Cette suggestion est toujours d'actualité, à l'image de la réflexion de Tomasello et Stahl (2004) indiquant que la notion d'échantillon de langage, telle qu'elle est pratiquée dans les approches psycholinguistiques de l'acquisition du langage, est susceptible d'être ébranlée, avec l'essor de la linguistique de corpus.

2 Un grand corpus d'acquisition en français pour étudier les complexités syntaxiques

Notre étude récente de la production des complexités syntaxiques chez l'enfant français s'appuie sur un grand nombre de corpus oraux archivés dans la base de données TCOF (« Traitement de Corpus Oraux en Français »)¹. Il s'agit de corpus d'interactions individuelles entre des adultes et 97 enfants âgés de 2 ans 3 mois à 6 ans 5 mois. Les enfants sont issus principalement des classes moyennes et ont le français pour langue première. Les corpus sont constitués de 87 enregistrements de conversations avec des enfants différents (12h) et de 10 corpus longitudinaux de narrations dialoguées (9h), dont 3 enfants sur une période d'un an, 6 enfants sur une période de 4 à 6 mois et 1 enfant sur une période de moins de 3 mois. Les conversations sont « spontanées » et toutes produites à un moment donné de la vie de chacun des enfants. Les narrations ont été recueillies à différents moments de la vie de dix enfants, à des intervalles à peu près réguliers. Il avait été demandé aux enfants de raconter, en s'appuyant sur les illustrations, l'histoire d'un livre que l'adulte lui avait lu préalablement. L'adulte avait pour « mission » d'interagir avec l'enfant pour l'aider à organiser sa narration².

Répertoire général des corpus sélectionnés pour l'étude							
Corpus	Type de discours	Nombre de enfants	Nombre corpus	Nombre corpus par âge *	Durée	Nombre mots**	Nombre énoncés**
Transversaux	Conversations libres entre un adulte et un enfant	87	87	2 ans = 12 3 ans = 25 4 ans = 26 5 ans = 24	12h02	Enfants : 50 937 Adultes : 61 721	Enfants : 8896 Adultes : 8953
Longitudinaux	Narrations dialoguées	10	90	3 ans = 17 4 ans = 39 5 ans = 34	8h54	Enfants : 26 003 Adultes : 47 441	Enfants : 4245 Adultes : 4520
TOTAL	-	97	177	-	20h56	Enfants : 76 940 Adultes : 109 162	Enfants : 13 141 Adultes : 13 473
*On considère que les enfants âgés entre 2 ans et 2 ans 11 mois sont classés dans le groupe des 2 ans, et ainsi de suite pour les autres tranches d'âge.							
** Le nombre de mots et d'énoncés a été calculé par le logiciel <i>Transcriber</i>							

Pour procéder à l'analyse, nous avons sélectionné les constructions complexes que nous avons considérées comme base pertinente pour l'étude de l'évolution syntaxique du langage de l'enfant, en nous appuyant sur la liste revisitée des « introducteurs de complexité » de Lentin, (1971) :

- Verbe suivi d'un verbe à l'infinitif (hors futur proche)
- Préposition (*à/de/sans*) suivie d'un verbe à l'infinitif (dorénavant Prep_Vinf), excepté *pour* (voir infra)
- *Que* conjonction
- Discours indirect (dorénavant disc. ind.)
- Interrogative indirecte (dorénavant int.ind.)
- *Parce que*
- *Puisque*
- *Comme* (dans le sens « étant donné que »)
- *Pour* suivi d'un verbe à l'infinitif (dorénavant *pour_Vinf*)
- *Pour que*
- *Quand*
- Gérondif
- Autres temporels (*pendant que, dès que...*)
- *Si* conditionnel
- Les relatives : *Qui, Que, Où*
- Les quantitatives (*plus que/moins que/pareil que...*)
- Les clivées
- Divers autres (*tellement que/ tant que/ alors que/tandis que...*)

Nous les avons réparties dans cinq catégories syntaxiques : complétives, relatives, circonstancielles, clivées et autres types de constructions.

De l'extraction opérée dans les corpus, à l'aide d'un concordancier *JConc*, développé dans le cadre du projet TCOF³, nous avons obtenu, toutes catégories confondues, 1692 occurrences de complexités syntaxiques chez les enfants et 2834 chez les adultes. Cependant, pour réaliser l'étude par groupe d'âge, nous avons harmonisé le nombre d'enfants : nous avons pondéré par deux le nombre de corpus d'enfants de 2 ans en conversation et par deux le nombre de corpus d'enfants de 3 ans en narration, puisque dans les deux cas ils étaient moitié moins nombreux que dans les autres groupes d'âges. *A fortiori* cette pondération a été reportée sur le nombre de complexités syntaxiques chez les enfants et les adultes⁴. Nous avons finalement travaillé à partir des valeurs suivantes : 1792 occurrences de complexités syntaxiques dans les énoncés des enfants dans 206 corpus et 3151 occurrences dans les énoncés des adultes dans les mêmes 206 corpus.

Par ailleurs, dans la mesure où les données entre les deux types de discours (narration et conversation) ne sont pas homogènes, nous avons procédé à des comparaisons intra-groupe. Nous avons ainsi réalisé une analyse descriptive des constructions complexes dans le langage de l'enfant, qui s'appuie sur des pourcentages mais nous l'avons doublée d'analyses statistiques pour mieux vérifier la significativité de certaines évolutions. Nous avons notamment utilisé le test statistique ANOVA pour l'analyse de la variance (réalisé avec SPSS 18.0)⁵.

3 Résultats généraux issus de l'extraction des constructions complexes dans les énoncés des enfants entre 2 ans 3 mois et 6 ans 5 mois

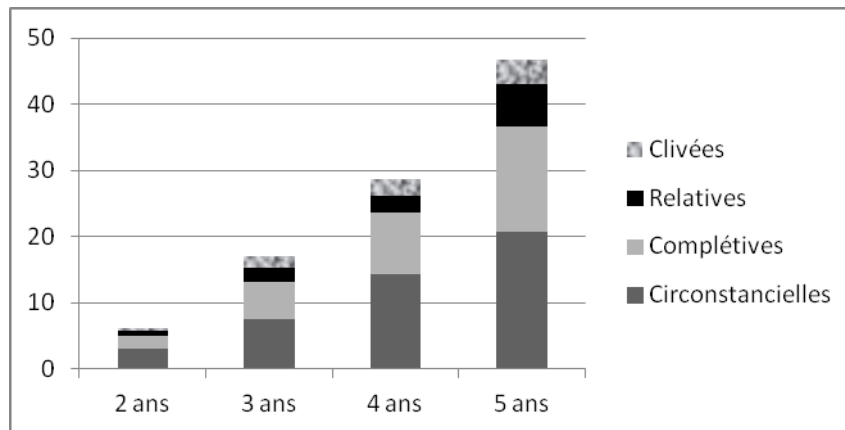
Les résultats issus de l'extraction des complexités pour le corpus dans sa globalité donnent les circonstancielles puis les complétives comme les constructions les plus fréquentes (par rapport au nombre total de constructions complexes pondérées avec le nombre de corpus d'enfants, soit 1792). Viennent ensuite les relatives et les clivées (dispositifs). Les comparatives, les corrélatives et les oppositions sont faiblement représentées dans le langage des enfants. Mais à l'intérieur de ces cinq catégories, certaines constructions émergent plus massivement (tableau 1) : dans la catégorie des complétives, ce sont les constructions infinitives (introduites par une préposition) et les conjonctions *que* qui sont proportionnellement les plus représentées. Dans la catégorie des circonstancielles, ce sont *parce que*, *pour* suivi d'un verbe à l'infinitif et *quand* qui sont les plus représentées, et pour la catégorie des relatives, il s'agit de *qui*.

Catégories	Types de constructions	Complexités syntaxiques	%
Complétives	Infinitives	Prep_Vinf (<i>à/de/sans</i>)	11,49
	Conjonction	<i>Que</i>	11,44
	Discours indirect		4,07
	Int. indirecte		5,86
Circonstancielles	Causalité	<i>Parce que</i>	19,58
		<i>Puisque</i>	0,06
		<i>Comme</i> (étant donné que)	1,23
	But	Pour_Vinf	9,43
		<i>Pour que</i>	1,56
	Temporalité	<i>Quand</i>	7,09
		Gérondif	1,73
		<i>Autres</i> (<i>pendant que, dès que...</i>)	1,62
Condition	<i>Si</i>	3,24	
Relatives		<i>Qui</i>	7,14
		<i>Que</i>	1,4
		<i>Où</i>	3,13
Quantitatives	Comparative	<i>Plus que/moins que/pareil que...</i>	0,73
Divers	corrélation	<i>Tellement vite que/ tant que</i>	0,22
	Opposition	<i>Alors que/tandis que...</i>	0,5
Dispositifs	Clivées	<i>extraction : c'est/il y a...qui/que/où</i>	8,48

Tableau 1 : Répartition en pourcentage des complexités syntaxiques dans les énoncés des enfants

4 Evolution de la production des complexités syntaxiques en fonction de l'âge

La proportion de constructions complexes augmente avec l'âge : on passe d'environ 6 % à 2 ans à 17 % à 3 ans, puis à 29 % à 4 ans pour finir à 47 % à 5 ans. Cette évolution est progressive pour toutes les catégories syntaxiques (graphique 1)⁶ :



Graphique 1 : Répartition des catégories syntaxiques complexes chez les enfants selon le groupe d'âge

Si l'on s'appuie sur les moyennes d'occurrences dans chaque groupe d'âge (tableau 2), l'analyse statistique de la variance (ANOVA) valide cette évolution pour les trois principales catégories syntaxiques : les différences entre les groupes d'âge sont significatives pour les complétives ($F(3, 202) = 12,03, p < 0,05$), les circonstanciellles ($F(3, 202) = 11,67, p < 0,05$) et les relatives ($F(3, 202) = 10,4, p < 0,05$).

Groupe d'âge		Complétives	Circonstanciennes	Relatives	Autres	Clivées
2 ans	Moyenne	1,42	2,33	0,58	0	0,25
	N	24	24	24	24	24
	Ecart-type	2,62	2,26	0,97	0	0,44
3 ans	Moyenne	1,71	2,29	0,63	0,10	0,54
	N	59	59	59	59	59
	Ecart-type	2,88	3,55	1,3	,3	1,15
4 ans	Moyenne	2,58	3,92	0,68	0,12	0,74
	N	65	65	65	65	65
	Ecart-type	3,04	3,79	1,03	,54	1,3
5 ans	Moyenne	4,93	6,38	1,96	0,21	1,14
	N	58	58	58	58	58
	Ecart-type	3,98	5,14	2,24	0,49	2,09

Tableau 2 : Evolution chez les enfants du nombre de complexités (regroupées en catégories) selon le groupe d'âge (moyennes et écarts-types)

Les tests *post hoc* de Bonferroni (différenciation des groupes d'âges 2 à 2) montrent néanmoins que l'évolution la plus significative s'effectue à partir de 5 ans : ainsi, la moyenne des complétives des enfants du groupe des 2 ans (1,42) est équivalente à celle des enfants du groupe des 3 ans (1,71) qui est équivalente à celle des enfants du groupe des 4 ans (2,58) mais ces trois moyennes sont inférieures à celle des enfants du groupe des 5 ans (4,93). Il en est de même pour les circonstanciennes et les relatives : 2 ans = 3 ans = 4 ans < 5 ans. Les moyennes des groupes d'âge des 2, 3 et 4 ans pour ces trois catégories ne se différencient donc pas et c'est la moyenne des enfants du groupe des 5 ans qui est supérieure aux autres.

Si l'on regarde de façon plus détaillée les complexités concernées, cette évolution à l'âge de 5 ans se confirme (tableau 3).

	2 ans	3 ans	4 ans	5 ans
Que conj/ disc. ind. / int. ind	1,79	3,01	5,92	10,66
Parce que	1,12	3,46	7,14	7,87
Prep_Inf (à/de/sans)	0,11	2,62	3,46	5,3
Pour_Vinf	1,23	1,84	2,06	4,3
Clivées	0,33	1,79	2,68	3,68
Quand	0,56	1,28	1,62	3,63
Qui relatif	0,78	1,67	1,45	3,24
Où relatif	0	0,28	0,22	2,62
Gérondif/ autres temporels (pendant que, dès que...)	0	0,22	1,12	2,01
Si condition	0,11	0,33	1,28	1,51
Puisque /Comme	0	0,06	0,5	0,73
comparative/corrélative/opposition	0	0,34	0,45	0,67
Pour que	0,11	0,33	0,5	0,61
Que relatif	0	0,11	0,78	0,5

Tableau 3 : Nombre moyen de chacune des complexités syntaxiques par groupe d'âge

L'analyse statistique à partir des moyennes d'occurrences dans chaque groupe d'âge affine encore ce résultat. Par exemple, dans la catégorie des complétives, l'analyse réalisée à partir des moyennes d'occurrences dans chaque groupe d'âge indique que la conjonction *que* ($F(3, 202) = 3,05, p < 0,05$), le discours indirect ($F(3, 202) = 9,38, p < 0,05$), l'interrogative indirecte ($F(3, 202) = 7,63, p < 0,05$) et les constructions infinitives ($F(3, 202) = 7,85, p < 0,05$) évoluent significativement. Il en est de même pour *parce que*, *pour* + verbe infinitif, les temporels (*quand*, gérondif,...), les relatifs. En revanche, les circonstancielles *si*, *pour que*, *comme* et *puisque*, les constructions clivées, les comparatives et les oppositions ne différencient pas significativement les groupes d'âge.

Parmi les complexités qui différencient les groupes d'âge, les tests *post hoc* de Bonferroni indiquent que les évolutions les plus significatives s'effectuent à partir de 4 ou 5 ans (tableau 4).

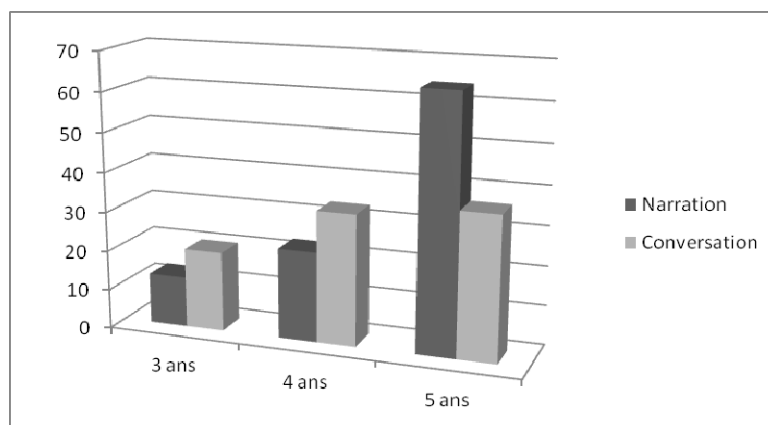
Test <i>post hoc</i> Bonferroni	Complexités syntaxiques
2 ans = 3 ans = 4 ans < 5 ans	Préposition + Verbe infinitif / Discours indirect / Que conjonction / Interrogative indirecte / Pour + Verbe infinitif / Quand / Qui relatif / Où relatif / Corrélatif
2 ans = 3 ans < 4 ans = 5 ans	Parce que / Gérondif et autres temporels / Que relatif

Tableau 4 : Evolution dans les énoncés des enfants du nombre de chacune des complexités selon le groupe d'âge (tests *post hoc* de Bonferroni)

Cette première série de résultats montre tout d'abord qu'il est justifié et indispensable de procéder à des analyses au-delà de l'âge de 3 ans puisque des évolutions dans le langage de l'enfant se poursuivent au-delà de 5 ans de façon significative. Ces évolutions ne sont pas uniformes pour toutes les catégories syntaxiques : il existe au fil du temps une progression des complétives, des relatives, des corrélatives et de certaines circonstancielles. Certaines constructions évoluent de façon significative à 4 ans puis se stabilisent, alors que d'autres évoluent de façon significative à 5 ans.

5 Evolution de la production des complexités syntaxiques en fonction du type de discours

En croisant les pourcentages de complexités dans chaque type de discours avec l'âge⁷ (graphique 2), il semblerait que la progression de la production de constructions complexes soit plus importante dans les narrations à partir de 5 ans que dans les conversations. Dans les conversations, la proportion générale de constructions complexes aurait tendance à stagner (on passe de 33 % à 4 ans à 36 % à 5 ans).



Graphique 2 : Constructions complexes dans les énoncés des enfants en fonction de l'âge et du type de discours

Dans les narrations, toutes les catégories de constructions complexes sont concernées par une évolution à l'âge de 5 ans (circonstancielle, complétives, relatives, et dans une moindre mesure clivées et autres constructions). Pour les conversations, l'évolution est faible au cours de cette période, elle n'est pas révélatrice d'une évolution, mais plutôt d'une stabilisation (tableau 5).

Catégories / groupe d'âge	3 ans		4 ans		5 ans	
	Narration	Conversation	Narration	Conversation	Narration	Conversation
Circonstancielle	4,56	9,44	11,27	16,13	26,52	16,87
Complétives	5,71	5,59	8,99	9,62	24,68	10,36
Relatives	2	2,11	1,85	2,84	10,7	3,57
Clivées	0,29	2,75	0,71	3,94	1,57	5,04
Autres	0,29	0,37	0	0,73	0,86	0,55

Tableau 5 : Répartition des catégories syntaxiques chez les enfants en fonction de l'âge et du type de discours

Si l'on s'appuie sur les pourcentages moyens d'occurrences dans chaque groupe d'âge et dans chaque type de discours (tableaux 6 et 7), l'analyse statistique valide cette évolution dans les narrations pour toutes les catégories syntaxiques : les différences entre les groupes d'âge sont significatives pour les complétives ($F(2, 104) = 19,94, p < 0,05$), les circonstancielle ($F(2, 104) = 26,95, p < 0,05$), les relatives ($F(2, 104) = 20,11, p < 0,05$), les clivées ($F(2, 104) = 4,27, p < 0,05$) et les autres constructions ($F(2, 104) = 3,4, p < 0,05$).

Les tests *post hoc* de Bonferroni (différenciation des groupes d'âges 2 à 2) montrent là encore que l'évolution la plus significative s'effectue à partir de 5 ans : ainsi, la moyenne des complexités produites par les enfants de 3 ans est équivalente à celle des enfants du groupe des 4 ans mais ces trois moyennes sont inférieures à celles des enfants du groupe des 5 ans (4,93). Il en est ainsi de toutes les catégories syntaxiques : 2 ans = 3 ans = 4 ans < 5 ans.

Groupe d'âge		Circonstancielles	Complétives	Relatives	Clivées	Autres
3 ans	Moyenne	0,13	0,17	0,06	0	0
	N	34	34	34	34	34
	Ecart-type	0,22	0,21	0,12	0,03	0,03
4 ans	Moyenne	0,29	0,23	0,05	0,02	0
	N	39	39	39	39	39
	Ecart-type	0,30	0,32	0,11	0,05	0
5 ans	Moyenne	0,78	0,72	0,31	0,04	0,02
	N	34	34	34	34	34
	Ecart-type	0,55	0,59	0,31	0,08	0,06

Tableau 6 : Evolution dans les narrations du nombre de complexités (regroupées en catégories) selon le groupe d'âge

A contrario, dans les conversations, si l'on s'appuie sur les pourcentages moyens d'occurrences dans chaque groupe d'âge (tableau 8), seules les circonstancielles différencient les groupes d'âge ($F(2, 72) = 3,26, p < 0,05$), et les tests *post hoc* de Bonferroni indiquent que leur évolution est significative à 4 ans : 3 ans < 4 ans = 5 ans.

Groupe d'âge		Circonstancielles	Complétives	Relatives	Clivées	Autres
3 ans	Moyenne	0,38	0,22	0,08	0,11	0,01
	N	25	25	25	25	25
	Ecart-type	0,42	0,37	0,15	0,14	0,03
4 ans	Moyenne	0,62	0,37	0,11	0,15	0,03
	N	26	26	26	26	26
	Ecart-type	0,36	0,32	0,10	0,15	0,07
5 ans	Moyenne	0,70	0,43	0,15	0,21	0,02
	N	24	24	24	24	24
	Ecart-type	0,59	0,35	0,21	0,26	0,049

Tableau 7 : Evolution dans les conversations du nombre de complexités (regroupées en catégories) selon le groupe d'âge

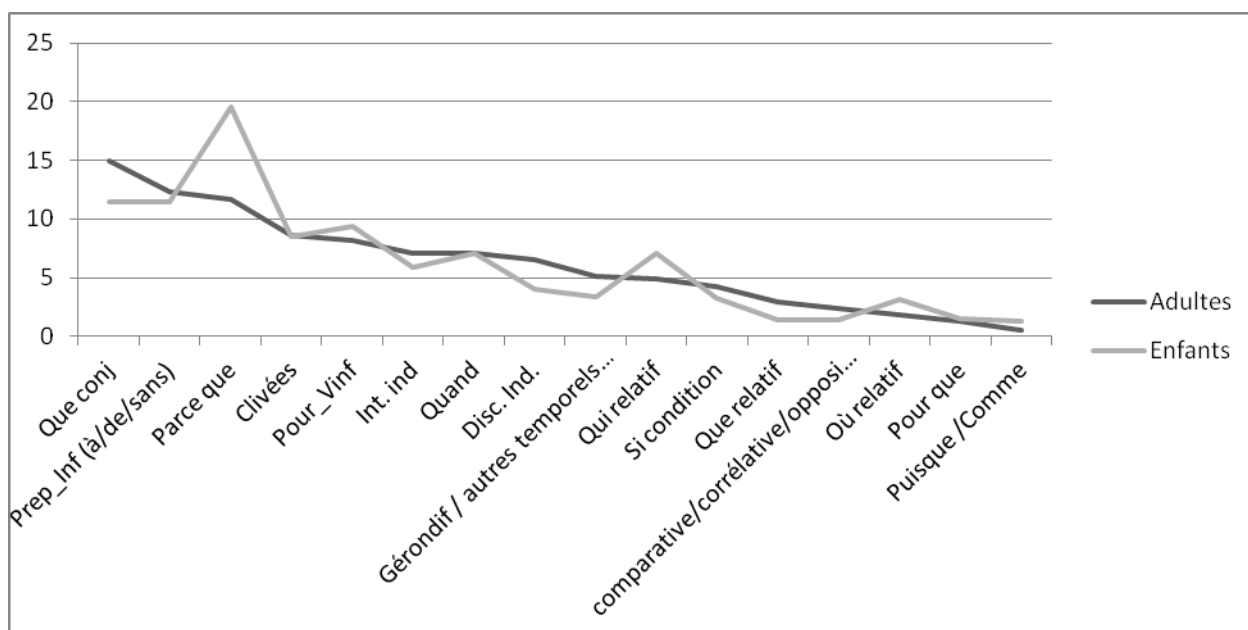
Cette deuxième série de résultats montre que, si l'on a bien une évolution selon l'âge, elle est à relativiser puisque, selon le type de discours, la production de constructions complexes diffère : elle augmente

significativement à 5 ans dans un autre type de discours que la conversation ordinaire. Comme nous l'avions déjà pressenti, la narration est un type de discours qui engendre dans le langage de l'enfant une plus grande densité de constructions complexes.

6 Etude comparative de la production des complexités syntaxiques chez les adultes et les enfants en fonction de l'âge et du type de discours

L'étude de l'acquisition de formes linguistiques ne doit pas faire l'impasse sur une autre variable : le langage adressé à l'enfant, qui peut expliquer des évolutions, en particulier du point de vue des reprises de complexités syntaxiques effectuées par l'enfant pour raconter ou expliquer. En comparant la répartition des constructions complexes selon leurs catégories syntaxiques, on remarque que les adultes et les enfants produisent des constructions dans des proportions similaires : majoritairement des circonstancielles et des complétives (79 % des constructions), dans une moindre proportion des relatives et des clivées (entre 8 et 12 %), et d'autres constructions de façon plus marginales. En revanche, les enfants produisent plus de circonstancielles (46 %) que de complétives (33 %), contrairement aux adultes pour lesquels la proportion est un peu plus équilibrée (38 % de circonstancielles et 41 % de complétives).

Si l'on compare plus en détail les complexités syntaxiques produites par les adultes et par les enfants, on constate une certaine homogénéité : les constructions peu nombreuses dans les énoncés des adultes le sont aussi dans les énoncés des enfants et réciproquement (graphique 3). La corrélation de Pearson entre les pourcentages (calculés au sein de chaque groupe) de chacune des constructions complexes employées par les adultes et par les enfants est de $r = 0,90$ ($p = 0,000$), valeur élevée, qui conforte l'idée que le langage adressé à l'enfant est en relation significative avec celui produit par l'enfant, du moins pour la verbalisation de ces complexités syntaxiques.



Graphique 3 : Complexités syntaxiques produites par les adultes et les enfants

En croisant ces données par âges et par types de discours (tableau 8), on constate que, dans les narrations, les corrélations augmentent avec l'âge des enfants : elles sont toutes significatives à $p = 0,00$, ce qui signifie que le nombre et le type de constructions complexes produits par les adultes et les enfants sont de plus en plus comparables lorsque l'âge des enfants augmente. En revanche, dans les conversations, la corrélation la plus élevée est à 3 ans : les enfants de 3 ans et les adultes qui interagissent avec eux produisent des proportions similaires de type de constructions complexes. Dans les autres groupes d'âge, les corrélations sont moins élevées et il n'y a pas d'augmentation régulière avec l'âge.

Corrélation de Pearson (N = 20)	3 ans	4 ans	5 ans
Narration	0,85	0,89	0,92
Conversation	0,92	0,75	0,73

Tableau 8 : Corrélation dans la production des complexités syntaxiques par les adultes et par les enfants en fonction de l'âge et du type de discours

En résumé, l'analyse tend à montrer que les narrations dialoguées entraînent dans les énoncés des enfants une plus grande production de constructions complexes, que cette évolution est marquée par une augmentation sensible à l'âge de 5 ans et qu'elle est corrélée à la production des mêmes complexités par l'adulte.

7 Conclusion

L'étude présentée ici a pour ambition de traiter globalement la question de la production des constructions complexes chez les enfants francophones de plus de trois ans, dans la voie tracée par la linguistique de corpus, c'est-à-dire en élargissant l'analyse à un nombre important de corpus oraux pouvant être, en partie, traités par des outils informatiques. L'étude s'est également voulue développementale, de façon à dégager des généralisations quant au processus d'acquisition du langage et au rôle de l'*input*. Si les résultats demanderaient bien sûr à être vérifiés sur un corpus plus homogène dans lequel les différentes variables seraient davantage équilibrées (nombre d'enfants et nombre de mots par catégories d'âge, temps de latence entre chaque enregistrement, nombre et durée des corpus...), les principales avancées de ce travail sont les suivantes :

- La production des constructions complexes augmente au fil du temps, c'est le cas des complétives, des relatives, des corrélatives et de certaines circonstancielles. Et on observe pour certaines constructions subordonnées une évolution significative à 4 ou 5 ans.
- Le type de discours joue un rôle majeur dans cette évolution : la narration dialoguée - c'est-à-dire le récit par l'enfant d'une histoire qui lui a été lue, avec le support des illustrations, dans un cadre interactif avec certaines formes d'étayage – permet une augmentation significative de la production de constructions complexes à 5 ans, ce qui n'est pas le cas dans les conversations ordinaires. La narration est donc un genre de discours plus propice à un développement langagier que la conversation. Nous pouvons interpréter ce résultat de la façon suivante : le discours de type narratif permet à l'enfant d'explorer d'autres fonctionnements linguistiques et favoriserait l'insertion des complexités syntaxiques dans de nouveaux cotextes de production. En ce sens, la narration serait vectrice simultanément d'un apprentissage textuel et linguistique : elle permettrait de tester de nouvelles acquisitions linguistiques organisant le discours tandis que la conversation privilégierait davantage la dimension communicationnelle.

- L'*input* joue un rôle fondamental dans le développement de la syntaxe au-delà de 3 ans : l'analyse montre de fortes corrélations entre les occurrences de constructions syntaxiques complexes dans les énoncés de l'adulte et de l'enfant, en particulier dans les narrations dialoguées. Le langage adressé par l'adulte à l'enfant est davantage vecteur d'évolution sur le plan grammatical dans les narrations que dans les conversations. Une étude quantitative et qualitative plus détaillée des reprises chez les enfants en fonction de l'âge et du type de discours (Canut, 2012 et 2013b) conforte ce résultat d'une syntaxe plus complexe dans le discours narratif de l'enfant en lien avec le langage qui lui est adressé. Ce résultat pour les enfants francophones est également à mettre en rapport avec l'étude anglo-saxonne de Cameron-Faulkner & Noble (2013) établissant que le *child directed speech* est plus complexe sur le plan grammatical en narration que l'*input* en conversation puisque les enfants mobilisent des ressources syntaxiques qui font écho à celles qu'ils entendent quand les adultes leur racontent des histoires.

Références bibliographiques

- André, V. & Canut, E. (2010). Mise à disposition de corpus oraux interactifs : le projet TCOF (Traitement de Corpus Oraux en Français). *Pratiques*, 147-148, 35-51.
- Blanche-Benveniste, C. (2010). *Le français. Usages de la langue parlée*. Leuven-Paris : Peeters.
- Bloom, L. (1991). *Language development from two to three*. Cambridge : Cambridge University Press.
- Bloom, L., Lahey, M., Hood, L., Lifter, K. & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child Language*, 7, 235-261.
- Bloom, L. & Capatides, J. B. (1987). Sources of meaning in the acquisition of complex syntax: The simple case of causality. *Journal of Experimental Child Psychology*, 43, 112-128.
- Bloom, L., Rispoli, M., Gartner, B. & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language*, 16, 101-120.
- Brown, R. (1973). *A first language: The early stages*. Cambridge : Harvard University Press.
- Cameron-Faulkner T. & Noble C. (2013). A comparison of book text and Child Directed Speech. *First Language*, 33-3, 268-279.
- Canut, E. (2001). *Évolution de la syntaxe et de l'ancrage énonciatif dans des narrations d'enfants de moins de 6 ans*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.
- Canut, E. (2012). *L'acquisition des constructions syntaxiques complexes chez l'enfant. Un processus interactionnel*. Habilitation à Diriger des Recherches, non publiée, Université Stendhal - Grenoble 3.
- Canut, E. (2013a). De l'apprentissage et de l'usage de la "subordination" chez l'enfant de moins de six ans. In Debaisieux, J.-M. (éd.), *Analyses linguistiques sur corpus : Subordination et insubordination en français*, Paris : éditions Lavoisier, 141-181.
- Canut, E. (2013b). Reprises et fonctionnement syntaxique : les fondements d'une compétence narrative chez l'enfant entre 3 et 6 ans. In *A.N.A.E.*, 124, 254-260.
- Clark, E. (1998). Lexique et syntaxe dans l'acquisition du français. *Langue Française*, 118, 49-60.

- Clark, E. (2003). *First language acquisition*. Cambridge : Cambridge University Press.
- Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge : Cambridge University Press.
- Eisenberg, S. & Cairns, H. S. (1994). The development of infinitives from three to five. *Journal of Child Language*, 21, 713-734.
- François, D., François, F., Sabeau-Jouannet, É. & Sourdot, M. (1977). *La syntaxe de l'enfant avant 5 ans*. Paris : Larousse.
- Hirsh-Pasek, K. & Golinkoff, R. M. (1996). *The Origins of Grammar: Evidence from Early Language Comprehension*. Cambridge : M.I.T. Press.
- Kail, M. (2000). Acquisition syntaxique et diversité linguistique. In Kail, M. & Fayol, M. (éds), *L'acquisition du langage. Le langage en développement au-delà de trois ans*, Paris : P.U.F, 9-44.
- 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. (1975). Problématique de l'acquisition de la syntaxe chez l'enfant. *Langue Française*, 27, 14-23.
- Lentin, L. & al (1984). *Recherches sur l'acquisition du langage*. Vol. 1. Paris : Presses de la Sorbonne Nouvelle - Paris 3.
- Lentin, L. & al (1988). *Recherches sur l'acquisition du langage*, Vol. 2. Paris : Presses de la Sorbonne Nouvelle - Paris 3.
- Le Normand, M.T., Moreno-Torres, I., Parisse, C., & Dellatolas, G. (2013). How do children develop early grammar and build multiword utterances ? A corpus study of French children aged 2 to 4. *Child Development*, 84(2), 647-661.
- Moreau, M.-L. & Richelle, M. (1981). *L'acquisition du langage*. Bruxelles : Mardaga.
- Peterson, C. & Mac Cabe, A. (1987). The connective 'and': Do older children use it less as they learn other connectives ? *Journal of Child Language*, 14, 375-381.
- Sekali, M. (2012). The emergence of complex sentences in a French child's language from 0;10 to 4;01: causal adverbial clauses and the concertina effect. *French Language Studies*, 22(1), 115-141.
- Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Cambridge : Harvard University Press.
- Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: how much is enough ? *Journal of Child Language*, 31-1, 101-121.

¹ Base de données rattachée à l'UMR 7118 CNRS ATILF. Les corpus sont en grande partie visibles, en accès libre et gratuit, sur le site du CNRTL : www.cnrtl.fr/corpus. Le choix de travailler sur un grand nombre de données attestées n'est pas le fruit du hasard, il est issu d'une longue réflexion et d'un long travail d'archivage, de transcription et de vérification, qui a démarré en septembre 2005. Pour une présentation détaillée du projet TCOF voir André & Canut, 2010.

² Ce choix de corpus est en adéquation avec les objectifs de l'étude : l'oralisation des textes écrits et les reformulations de l'adulte procurent à l'enfant une expérience langagière diversifiée et permettent d'appréhender l'entière variété de productions dont sont capables les locuteurs (Blanche-Benveniste, 2010).

³ Ce logiciel, libre et gratuit, a été développé (en Java) par Nicolas Guth et Vincent Poutissou ; sa conception a été supervisée par O. Mella, C. Benzitoun et V. André.

⁴ Cette pondération ramène le nombre de corpus d'enfants de 2 ans à seulement 24 au total (narration et conversation) alors que dans les autres groupes d'âge, il y en a au total une soixantaine. Les données du groupe des 2 ans sont donc sous-estimées, mais nous n'avons pas jugé utile de les pondérer jusqu'à 60 car le but premier de l'analyse est davantage de se focaliser sur les enfants au-delà de 3 ans. Pour valider cette pondération, nous avons doublé le nombre d'occurrences de complexités syntaxiques présents dans les énoncés des adultes et des enfants dans les corpus de conversation du groupe des 2 ans et dans les corpus de narration du groupe des 3 ans.

⁵ Nous n'avons pas procédé au calcul de la longueur moyenne d'énoncés (MLU) car ce n'est plus un indice suffisant au regard de l'évolution syntaxique après l'âge de 3-4 ans. L'analyse statistique a été supervisée par C. Bocéréan.

⁶ Le pourcentage de chaque catégorie de constructions complexes est calculé par rapport au nombre total de constructions complexes chez les enfants, soit 1792.

⁷ La catégorie du groupe des enfants de 2 ans est exclue de cette analyse croisée âge/type de discours puisque nos données à cette période d'âge ne comportent que des conversations.