

Des données représentatives... de quoi en acquisition du langage? Constitution de données à observer et objectifs d'analyse

Emmanuelle Canut, Martine Vertalier

► To cite this version:

Emmanuelle Canut, Martine Vertalier. Des données représentatives... de quoi en acquisition du langage? Constitution de données à observer et objectifs d'analyse. *Verbum* (Presses Universitaires de Nancy), Université de Nancy II, 2010, pp.150-154. hal-00523406

HAL Id: hal-00523406

<https://hal.archives-ouvertes.fr/hal-00523406>

Submitted on 21 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.

DES DONNÉES REPRÉSENTATIVES... DE QUOI, EN ACQUISITION DU LANGAGE ? CONSTITUTION DE DONNÉES A OBSERVER ET OBJECTIFS D'ANALYSE

Emmanuelle Canut (Nancy Université - ATILF)
Martine Vertalier (Université Sorbonne nouvelle - Paris 3)

RÉSUMÉ

Nous aborderons la question de la constitution et de l'analyse des données dans le domaine de l'acquisition du langage en tentant de faire ressortir les différences entre une approche psycholinguistique et une approche linguistique de l'étude du langage de l'enfant de moins de six ans. Un détour par l'histoire des recherches en acquisition du langage et quelques notes sur un numéro récent de la revue Journal of Child Language nous permettront de comparer les modalités et les objectifs des travaux en psycholinguistique développementale à ceux de l'approche qualitative des recherches en linguistique de l'acquisition, du point de vue tant de la conception du langage que du recueil, de la sélection et de l'analyse des données.

ABSTRACT

The aim of this paper is to compare data construction and analysis from the point of view of developmental psycholinguistics and from the point of view of the qualitative approach of the linguistics of language acquisition, with reference to the history of language acquisition studies and to recent papers in the Journal of Child Language.

1. L'HÉTÉROGÉNÉITÉ DES RECHERCHES EN ACQUISITION DU LANGAGE : PETIT RETOUR EN ARRIÈRE

Le domaine des études en acquisition du langage présente une grande hétérogénéité, du fait de la diversité des approches théoriques, des méthodes mais aussi des grandes disciplines qui les inspirent : psychologie ou linguistique, et plus récemment psycholinguistique. Le foisonnement des recherches sur des phénomènes à caractère limité, dans des contextes méthodologiques différents, rend souvent difficile la confrontation des résultats. Les études tendent à s'organiser selon deux grands types d'approches, issues de l'histoire de la constitution de ce qu'on peut, peut-être, appeler aujourd'hui une discipline : les approches de type qualitatif et les approches de type quantitatif. David Ingram (1989), dans une des rares tentatives de rendre compte de l'histoire et de l'ampleur du domaine, a présenté l'histoire des études en acquisition du langage comme marquée par

trois grandes époques, de 1876 à nos jours.

1.1. 1876 – 1926

L'époque des monographies parentales, *baby biographies* ou *diary studies*, journaux de bord des pionniers des études en acquisition du langage, voit paraître des recueils longitudinaux constitués par des amateurs éclairés, philosophes, scientifiques, puis essentiellement par des psychologues, observant le langage de leurs propres enfants. La visée de ces notes était d'abord de saisir le génie créatif propre à chaque enfant, en accédant à l'évolution de ses affects, à son développement psychologique, le langage ne représentant qu'un aspect du développement général de l'enfant.

Les données sont alors constituées de fragments plus ou moins étendus du langage de l'enfant, captés « à la volée », accompagnés de notes sur le comportement de l'enfant, en l'absence de toute ressource technologique de recueil visuel ou sonore. On a pu taxer ces données de partialité, d'incomplétude, et pourtant elles étonnent encore aujourd'hui par leur minutie et leur volonté d'exhaustivité. Ingram fait remonter le début de cette période à la publication des « Notes sur l'acquisition du langage chez les enfants et dans l'espèce humaine » d'Hippolyte Taine, dans le numéro 1 de la *Revue Philosophique*, en 1876. Une partie de ces notes sur le développement du langage de sa fille, de la naissance à 3 ans, ont été traduites en anglais l'année suivante dans le numéro 1 de *Mind*.

Le numéro 2 de *Mind* publie des extraits du journal d'observations minutieuses tenu par Charles Darwin 37 ans auparavant, sur les actes instinctifs, les sensations, les affects et les « moyens de communication » d'un de ses fils, relevés avec notation du contexte, dès la première semaine de la vie, sur une période d'un an. Dans « A biographical sketch of an infant », Darwin rapporte, par exemple, l'invention de *mum*, à un an, pour désigner de la nourriture, *shu-mum* pour désigner le sucre et *black shu-mum* pour la réglisse. La question de la représentativité des données ne se pose pas et Darwin affirme : « Je suis persuadé, d'après ce que j'ai pu observer auprès de mes propres enfants, que la période de développement des différentes facultés révélera des différences considérables d'un enfant à l'autre » (1877, 285).

Parmi les nombreuses observations d'enfants du 19e siècle, on peut mentionner, en France, le mémoire de Jean Itard écrit en 1801, sur Victor « le jeune sauvage » de l'Aveyron et le journal de Michel Bréal *Mémoires*

de mon fils Edouard Auguste Bréal (1869, voir Ottavi 2007).

Cette période a vu la production d'observations des plus minutieuses, détaillées, de la part de psychologues, comme celles de Wilhelm Preyer (*Kinder Seele*, 1882) sur le développement général de son fils Axel de 0 à 3 ans. On y trouve la notation de la production de deux mots seulement, depuis 1 an 1 mois : « *atta* » pour « aller » et « *heiss* » pour « chaud », l'identification de réactions appropriées à des questions comme « où est papa ? », une liste de mots compris régulièrement par l'enfant (« oreille, chaussure, chaise, épaule... »), et de verbes auxquels l'enfant obéit : « cours, touse, souffle, apporte, donne, fait un baiser... ». *Die Kindersprache* de Clara et Wilhelm Stern (1907), psychologues, est le premier grand ouvrage consacré au langage de l'enfant et la première tentative de dégager des étapes d'apprentissage.

Même si 1926 voit s'ouvrir une autre ère, celle des recherches de type comparatif, les monographies parentales ne disparaissent pas. Elles sont fondées sur des prises de notes et, plus tard, sur des enregistrements. Ainsi la double monographie d'Antoine Grégoire, linguiste, (*L'apprentissage du langage*, 1937, 1947) sur le langage de ses deux fils, de la première année à trois ans et au-delà, marque une étape dans la volonté de systématisation de l'observation. Les deux volumes de près de 800 pages contiennent des notations de tous les aspects du langage : développement phonétique et phonologique, mots, phrases, catégories du discours, temps, avec des remarques sur le contexte du « français wallonisé » de l'entourage. Dans son avant-propos, Antoine Grégoire fait référence aux injonctions de Marcel Cohen (1925) quant à la nécessité d'une collecte du « langage total » de l'enfant. M. Cohen écrit : « Si l'on en conclut que seul un ménage de linguistes élevant ses enfants et qui n'est distrait par aucune autre occupation peut réunir toutes ces conditions, on ne s'étonnera pas que les descriptions complètes de langages d'enfants soient si rares jusqu'à présent et on n'espérera pas trop les voir se multiplier » (1993 [1925], 34).

Antoine Grégoire fait passer le langage au premier plan, comme moyen d'accès au développement psychologique de l'enfant :

« L'âme s'éveille de plus en plus, à partir de la naissance ; personne n'en doutera ; mais reconnaissons que l'un des moyens efficaces de nous en rendre compte nous est fourni par la parole. Il vaut la peine de noter les manifestations de celle-ci, le plus fidèlement qu'il est possible, et d'en chercher l'interprétation la plus vraisemblable. C'est là une tâche qui demande pour le moins la qualité fondamentale de tout travail scientifique, à savoir l'exactitude poussée très loin, tant dans la notation des phénomènes linguistiques que dans l'établissement des

conditions où ils se produisent. » (Grégoire, 1937, 6)

Pour les deux premières années, l'auteur recourt à la transcription phonétique des éléments recueillis, selon le système de l'Association internationale de phonétique. A partir de la troisième année, Antoine Grégoire précise :

« la prononciation des enfants tend à se rapprocher de plus en plus de la prononciation normale, et l'on peut d'ordinaire conserver aux mots qu'ils disent l'orthographe traditionnelle, en donnant à celle-ci la valeur phonétique qui lui revient. Ainsi l'*e* final des mots « charrette », « vase » est réellement muet ; [...]. Par contre, lorsque l'enfant remplacera le *ch* final de vache par une sifflante se rapprochant de *s*, on écrira *vasse*, mais pas : *vase* (= *vaze*), ni : *vas* (= *va*), qui prêteraient à confusion. » (1947, 8)

Bénéficiant des progrès de la technique des années 1960, Ruth Weir, enregistre les monologues de son fils pendant la période d'endormissement dans son berceau, sur deux mois, entre 2;2 et 2;4. *Language in the Crib*, 1962, donne lieu à une analyse phonologique d'abord mais aussi grammaticale, et discursive. La transcription phonétique complète du corpus « *in rough phonetics* », (1962, 147) – en phonétique approximative, est fournie en fin d'ouvrage.

1.2. 1926 - 1957

La seconde période historique des études en acquisition du langage est marquée par l'explosion des études de type expérimental, quantitatives, comparatives, transversales (*cross-sectional*), fondées sur le recueil d'échantillons de langage de nombreux sujets d'âges différents (124 dans l'étude de Madorah Smith (1926), 430 enfants entre 3 et 8 ans, dans celle de Mildred Templin (1957). Ces études sont inspirées par l'avènement du behaviorisme, avec les travaux de Pavlov sur les réflexes conditionnels (1924), et l'ouvrage de John B. Watson (1914) : *Behavior : an introduction to comparative psychology*, puis *Verbal Behavior* (le comportement verbal) de Skinner (1957). Le langage est traité comme un comportement parmi d'autres, testé par des séries d'épreuves standardisées, portant sur l'articulation, testée par la répétition de mots isolés ou de logatomes, la discrimination de paires de syllabes, le développement du vocabulaire ou de la phrase (étudié sur un échantillon de 50 phrases pour chaque enfant, produites à propos d'images ou de jouets et transcrites sur-le-champ). Les études cherchent à mesurer le comportement observable, à déterminer la nature du comportement « normal » et débouchent sur la production de normes de comportement, qui servent encore aujourd'hui de normes de

référence pour l'évaluation du langage des enfants : les résultats de l'étude de Mildred Templin servent de norme du développement articulatoire et ceux de l'étude de Madorah Smith de normes du développement du vocabulaire, pour l'anglais.

Dans le *Manuel de psychologie de l'enfant* de Leonard Carmichael (1946, tr. fr. 1952), Dorothea Mac Carthy traite « Le développement du langage chez l'enfant ». Elle combine les résultats de 8 des principales études réalisées entre les années 20 et les années 40, pour tenter de constituer une séquence de développement du langage à partir des 126 aspects du développement linguistique relevés dans ces études. Deux des études sur lesquelles elle s'appuie sont entièrement longitudinales et les six autres sont essentiellement transversales. Cette synthèse permet d'observer quels éléments sont recherchés, repérés et datés dans les différentes études : productions, réactions, répétitions, cri, imitation des sons, syllabes, apparition du premier mot (avec désaccord sur la date dans les deux seules études qui le mentionnent), catégories du discours : pronoms, prépositions, phrases, propositions...

1.3. 1957 et au-delà

La troisième période ouvre l'époque des recueils « d'échantillons de langage longitudinaux ». Portant sur un nombre d'enfants supérieur à un mais beaucoup plus restreint que dans les études quantitatives (soit trois à cinq), ces méthodes visent à recueillir, à intervalles réguliers prédéterminés et pour une durée prédéfinie, des échantillons de langage dits « représentatifs ». De plus, s'appuyant sur les étapes d'acquisition déterminées par les études quantitatives, les chercheurs, désireux de ne pas manquer l'émergence d'éléments dans les tranches d'âge sensibles, n'hésitent pas à solliciter les parents pour qu'ils constituent un journal parental, où ils ont pour consigne de repérer, par exemple, les énoncés à un mot, l'apparition d'énoncés à deux mots, etc. (Martin Braine, 1963).

A la différence de ce qui se passait pour les monographies de la première période, les sujets des études longitudinales ne sont pas les enfants de l'expérimentateur, ils sont sélectionnés en fonction du stade de développement où sont censés se produire les phénomènes à étudier. Les enfants sont enregistrés au magnétophone, dialoguant avec l'expérimentateur ou avec leur mère. Nous évoquerons, pour cette période, l'étude de Roger Brown (1973), qui a donné lieu à des normes systématiquement utilisées dans la plupart des études actuelles. Les trois enfants de l'étude ont été

sélectionnés parmi une trentaine parce qu'ils commençaient à produire des énoncés à plusieurs mots. Les principales données sont des transcriptions du langage spontané, non contrôlé, de l'enfant, en conversation avec sa mère, à la maison, d'au moins deux heures par mois. Les transcriptions sont du niveau du morphème, donc généralement en orthographe « normale » de l'anglais. Les trois enfants n'avaient pas le même âge au début de l'expérimentation : Eve, 1;6, Adam et Sarah 2 ;3. Du fait de la variation dans le rythme d'acquisition des enfants, l'étalonnage ne s'est pas fait en fonction de l'âge mais en fonction des longueurs d'énoncés. Les calculs ont débouché sur l'identification de 5 stades de développement du langage, correspondant aux longueurs moyennes d'énoncé (MLU), et longueur de l'énoncé le plus long, calculées en morphèmes, selon des règles de sélection d'échantillons au long des corpus. Aujourd'hui ces stades servent au classement préalable des sujets des études expérimentales, non par groupes d'âge mais en fonction des longueurs moyennes d'énoncés.

2. AVÈNEMENT DE LA PSYCHOLINGUISTIQUE : TENDANCES ACTUELLES

D'un point de vue théorique, on peut distinguer globalement deux grandes tendances. D'une part, les études de type psycholinguistique abordent le langage comme un comportement mesurable et s'appuient directement sur les données, sans afficher une position théorique de départ, bien qu'on puisse inférer une théorie implicite du langage d'après la démarche suivie. D'autre part, les approches de type linguistique envisagent le langage comme une fonction humaine, soit plus fondamentalement biologique (telles les approches générativistes), soit à la fois biologique et sociale se développant en contexte (telles les approches socio-interactionnistes).

Les études de type linguistique se préoccupent de déceler des processus d'acquisition, tandis que les études de type psycholinguistique ont généralement pour objet de dégager des étapes et des normes de développement du langage de l'enfant. Les études d'orientation générativiste s'appuient sur des modèles de compétence et ont souvent recours à des énoncés fictifs, produits de l'introspection et de l'intuition du linguiste. Il leur arrive aussi de faire appel à des données recueillies selon les méthodes issues de la psychologie et utilisées en psycholinguistique : épreuves de compréhension et de production, ce qui brouille un peu les pistes... (Christophe, 2002)

On peut tenter de rendre compte des différences d'un point de vue épistémologique : les recherches en psycholinguistique développementale (dont les approches générativistes) s'inscrivent dans le champ des sciences cognitives « qui tente d'élucider par l'expérimentation, la modélisation et l'usage de technologies de pointe, le “Mystère de l'Esprit” dans ses rapports avec la Matière : le cerveau, le corps et l'ordinateur. » (Houdé, Kayser, Koenig, Proust, Rastier (1998, 1). L'objectif sera donc plutôt de rechercher des mécanismes généraux de traitement de l'information par l'enfant, autrement dit ce qui est acquis à un moment donné en fonction des capacités cognitives de l'enfant. Les procédures expérimentales (de la succion non nutritive au rappel de texte en passant par des tests variés) et les études transversales semblent la méthodologie la plus adéquate pour atteindre cet objectif. Quand le recueil de données est “authentique” (enregistrements de dialogues), seul le langage de l'enfant retient l'intérêt des chercheurs (Kail & Fayol, 2000).

A l'inverse, la prise en compte du langage adressé à l'enfant est fondamentale pour les études s'inscrivant dans une approche socio-interactionniste : les interventions de l'entourage sont considérées comme jouant un rôle décisif dans le développement des capacités cognitivo-langagières de l'enfant, dans la mesure où “toute activité mentale est culturellement située, c'est-à-dire qu'elle n'est jamais isolée ou menée sans assistance” (Bruner, 1991). On retrouve toutefois dans ce cadre deux façons d'analyser les données :

- une analyse quantitative (études anglo-saxonnes sur l'*input*) dont l'objectif est de rechercher des étapes de développement à partir d'études de corrélation entre langage de l'adulte et langage de l'enfant (Gallaway & Richards, 1994) ;

- une analyse qualitative (études francophones en linguistique de l'acquisition, entre autres) dont l'objectif est de déterminer des processus dialogiques d'acquisition et des modalités d'interaction contribuant au développement cognitivo-langagier de l'enfant (Lentin, 2009).

2.1. Une illustration : les données dans *Journal of Child Language*

Il y a une dizaine d'années (Karnoouh-Vertalier, 1997), nous nous étions livrées à un exercice, un peu masochiste, qui consistait à inventorier les méthodes de recueil et traitement des données dans la principale revue de référence en acquisition du langage : *Journal of Child Language* (23-2, 1996). Nous avons refait l'exercice avec le n° 36-2 (2009). Les articles sont tous construits sur le même plan, fourni par la revue. Généralement une

introduction fait l'état de la question en termes de résultats des études antérieures sur le même sujet, sans pour autant poser une théorie du langage sur laquelle fonder la démarche. Puis l'exposé de la méthode : recrutement, sélection d'échantillons d'enfants, grâce à divers tests, comme le MacArthur-Bates Communicative Development Inventory, ou les stades de développement du langage selon le MLU de Brown (par exemple 10 enfants de 4, 5, 6, 7 ans et 10 adultes pour comparaison) ; ou bien recours à des corpus déjà constitués archivés dans la base de données de CHILDES¹. La méthode de transcription est fonction de l'objectif de l'analyse, mais, même si le langage de l'enfant est recueilli en dialogue avec un adulte, généralement la mère, parfois seul le langage de l'enfant est transcrit (voir par exemple Taelman, Durieux, Gillis, 2009, 326). Vient ensuite la mention des supports, ou stimuli, des images par exemple, sur lesquelles l'enfant est censé produire une phrase.

Puis vient le protocole. Certaines études se font au domicile de l'enfant, d'autres en laboratoire. Un exemple : dans "Maternal control strategies, maternal language usage and children's language usage at two years" (Taylor, Donovan, Miles & Leavitt, 2009), l'étude concerne 61 dyades mère-enfant, soumises à un "*clean-up task*" en laboratoire. Chaque mère est censée se comporter comme elle le ferait "typiquement" à la maison, pour faire ranger à l'enfant des jouets avec lesquels il a joué pendant qu'elle remplissait un questionnaire.

Les conventions de transcription sont souvent celles de CHILDES. Vient enfin la sélection, ou réduction des données et le codage pour l'analyse et les mesures du langage de l'enfant et/ou de l'adulte. La plupart du temps, seules les données correspondant aux catégories bien identifiées de la grammaire sont retenues comme "correctes". Tout ce qui pourrait relever de tâtonnements d'acquisition est éliminé. Les éléments étudiés sont toujours extraits du contexte de production. Dans l'ensemble, les études sont de plus en plus clairement affichées comme "comportementales".

Ce qui est frappant, c'est que les données langagières étudiées ne sont pas visibles : les articles contiennent des tableaux de statistiques, des courbes d'évolution concernant les occurrences relevées dans le corpus, sans qu'elles soient citées. Il s'agit donc de résultats chiffrés, mais l'absence de données ne permet pas de vérifier la manière dont elles ont été analysées. Tout ce qui est possible, comme le veut la tradition, c'est que l'expérimentation puisse être reproduite ("falsifiée"), pour confronter les résultats eux-mêmes mais

1 <http://childes.psy.cmu.edu>

pas les analyses des éléments étudiés.

Quant au langage de l'adulte, il n'est souvent qu'un prétexte au recueil du langage de l'enfant. Quand il est inclus dans l'étude, il est généralement considéré comme "*input*", analysé séparément du langage de l'enfant et comparé au langage de l'enfant à l'aide de recherche de corrélations entre certains aspects du langage de chaque locuteur.

2.2. Prise en compte du langage de l'adulte : une évolution ?

Dans la conclusion de l'une de ces études, sur l'acquisition des *clusters* en français, on peut déceler l'amorce d'une nouvelle évolution, avec une timide évocation de la prise en considération du langage de l'entourage. L'article (Demuth, McCullough, 2009) contient une seule allusion, fugitive, au langage de l'entourage : "*the Lyon dialect examined in this corpus contains few word-final schwas and we found few in children's productions.*" Cependant la conclusion laisse entrevoir la nécessité d'examiner l'*input* entendu par les enfants et l'impact qu'il pourrait avoir sur le développement des groupes consonantiques morphémiques ou non morphémiques. L'article se termine par la remarque suivante :

"The field of language acquisition is only now beginning to acknowledge the importance of understanding the nature of the input learners hear, and the potential impact this has on the process of language acquisition. A fuller investigation of these issues, including a closer analysis of both adult and child productions at the phonetics/phonology interface, is long overdue" (446)².

Dans leur article sur la dimension des échantillons de langage d'enfants à collecter, Michael Tomasello & Daniel Stahl (2004) évoquent l'évolution des méthodes, depuis les journaux parentaux jusqu'aux enregistrements audio ou vidéo d'interactions spontanées d'enfants avec un parent ou un autre interlocuteur pratiqués aujourd'hui. Ils soulèvent la question de la fréquence, de la durée et du nombre d'enfants des échantillons de langage. Ils signalent comme nouvelle la préoccupation des chercheurs pour l'usage, le traitement et l'apprentissage (*usage, processing and learning*) et la nécessité d'évaluer la fréquence et la date d'émergence des

² "Le champ de l'acquisition du langage ne fait que commencer à reconnaître l'importance qu'il y aurait à comprendre la nature de l'*input* qu'entendent les apprenants, et l'impact potentiel que cela pourrait avoir sur le processus d'acquisition du langage. Des recherches plus complètes sur ces questions, ne se sont fait que trop attendre, y compris une analyse plus précise des productions de l'adulte comme de celles de l'enfant à l'interface phonétique/phonologie." (Notre traduction).

phénomènes (118), ce qui confirme que jusqu'ici la psycholinguistique se préoccupait plus de décrire le langage de l'enfant à différents stades de développement, que d'identifier des processus d'apprentissage. Dans leur conclusion, ils font référence au rôle des corpus, en particulier pour la construction de grammaires "de base", comme la *Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad & Finegan, 1999) et ils annoncent (en 2004, donc) des temps nouveaux pour la recherche en linguistique grâce aux corpus : "The coming decades in linguistics in general will almost certainly be dominated by the analysis of corpora" (120).

Dans ce "nouveau" contexte, mais pas si nouveau somme toute, il semble que la notion d'échantillon de langage spontané, telle qu'elle est pratiquée dans les approches psycholinguistiques de l'acquisition du langage, soit susceptible d'être ébranlée, à la lumière de la linguistique de corpus, qui englobe la réflexion déjà ancienne, menée depuis plus de vingt ans, sur les rapports de l'oral et de l'écrit, elle-même éclairée par la notion de genre de discours développée par Bakhtine. Comme l'a écrit Claire Blanche-Benveniste déjà en 1997 :

"La linguistique contemporaine a proposé des classements [des usages] qui dépassent l'opposition trop rigide entre les deux seuls pôles de l'oral et de l'écrit. Il n'y aurait pas une opposition tranchée mais un continuum de pratiques différentes de la langue, tant pas écrite que par oral. En analysant certaines caractéristiques, on peut classer ces pratiques en autant de genres." (35-36).

Une approche linguistique de l'acquisition du langage a tout à gagner à s'appuyer sur ces réflexions.

3. LA LINGUISTIQUE DE L'ACQUISITION : UNE PRISE EN COMPTE DES PRATIQUES D'INTERACTION LANGAGIÈRE

Dans la recherche des premières modalités d'apprentissage du langage, les travaux de la linguistique de l'acquisition se fixent comme objectif principal l'évolution de l'organisation sémantico-syntaxique du langage de l'enfant en lien avec le langage qui lui est adressé, ou dit autrement, les caractéristiques de l'interaction favorables au développement langagier (Canut & Vertalier 2009).

Pour atteindre cet objectif, les chercheurs recueillent des corpus longitudinaux d'interaction entre adulte et enfant âgés de 2 à 7 ans, dans des situations de vie quotidienne ou des situations d'apprentissage (recueil semi-expérimental), ce sont soit des conversations, soit des co-narrations avec le

support de livres illustrés.

L'étude de ces corpus consiste en une analyse qualitative des dialogues adulte-enfant avec une focalisation sur l'articulation du discours c'est-à-dire sur les formes linguistiques en lien avec le développement cognitif. Pour ce faire, nous repérons simultanément :

- une évolution générale du fonctionnement syntaxique. D'abord l'évolution de la longueur d'énoncé, étudiée à l'aide d'une méthodologie statistique classique (médiane, quartiles, écart type). Ensuite l'évolution des catégories syntaxiques d'énoncés, quand l'enfant commence à produire des énoncés contenant plusieurs constructions syntaxiques, simples et complexes. Un répertoire de constructions, les "Introducteurs de Complexité" repérés expérimentalement (Lentin 1971) et une grille d'analyse spécifique (en catégories syntaxiques d'énoncés) permettent de tenter d'apprécier dans un même corpus ou d'un corpus à l'autre la diversité des constructions, leur profondeur syntaxique et l'évolution de leur combinatoire (Lentin et al. 1988).

- des phénomènes de reprises et de reformulations. La grille d'analyse en catégories syntaxiques d'énoncés a l'avantage de permettre de visualiser dans les énoncés de l'enfant les constructions complexes énoncées par l'adulte et reprises immédiatement ou en différé par l'enfant et inversement de visualiser dans les énoncés de l'adulte les reformulations de constructions en cours d'appropriation chez l'enfant.

L'analyse de l'extrait de corpus d'un enfant (E) âgé de 6 ans 3 mois et 18 jours, racontant avec l'adulte l'histoire de "La voiture téléguidée de Vincent"³ permet d'illustrer cette méthodologie.

Extrait du dialogue

E8- i(l) dit à maman que + euh ma mes piles i(ls) sont i(ls) marchent pas du tout

A8- oui il dit que les piles ne marchent pas

E9- i(l) dit à maman que les piles i(ls) marchent pas

A9- très bien

E10- e(lle) dit que voilà ! J(e) vais je vais appuyer trop fort

Les énoncés sont répertoriés dans la grille d'analyse en catégories syntaxiques ci-après.

3 LENTIN L., BERTIN M, ALLAIN G. (1988), *Collection Grenadine*, Illustrations : CRAPON A., Paris, Fleurus-enfants (épuisé, réédition en cours, AsFoReL 2009).

Numéro de l'énoncé E = Enfant A = Adulte	Enoncés comportant une construction simple		Enoncés comportant des constructions simples multiples (juxtaposées ou coordonnées)		Enoncés comportant des constructions complexes (voir liste des IC) + : juxtaposition / : emboîtement inc : cstrn incomplète		Enoncés comportant des essais (voir liste IC)
	E	A	E	A	E	A	E
E8							<u>13</u>
A8						<u>13</u>	
E9					<u>13</u>		
A9		X					
E10							<u>13</u>

Les numéros sont ceux des Introduteurs de Complexité (IC) répertoriés ci-dessous (par ordre alphabétique). Le numéro 13 désigne la construction du discours indirect.

à + verbe à l'infinitif	17	Ex : elle donne à manger
Comme (= étant donné que)	16	Ex : comme la cloche a sonné, ils sont rentrés dans la classe
Comparative (= comme)	21	Ex : elle écrit comme elle parle
de + verbe à l'infinitif	18	Ex : elle n'est pas contente de voir sa maman
discours indirect (paroles rapportées)	13	Ex : elle dit de faire la vaisselle ; elle dit que je suis beau
divers	24	Opposition (tandis que, alors que), tellement que, sans que, surtout que, déjà que, sinon, etc
Extraction	1	c'est, voilà, il y a ... qui, que, où
Gérondif	20	Ex : en marchant
il faut que	4	Ex : il faut que j'aide mes amis
interrogative indirecte	12	Ex : je sais ce que, je vais voir comment, je (ne) sais (pas) si je vais aller
introduteurs temporels (autres que « quand »)	7	dès que, après que, chaque fois que, pendant que, avant que, etc
Où relatif	14	Relatif ; ex : elle a trouvé le paquet où il y a les couches du bébé
parce que	8	Ex : le bébé ne boit pas parce qu'il a déjà bu son lait
pour + verbe à l'infinitif	19	Ex : papa sort sa clé pour ouvrir la porte
pour que	15	Ex : elle appuie un petit peu pour que le dentifrice sorte du tube
puisque	23	Ex : il va jouer tout seul puisque sa copine est partie
Quand (= lorsque)	6	Ex : ils sont contents quand ça roule

Quantitative (comparatif)	11	Plus, moins, autant... que ; ex : Paul est plus grand que Pierre
Que conjonction	5	Ex : elle trouve que c'est drôle (« que » après un verbe conjugué)
Que relatif	22	Ex : elle ouvre le cadeau que son papa lui a donné
Qui relatif	3	Ex : il sort du camion un canapé qui est très lourd
Si (supposition et condition)	9	Ex : si le chat griffe le ballon il peut éclater
V + verbe à l'infinifit	17	Ex : elle veut manger

La grille d'analyse en catégories syntaxiques permet de visualiser le degré de maîtrise des constructions par l'enfant. Ainsi dans l'extrait étudié nous voyons pour le discours indirect :

- une alternance d'essais : en E8 et E10, il y a insertion de discours direct avec un possessif ou un pronom de première personne dans la forme de discours indirect *dire que*, qui aurait ici nécessité une transposition avec un possessif et un pronom de troisième personne.

- une reprise immédiate d'un énoncé de l'adulte : en E9, l'enfant reprend la reformulation de la construction du discours indirect par l'adulte (A8).

Ainsi, cette grille permet de localiser un élément en cours d'acquisition pour faire ensuite une analyse plus précise de l'évolution de ses contextes d'emploi. Pour l'exemple proposé ci-dessus, la construction du discours indirect peut être repérée sur l'ensemble du corpus longitudinal des interactions de cet enfant avec l'adulte. Il est ainsi possible de saisir à quel moment la construction du discours indirect se stabilise en ce qui concerne la forme syntaxique, et aussi d'évaluer son autonomie, c'est-à-dire son indépendance vis à vis des verbalisations de l'adulte et sa production dans des contextes linguistiques variés.

4. UNE AIDE INFORMATIQUE POUR LE RECUEIL ET L'ANALYSE DES DONNÉES ?

4.1. Archivage de données numérisées : le projet TCOF

Les données recueillies depuis de nombreuses années dans cette perspective linguistique font depuis peu l'objet d'un archivage (stockage et sauvegarde numérique) dans le cadre du projet TCOF « Traitement de Corpus Oraux en Français » à l'ATILF. 63 corpus longitudinaux d'interaction adulte-enfant, soit 175h d'enregistrement, sont en cours d'archivage. Ces corpus ont fait l'objet d'une transcription orthographique intégrale des

énoncés des deux locuteurs⁴ avec le logiciel d'alignement texte-son *Transcriber*⁵. La visée est double :

- valider les hypothèses et les analyses réalisées jusqu'ici en linguistique de l'acquisition en travaillant sur de grands corpus ;
- mettre les corpus à disposition publique (première phase en 2010) pour une plus grande mutualisation des données pour la recherche.

Les données actuellement disponibles au plan international présentent une étendue réduite, la base de données CHILDES restant la référence, soit 230 corpus d'interactions verbales adulte-enfant (majoritairement non provoquées) dans trente langues différentes. Pour le français, le contenu est relativement mince puisque dix corpus longitudinaux sont archivés, dont seulement deux au-delà de l'âge de 3 ans.

Un archivage dans la base de données TCOF de plusieurs corpus longitudinaux d'interaction avec des enfants de plus de 3 ans devrait combler ce manque et permettre des descriptions syntaxiques plus approfondies et des analyses validées statistiquement. Par ailleurs, dans la mesure où la base de données TCOF comporte aussi des corpus de français parlé par des adultes (environ 200h en cours d'archivage), elle devrait nous permettre de comparer les caractéristiques linguistiques des productions d'adultes avec celles d'adultes s'adressant aux enfants.

4.2. Repérage d'éléments à l'aide d'un concordancier

Un concordancier permet de répertorier dans un ou plusieurs corpus un même élément (ou une chaîne de caractères) et d'afficher les occurrences en contexte, par exemple les formes en « dire que ». Un concordancier permet de travailler sur un grand nombre de corpus, ce qui représente un gain de temps et une certaine fiabilité. Il permet donc de dépasser les limites d'une analyse manuelle où la recherche d'éléments est souvent limitée à quelques corpus (en raison de son coût en temps) et sujette à erreurs (comme des oublis des occurrences et des vérifications aléatoires).

Néanmoins, pour les corpus d'interaction adulte-enfant, un

⁴ En accord avec les principes de base des travaux sur le français parlé (Blanche-Benveniste, 1997).

⁵ Le choix de *Transcriber* est lié au fait qu'il permet d'associer le texte au son, qu'il permet de diminuer le temps de transcription mais également qu'il facilite le passage à un format TEI (*Text Encoding Initiative*), utile pour un archivage sur une plateforme.

concordancier présente des limites :

- il ne répertorie pas d'emblée les formes en essai : il faudrait les connaître par avance ou les découvrir par hasard. Il ne répertorie pas non plus les formes en émergence non réalisées comme dans : *elle voit Pilou est plus là*.

- il ne donne pas une vision globale des phénomènes de reprises et reformulations, et en particulier les interactions différées.

5. CONCLUSION

Nous avons vu que les méthodologies de recueil et d'analyse des données en acquisition du langage sont fortement dépendantes des objectifs d'étude et surtout du cadre théorique sous-jacent, et qu'il est difficile de concilier les positions des psycholinguistes développementaux et celles des linguistes de l'acquisition du fait de conceptions du langage divergentes. Néanmoins, on peut espérer que ces différences s'estompent avec l'exploitation de grands corpus, qui permettrait entre autres de concilier certains aspects des méthodologies qualitatives et quantitatives et peut-être de vérifier les données expérimentales.

Mais d'autres problèmes émergent : il reste notamment à déterminer la quantité de données suffisante à une généralisation (Tomasello & Stahl, 2004) et la représentativité des corpus, ce qui nous ramène à la question des « genres » : un échantillon de 1000 mots est-il suffisant pour identifier la distribution des caractéristiques d'un texte (Biber & Finegan, 1991) ?

BIBLIOGRAPHIE

- BIBER D. (1988), *Variation across Speech and Writing*, Cambridge, Cambridge University Press.
- BIBER D. & FINEGAN E. (1991), "On the exploitation of computerized corpora in variation studies", in Aijmer K. & Altenberg B. eds., *English Corpus Linguistics*, 204-220
- BIBER D., JOHANSSON S., LEECH G., CONRAD S. & FINEGAN E. (1999), *Longman Grammar of spoken and written English*, Harlow, Pearson.
- BLANCHE-BENVENISTE C. (1997), *Approches de la langue parlée en français*, Paris, Ophrys.
- BRAINE M. D. S. (1963), "The ontogeny of English phrase structure : the first phase", *Language* 39, 1-13.
- BROWN R. (1973), *A First Language : the early stages*, Cambridge, MA, Harvard University Press.
- BRUNER J. S. (1991), *...car la culture donne forme à l'esprit. De la révolution cognitive à la psychologie culturelle*, Paris, Eshel.

- CANUT E & 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.
- CHRISTOPHE A. (2002) « L'apprentissage du langage (les bases cérébrales du langage) », Y. Michaud dir., *Université de tous les saviors 5, Le cerveau, le langage, le sens*, Paris, Odile Jacob, 195-213.
- COHEN M. (1993 [1925]), « Sur les langages successifs de l'enfant », *Mélanges linguistiques offerts à M. J. Vendryès par ses amis et ses élèves*, Paris, Champion, republié dans *L'Acquisition du Langage Oral et Ecrit*, 31, 33-43.
- DARWIN C. (1877) « A biographical sketch of an infant » *Mind* 2-7, 285-94.
- DEMUTH K., McCULLOUGH E. (2009), "The longitudinal acquisition of clusters in French", *Journal of Child Language* 36-2 425-448.
- GALLAWAY C. & RICHARDS B. J. (1994), *Input and Interaction in Language Acquisition*, Cambridge, Cambridge University Press.
- GREGOIRE A. (1937), *L'apprentissage du langage : les deux premières années*, Paris, Droz.
- GREGOIRE A. (1947), *L'apprentissage du langage II : la troisième année et les années suivantes*, Paris, Droz.
- HOUDE O., KAYSER D., KOENIG O., PROUST J., RASTIER F. (1998), *Vocabulaire de sciences cognitives*, Paris, P.U.F.
- INGRAM D. (1989), *First Language Acquisition. Method, Description and Explanation*, Cambridge, Cambridge University Press.
- ITARD J.-M. (1801), *De l'éducation d'un homme sauvage ou des premiers développements physiques et moraux du jeune sauvage de l'Aveyron*, Paris, Goujon.
- Journal of Child Language* 36-2 (March 2009).
- KAIL M. & FAYOL M. (2000), *L'acquisition du langage. Le langage en développement. Au-delà de trois ans*, Paris, P.U.F.
- KARNOUOH-VERTALIER M. (1997), "Constitution des données à observer en acquisition du langage : analyse du n° 23-2, 1996, du *Journal of Child Language*", intervention au séminaire "Vers une linguistique des pratiques langagières", Université Sorbonne Nouvelle - Paris 3.
- KARNOUOH-VERTALIER M. (1998), "Evolution 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.
- LENTIN L. (1971), "Recherche sur l'acquisition des structures syntaxiques chez l'enfant entre 3 et 7 ans", *Etudes de Linguistique Appliquée* 4, 7-52.
- LENTIN L. (2009), *Apprendre à penser, parler, lire, écrire*, Paris, E.S.F.
- LENTIN L. et al. (1988), *Recherches sur l'acquisition du langage*, t. 2, Paris, Presses de la Sorbonne Nouvelle.
- McCARTHY D. (1952 [1946]), "Le développement du langage chez l'enfant", *Manuel de psychologie de l'enfant*, t. 2, Carmichael L. (Ed.), Paris, P.U.F., 751-916.
- OTTAVI D. (2007), "L'apprentissage du langage chez Michel Bréal", *Actualités de la Recherche en Education et Formation*.
- PAVLOV I. (1963 [1924]), *Réflexes conditionnels et inhibitions*, Genève, Gonthier.

- PREYER W. (1887), *L'âme de l'enfant*, Paris, Alcan (*Kinder Seele*, 1882).
- SKINNER B. F. (1957), *Verbal behavior*, New York, Appleton Century-Crofts.
- SMITH M. (1926) "An investigation of the development of the sentence and the extent of vocabulary in young children", *University of Iowa Studies in Child Welfare*, vol. 3, n° 5.
- STERN C. et W. (1907), *Die Kindersprache*, Leipzig, Quelle Meyer.
- TAELMAN H., DURIEUX G., GILLIS S. (2009), "Fillers as signs of distributional learning", *Journal of Child Language* 36-2, 323-353.
- TAINÉ H. (1876), "Notes sur l'acquisition du langage chez les enfants et dans l'espèce humaine", *Revue philosophique*, 1.
- TAYLOR N., DONOVAN W., MILES S. & LEAVITT L. (2009), "Maternal control strategies, maternal language usage and children's language usage at two years", *Journal of Child Language* 36-2, 381-404.
- TEMPLIN M. (1957) *Certain language skills in children*, University of Minnesota Institute of Child Welfare Monograph Series 26, Minneapolis, University of Minnesota Press.
- TOMASELLO M. & STAHL D. (2004), « Sampling children's spontaneous speech : how much is enough ? », *Journal of Child Language* 31-1, 101-121.
- WATSON J.B. (1914), *Behavior : an introduction to comparative psychology*, Holt.
- WEIR R. (1962), *Language in the Crib*, The Hague, Mouton.