



HAL
open science

Corpus ESLO-Enfants : de sa création aux premiers résultats

Jennifer Ganaye

► **To cite this version:**

Jennifer Ganaye. Corpus ESLO-Enfants : de sa création aux premiers résultats. Corpus, 2021. halshs-03162347

HAL Id: halshs-03162347

<https://shs.hal.science/halshs-03162347v1>

Submitted on 22 Nov 2021

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.

Corpus ESLO-Enfants : de sa création aux premiers résultats

Jennifer Ganaye – Laboratoire Ligérien de Linguistique UMR 7270

Résumé

Le module ESLO-Enfants est un corpus variationniste longitudinal issu du grand corpus de langue française : Enquêtes SocioLinguistiques à Orléans (ESLO). Variationniste car il s'appuie sur un public varié (enfants de 2 ans à 7 ans avec leur entourage proche) provenant de familles de différentes catégories socio-économico-culturelles et enregistré dans des situations naturelles variées formant le quotidien des enfants. Ce corpus, qui sera mis à disposition à la fois sur ESLO et Childes, a été créé dans le but d'étudier l'impact de l'environnement langagier (input) sur les productions enfantines (output). Pour cette étude, la liaison, objet à la fois stable et variable, s'est révélée être un phénomène sociolinguistique propice pour observer sa transmission à des enfants en cours d'acquisition du langage mais aussi pour observer d'éventuelles différences entre les familles.

Abstract

The ESLO-Children module is a longitudinal variationist corpus derived from the large French-language corpus: Enquêtes SocioLinguistiques à Orléans (ESLO). It is variationist because it is based on a varied audience (children aged 2 to 7 with their close relatives) from families of different socio-economic and cultural categories and recorded in a variety of natural situations that form the daily lives of children. This corpus, which will be made available on both ESLO and Childes, was created with the aim of studying the impact of the language environment (input) on children's productions (output). For this study, liaison, an object that is both stable and variable, has been shown to be a sociolinguistic phenomenon that is conducive to observing its transmission on children in the process of language acquisition, but also to observing possible differences between families.

Mots clefs :

Acquisition du langage, corpus variationniste, liaison, ESLO, parole enfantine, situation naturelle

Keywords :

Origines du corpus ESLO-Enfants

L'initiative de mettre au point un corpus intégrant dans un même projet la parole des enfants, la parole de leur entourage et la prise en compte de l'environnement dans lequel les échanges ont lieu trouve son origine dans notre volonté d'élargir le périmètre actuel des études de corpus enfantins en français, mais aussi d'étudier, l'impact de l'input sur l'output en fonction de différents environnements langagiers.

Les nombreux travaux menés dans le cadre des Théories Basées sur l'Usage (UBT) mettent en évidence l'importance de l'environnement (ou input) des enfants pour l'apprentissage de leur(s) langue(s) (Tomasello, 2003). Cet input constituant d'une certaine façon la matière première (Bybee & Hopper, 2001) à partir de laquelle l'enfant va pouvoir construire la grammaire de sa langue, développer son lexique. Dès le plus jeune âge, l'enfant reçoit une grande quantité de paroles de la part de son entourage qui l'aidera à développer ses compétences lexicales, dont une partie provient des scènes d'attention conjointe (Goldwin-Meadow, 2010).

Pour l'heure, les corpus enfantins en français (citons par exemple : CoLaJE, ALIPE, TCOF et *Corpus de Lyon*) ne peuvent rendre compte de cela que de façon parcellaire. Nous constatons que les enfants participants à ces différents projets sont souvent issus de milieux favorisés, voire très favorisés. Cela est souvent dû à la facilité de trouver des familles ayant un niveau socio-économique plus élevé dans l'entourage du chercheur. Mais qu'en serait-il avec des enfants évoluant dans un milieu moins favorisé ? C'est en cela que le module ESLO-enfants se veut novateur et original puisque nous souhaitons élargir le cercle des enfants enregistrés en veillant à intégrer, dans notre échantillon, des enfants issus de milieux sociaux variés à l'image de ce que peut être la population d'une ville. Constituer un tel panel permettra d'accéder à des interactions familiales pour lesquelles nous n'avons pour l'heure que très peu de données et peut-être de questionner l'influence des milieux socio-économiques sur l'usage du langage.

En effet, s'inscrire dans une approche où l'input est au centre de la construction du langage a des conséquences théoriques et méthodologiques fortes. Dans ce cadre, nous empruntons certains principes de la linguistique variationniste (Labov, 1976) qui sont en totale cohérence avec ceux des UBT (Kemmer & Barlow, 2000).

Il faut aussi prendre en compte que notre corpus fait partie d'un projet de laboratoire plus large de recueil, documentation et mise à disposition d'un grand corpus de français parlé (Enquêtes SocioLinguistiques à Orléans : ESLO) collecté à deux périodes : dans les années 1970 et dans les années 2010. Dès la première phase de collecte, les chercheurs avaient pour objectif de créer un corpus de type sociolinguistique :

“Selon nous une recherche sociolinguistique impliquait une étude de la langue dans sa diversité plutôt que comme un tout homogène et figé. En effet, même si on étudie un état de langue à un moment précis de l'histoire, il n'empêche qu'il offre une variété à plusieurs niveaux : différences entre les générations ; différences dialectales entre communautés ; différences entre les milieux sociaux ; différences liées aux conditions de production du discours.” (Blanc & Biggs, 1971: 16)

Par la suite, l'approche variationniste a orienté le travail des chercheurs, en mettant notamment en avant la prise en compte de l'hétérogénéité des pratiques langagières, et en cherchant à rendre compte de la structure des variations (Abouda & Baude, 2005). L'importance accordée aux données et à la possibilité de faire varier les contextes/situations a grandement guidé notre travail. Nous empruntons à ESLO également sa chaîne de traitement de corpus.

Dans ce corpus variationniste, nous centrerons notre étude sur un phénomène linguistique : l'usage des liaisons nous permettant d'observer ou non l'influence de l'input sur les productions enfantines, point sur lequel nous reviendrons plus en détails par la suite.

Apports du corpus ESLO-Enfants

Les apports du corpus ESLO-Enfants se situent à deux niveaux. Ce module constitue en premier lieu un apport pour ESLO. Nous pouvons en effet constater qu'aujourd'hui, dans ESLO, il n'y a pas de données contenant des paroles d'enfants en interaction. Précisons toutefois l'existence de deux modules qui intègrent des enfants :

- le module « Ecole », où il est possible d'entendre des enfants, mais ces enregistrements ne sont pas focalisés sur la parole enfantine, mais plutôt sur l'ensemble de la vie d'une école.

- le module « Livres pour enfants » en cours de traitement, dans lequel l'objectif était d'enregistrer des moments de lecture faits à des enfants, les lecteurs pouvaient être les parents ou bien des personnes de l'entourage proche des enfants. Dans ce dernier, les enregistrements sont focalisés sur les paroles de l'adulte et non celles des enfants.

Notre corpus élargit le périmètre actuel d'ESLO avec de nouveaux locuteurs d'Orléans qui représentent une partie non négligeable de la population de la ville.

Par ailleurs, nous ajoutons également un nouvel horizon aux corpus enfantins francophones déjà existants, notre corpus comprenant une grande variété de situations d'enregistrement, tout en diversifiant les personnes avec lesquelles l'enfant est amené à échanger : par exemple, lorsque l'enfant est amené à passer du temps avec ses grands-parents que ce soit chez eux ou bien chez l'enfant, mais aussi dans sa classe. Nous cherchons donc à obtenir des scènes du quotidien où l'enfant peut jouer avec des amis, sa famille, etc. ou encore lorsqu'il est gardé par une baby-sitter. Nous ne nous restreignons pas aux moments de loisirs, nous captions également des moments simples mais qui font partie d'un quotidien rythmé par des habitudes comportementales et langagières. Ainsi, l'enregistrement de scènes contenant les repas en famille, le bain ou encore le coucher peut être primordial puisque ces dernières contiennent un lexique récurrent où l'enfant pourrait alors acquérir des routines langagières spécifiques. Notre but étant de nous rapprocher au mieux des conversations que pourrait avoir quotidiennement un enfant, pour pouvoir étudier ses différentes sources d'input.

Pour résumer, la création d'un module aussi varié va nous permettre de mieux connaître les influences que les différences socioculturelles peuvent avoir sur l'acquisition de la langue. Nous approfondirons ce point par l'étude du lien input/output à travers la diffusion de l'usage de la liaison. Notre projet permet donc de compléter le corpus ESLO2 avec des locuteurs qui ne sont que marginalement présents à l'heure actuelle, et dans l'esprit variationniste d'ESLO en veillant à prendre en compte des niveaux de variations différents tout en complétant les corpus enfantins francophones existants.

Objet d'étude : la liaison

Pour observer la question du lien entre input et output, nous décrirons l'usage des liaisons à la fois dans l'environnement langagier des enfants et dans leurs premières productions. A l'image du travail mené par Liégeois (2014), nous chercherons à rendre

compte de l'impact de l'environnement langagier sur les formes produites par les enfants et sur sa propre courbe d'acquisition par rapport au modèle de Chevrot et al. (2009).

1. La liaison : définition

La liaison est un phénomène oral variable typique du français. Il s'agit de l'oralisation de la consonne finale d'un Mot1 devant un Mot2 à voyelle initiale, consonne muette quand le Mot1 est produit en finale d'énoncé ou devant un mot à initiale consonantique (Soum-Favaro et al. 2014 ; Chevrot et al. 2007) :

1. *le gros éléphant*
[lɔgʁozɛlefã]
2. *le gros camion*
[lɔgʁokamjɔ̃]
3. *il est gros*
[ilɛgʁɔ]

Dans les cas ci-dessus, nous voyons que le « s » de « gros » se prononce sous la forme d'une liaison en /z/ uniquement lorsqu'il précède un mot à initiale vocalique (1).

Plusieurs études de corpus avec des méthodes différentes et sur des corpus différents (Corpus d'Orléans : De Jong 1994, ESLO : Dugua & Baude 2017, PFC : Durand & Lyche 2008, Coquillon et al. 2010, Durand et al. 2011) décrivent d'ailleurs les variations d'usages des liaisons chez l'adulte. Toutes s'accordent sur le fait qu'il existe au moins deux types de liaisons :

- Des liaisons systématiques/invariables (classiquement appelées Liaisons obligatoires - LO) : les liaisons toujours réalisées quels que soient le locuteur et la situation (De Jong, 1994), exemple :
 - après déterminant (*des [z] enfants*)
 - entre pronom personnel et verbe (*ils [z] appellent*)
 - entre verbe et pronom (*prends [z] en*)
- Des liaisons variables ou facultatives (LF) : exemple : « un gros [z] éléphant », qui peut être réalisée ou non. Ce type de liaison est considéré comme une variable sociolinguistique du français dans le sens où, comme le précise Gadet (1989 : 71), elle « permet [à elle seule] de classer socialement un locuteur ». L'usage des liaisons variables est conditionnée par des facteurs linguistiques comme les catégories grammaticales des mots, la longueur des mots, la nature de la consonne de liaison,

etc. et des facteurs extralinguistiques (ou sociaux) comme le milieu social, les situations de communications (Gadet, 1989). Dans le cas du français, la langue des classes favorisées correspond, dans les esprits, à la norme ou la langue standard (Gadet, 1997), il est donc possible de distinguer des personnes issues de classes différentes par leurs pratiques langagières. Pour les liaisons, par exemple, les classes favorisées ont tendances à plus prononcer les liaisons facultatives (Encrevé, 1988).

2. La liaison : phénomène linguistique percutant

Pour étudier le lien input/output, s'appuyer sur un objet qui peut être à la fois stable et variable nous fournit un cadre d'analyse riche. Nous pourrions observer par exemple si, ce qui est stable/invariable chez l'adulte, l'est aussi chez l'enfant et à partir de quel âge ; si ce qui est stable est reçu par tous les enfants de la même manière, et ce qui peut différer (la quantité de liaisons stables, la diversité des liaisons stables). Quant aux liaisons variables nous étudierons comment elles varient d'une famille à l'autre, d'un enfant à l'autre et d'une situation à l'autre.

Notre approche est fondamentalement basée sur l'usage, sur des données de parole recueillies, et non sur des conceptions normatives ou basées sur des exemples. En étudiant l'usage des liaisons à la fois dans la parole adulte et dans la parole enfantine et en décrivant ce que l'enfant peut récupérer, et cela dans différentes familles ayant un rapport à la langue varié, nous pourrions voir comment la variation se réalise (côté adultes) et se construit (côté enfant) en perception et en production.

Pour convenir à la démarche variationniste recherchée et pour répondre à nos objectifs sur le plan de l'acquisition du langage, nous avons mis en place une procédure de collecte particulière pour accéder au quotidien des enfants de la façon la moins invasive possible, procédure que nous présentons dans la section suivante.

Méthodologie d'enquête

Rappelons que le corpus ESLO-Enfants s'inscrit dans le plus grand corpus de français parlé en France, le corpus ESLO (<http://eslo.huma-num.fr/>). Nous avons donc suivi le cadre méthodologique de ce projet tout en l'adaptant à notre nouveau public infantin. Tout d'abord, l'architecture du corpus ESLO (Baude & Dugua, 2011), à savoir la façon dont les différents modules coexistent et la forme que cette construction donne au corpus, est le reflet du « parti pris variationniste des auteurs d'ESLO1 et ESLO2 [...] afin de permettre la prise en compte de différents types de variations » (Baude & Dugua, 2015 : 357). Dans cette architecture, le

corpus ESLO-Enfants constitue un module particulier puisque nous l'envisageons comme un modèle réduit d'ESLO dans le sens où notre objectif est d'intégrer en son sein la diversité des situations auxquelles un enfant participe au quotidien. Nous cherchons en effet par la constitution de ce corpus à balayer une large palette de situations familiales des enfants, sachant toutefois que les contraintes liées à la collecte limiteront les possibilités.

1. Présentation des familles : caractéristiques sociales

Les locuteurs que nous souhaitons enregistrer sont des enfants compris dans une tranche d'âge entre 2 ans et 7 ans, ce qui correspond à la période qui précède l'entrée à la maternelle, la maternelle et l'entrée dans l'écrit. La variété des âges nous permet d'étudier et de comparer des enfants à différents stades de leur acquisition du langage. De plus, comme le phénomène de liaison est lié à la forme graphique des mots (la présence d'une liaison se déclenche par une lettre écrite non oralisée), il nous semblait intéressant de pouvoir observer les enfants avant et après l'entrée dans l'écrit (Fijalkow, 1997), afin de voir si l'émergence de l'écrit et le rapport à l'écrit peuvent influencer les usages des liaisons.

Les enfants de cette enquête sont issus de familles choisies avec minutie. Les appartenances socio-économico-culturelles des familles sont diversifiées pour rendre ce corpus représentatif de la population française et plus précisément, représentatif de la population de l'agglomération orléanaise, mais aussi pour apporter de nouvelles données quant aux paroles enfantines.

Pour décrire ces familles, nous nous appuyons sur différents critères car nous estimons que les catégories socioprofessionnelles de l'INSEE ne sont pas suffisantes. Classifier les personnes seulement sur leurs catégories socioprofessionnelles ne suffit pas à déterminer leur capital culturel et économique qui sont des éléments qui pourraient nous éclairer sur le capital hérité des enfants, comme Bourdieu le défend en 1979.

C'est pour cela que nous avons décidé de créer notre propre échelle qui fait intervenir trois grands axes : l'axe formation(s)-emploi(s), l'axe temps libre (sorties) et l'axe médias. Ainsi, nous proposons une échelle socio-linguistico-culturo-professionnelle (SLCP), pour classer les différentes familles.

A l'aide de ces trois axes, nous avons établi une fiche (voir un extrait ci-dessous) déclinant les principales particularités de chacun des axes. Chaque parent a rempli la fiche informative en précisant des échelles qui varient de 1 à 10. Ainsi, les parents inscrivent eux-mêmes la

proportion que chaque axe représente dans leur quotidien. Pour cela, nous nous sommes fondée en partie sur les travaux de Bourdieu (1994) sur la théorie du monde social, et l'avons adapté à nos exigences en termes de public visé et d'informations sur leur input quotidien. Pour chaque axe, nous prenons en compte le temps et la fréquence des activités pratiquées.

Culture/Loisirs : A quelle fréquence y allez-vous ?

Spectacles										
0	1	2	3	4	5	6	7	8	9	10
Cinéma										
0	1	2	3	4	5	6	7	8	9	10
Concert										
0	1	2	3	4	5	6	7	8	9	10
Festivals										
0	1	2	3	4	5	6	7	8	9	10
Théâtre										
0	1	2	3	4	5	6	7	8	9	10
Musées										
0	1	2	3	4	5	6	7	8	9	10
Médiathèque										
0	1	2	3	4	5	6	7	8	9	10

Extrait de la fiche informative

Dans l'axe formation(s)-emploi(s), nous faisons apparaître la/les formation(s) que les parents des familles participantes ont suivies, les éventuelles reprises d'études, le lieu d'étude (pays, région, ville). De la même manière, nous prendrons en compte les emplois qu'ont effectués les parents, la période durant laquelle ils les ont effectués mais aussi, si c'est un travail solitaire ou bien en équipe.

L'axe du temps libre concerne toutes les sorties et activités que les participants peuvent effectuer. Nous les regroupons dans différentes catégories : les sorties culturelles (telles que le théâtre, le cinéma, etc.), les sorties sociales (retrouver des amis pour boire un verre, aller au restaurant, etc.) et des sorties et activités qui comprennent les activités extra-professionnelles (concerne les activités en club mais aussi si les personnes pratiquent des activités à la maison comme par exemple le tricot, les jeux avec leurs enfants, etc.).

Le troisième axe concerne le rapport aux médias. Dès 1950, Queneau faisait remarquer que les médias influencent fortement notre manière de nous exprimer. Ainsi, depuis l'arrivée de la télévision, les individus calquent leur manière de s'exprimer sur celle des présentateurs de télévision car cette forme de langue est jugée plus légitime (Queneau, 1950). Au-delà du petit écran avec le type d'émission que les participants regardent, nous allons nous intéresser aux

types de radio, et de lectures, mais aussi la question des usages des vidéos et jeux vidéo, des réseaux sociaux.

Voici une des représentations sous forme de radars obtenues avec le questionnaire que nous venons de décrire. Les exemples exposés ci-dessous concernent les familles des enfants que nous étudierons dans la suite de cet article.

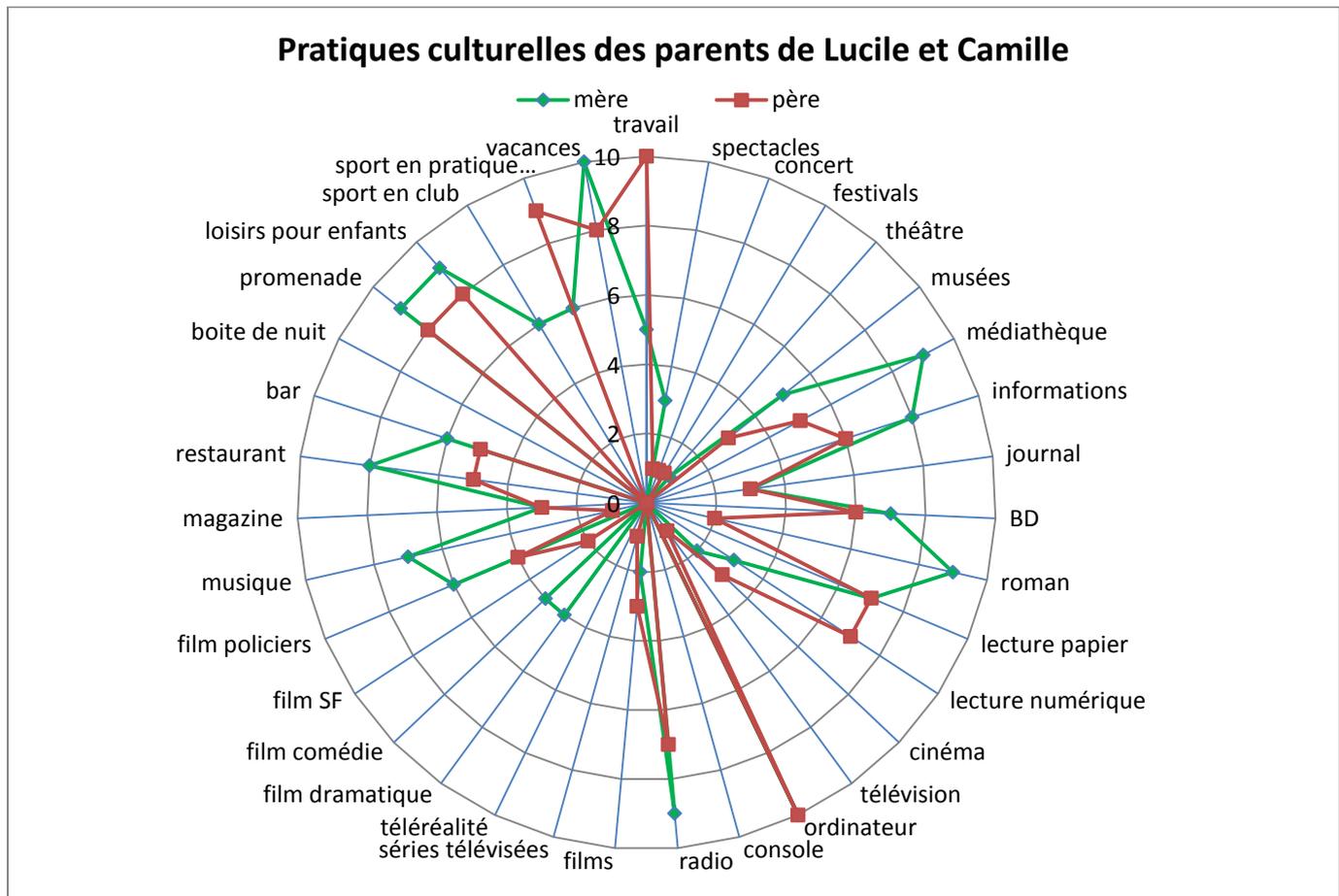


Figure 1 : Représentation des pratiques culturelles des parents de Lucile et Camille

Les tracés révèlent que les parents de Lucile et de Camille ont des habitudes de sorties à tendances socialisantes hors du cadre familial proche avec de hauts résultats concernant les vacances, le sport en club, les promenades, les restaurants, etc. Tout en ayant quelques habitudes littéraires.

Voyons à présent les habitudes des parents de Roxane :

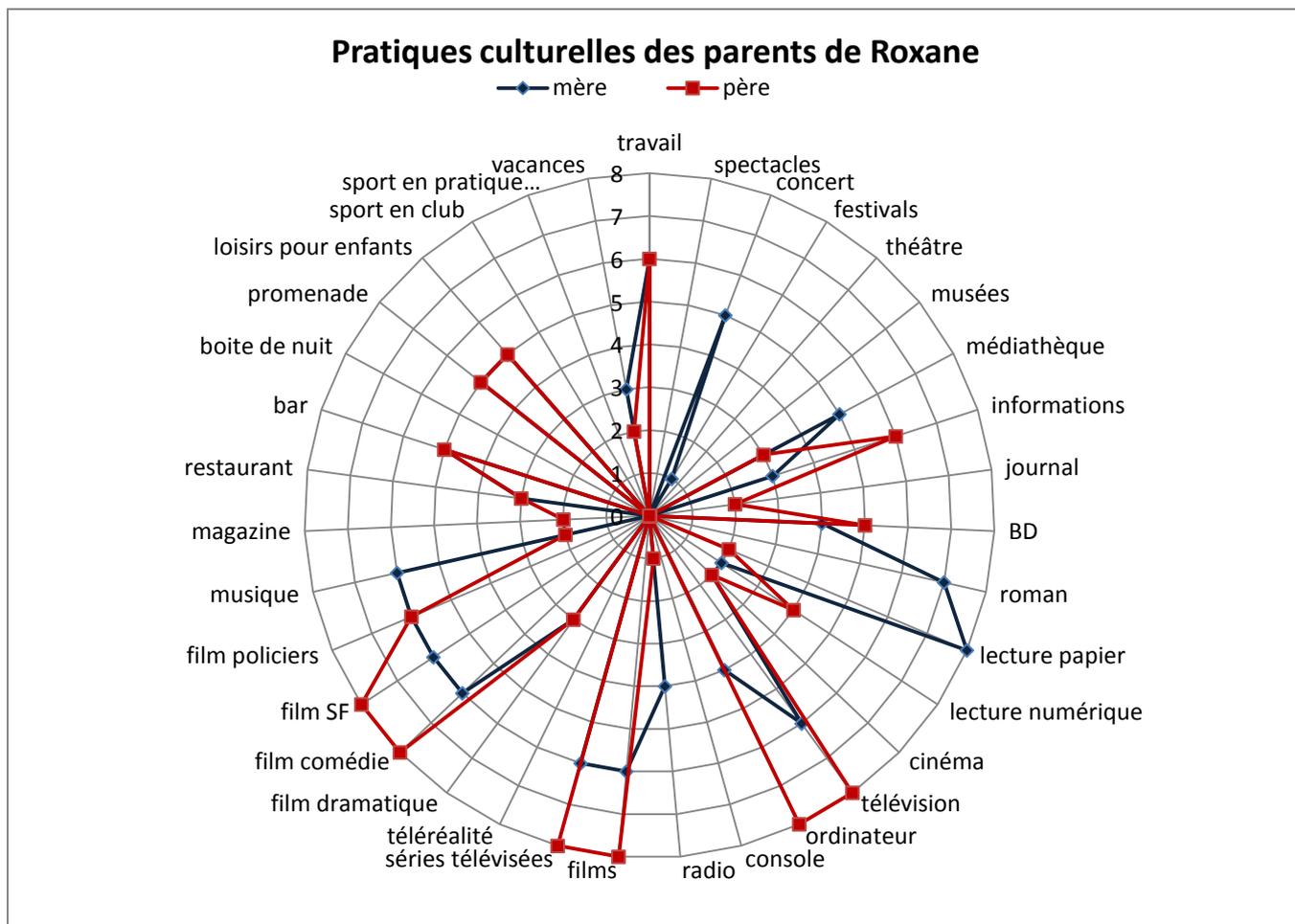


Figure 2 : Représentation des pratiques culturelles des parents de Roxane

Dans le cas de cette famille, nous voyons que les parents de Roxane aiment les activités audiovisuelles, mais aussi la lecture. Les sorties à tendances sociales hors du cadre familial sont néanmoins présentes mais avec des résultats moins élevés que pour la famille précédente.

Nous retiendrons que les pratiques culturelles de la famille de Camille et Lucile sont plus des activités en extérieur, à visée socialisante alors que les parents de Roxane sont sur des activités qui peuvent se pratiquer à l'intérieur même de leur foyer, qui ne sortent donc pas du cadre familial. Ces radars sont un extrait de ce que nous pouvons obtenir à l'aide de la fiche de renseignements, nous pourrons les croiser avec les résultats obtenus pour les usages des liaisons dans le corpus afin de voir dans quelle mesure l'usage des liaisons facultatives peut être influencé par les pratiques d'activités différentes. Ainsi, notre méthode permettra de croiser des représentations issues du questionnaire et des pratiques à partir de l'usage des liaisons.

2. Panel de locuteurs

La contrainte consistant à diversifier les familles sur le plan socioculturel a influencé notre façon d'aborder les prises de contact. Il est évident que les CSP défavorisées restent les familles les plus difficiles à atteindre. En effet, différentes caractéristiques de notre protocole peuvent freiner tout particulièrement les locuteurs appartenant à des CSP défavorisées :

- l'insécurité linguistique de cette classe (Bourdieu, 1982),
- le jugement des pratiques linguistiques et sociales qui peut être imposé sur leur vie, sur l'éducation qu'ils donnent à leur(s) enfant(s) (Neill, 1966 ; Gagnier, 2011),
- l'enregistrement de mineurs et de leur milieu intime que constitue la famille (Linard & Prax, 1984).

Tous ces points compliquent la façon d'approcher ces locuteurs, et d'obtenir leur accord. Nous avons alors mis en place différents modes d'approche.

L'un d'eux a consisté à nous rapprocher d'une école élémentaire dont nous connaissions la directrice et où la proportion d'enfants issus de familles de CSP basses est importante. Nous sommes entrée directement en contact avec les familles lors d'une présentation orale du projet au moment de la réunion de rentrée dans les deux classes de CP. Ce processus a permis d'avoir un plus grand impact sur les familles que de simplement leur donner un prospectus récapitulatif. Nous avons obtenu par ce biais l'accord de la famille de Marise.

L'autre moyen d'accéder à une diversité de familles a été d'utiliser nos réseaux, ce qui nous a permis d'enregistrer la famille de Fréïhr et Briséis (réseau proche) et la famille de Roxane (réseau plus distant). Les deux dernières participantes, Lucile et Camille, sont les filles d'une enseignante chercheuse participant au corpus ESLO.

Voici le panel d'enfants du corpus :

Enfant étudié	Age période 1	Age période 2	Age période 3	Membres de la famille	
Marise	6;8	7;5	Non disponible	- Mère - Père	- Sœur aînée
Fréïhr	1;11	2;10	Non disponible	- Mère - Père	- Sœur Briséis - Sœur aînée
Briséis	7;1	7;11	Non disponible	- Mère - Père	- Frère cadet Fréïhr - Sœur aînée
Roxane	3;10	4;8	5;3	- Mère - Père	- Sœur cadette
Lucile	3;9	4;9	5;2	- Mère - Père	- Sœur aînée (Camille)

Camille	6;11	7;10	8;3	- Mère - Père	- Sœur cadette (Lucile)
---------	------	------	-----	------------------	-------------------------

3. Formulaire de consentement / aspects juridiques / éthiques

Lors de la rencontre avec les familles pour finaliser leur participation, nous faisons remplir et signer un formulaire de consentement éclairé qui reprend les grandes idées présentées aux familles. Il y est par exemple précisé par écrit (liste non exhaustive) :

- en quoi consiste le projet ESLO-Enfants,
- l'anonymisation des données identifiantes, hormis les prénoms des enfants (en accord avec les familles)
- le fait que la vidéo ne sera pas rendue disponible
- que les documents audio et les transcriptions seront diffusés sur deux plateformes pour le monde de la recherche,

Ce document est nécessaire pour rendre compte de ce qui va être fait et le devenir des données et métadonnées qui vont être récoltées, mais aussi pour protéger les droits des familles participantes et du Laboratoire Ligérien de Linguistique (LLL).

Matériel d'enregistrement et mise en place

Pour mener à bien ces enregistrements et la transcription qui en découlera, nous procédons à deux types d'enregistrements. Le premier est un enregistrement vidéo qui permettra de mieux comprendre les scènes puisque avec un public enfantin, il est souvent nécessaire de visualiser la scène pour savoir à qui l'enfant parle et de quoi il parle. Pour cela, nous choisissons d'utiliser une caméra d'action car elle possède un champ d'enregistrement large et parce qu'elle est très petite, ce qui permet d'être moins intrusif quant à l'intimité des familles (Linard & Prax, 1984). Nous espérons, ainsi, réduire le « paradoxe de l'observateur » (Labov, 1972). Le second est un enregistrement audio pour suivre au mieux les paroles des participants principaux. Nous choisissons d'utiliser deux micros cravate HF sans fil qui seront portés par l'enfant et un membre de la famille.

Afin de réduire l'effet du « paradoxe de l'observateur » et d'accéder aux situations les plus écologiques possibles, nous ne sommes pas présente lors des enregistrements. Notre absence permettra aux familles de rester alors dans un cadre intime, de leur quotidien qui permet d'obtenir des données dans leur contexte social ordinaire et routinier (Mondada, 2012). Nous

fournissons le kit d'enregistrement et le mode opératoire aux familles, avec des fiches de suivi que les familles devront remplir¹ pour chaque enregistrement effectué. Dans ces fiches, nous demandons de noter les participants à l'enregistrement, la date, le lieu et ce qui se passe durant l'enregistrement. Les familles peuvent préciser s'il y a des passages qu'elles souhaiteraient ne pas voir apparaître dans le corpus. Ces informations serviront à remplir les métadonnées des enregistrements.

En ce qui concerne le rythme de collecte des enregistrements, nous tentons de réduire le temps d'emprunt du matériel par les familles, ce qui n'est pas toujours aisé. Nous laissons tout de même des libertés aux familles pour ne pas qu'elles se sentent pressées, ni opprimées par nos demandes. Cette démarche permet de pas ou peu impacter le comportement des familles pour la réalisation des enregistrements mais présente l'inconvénient de ne pas avoir des temps de collecte homogènes d'une famille à l'autre.

Nous demandons aux familles d'enregistrer au moins 4 scènes par période représentant des moments clefs de la vie de leurs enfants, tels que les repas, les moments de jeux, les devoirs, le rituel de la douche, etc. avec les personnes que les enfants côtoient fréquemment.

Voici les caractéristiques des enregistrements récoltés pour la première et la deuxième période :

¹ Ces informations auraient pu être enregistrées si nous avons rencontré des familles dans l'impossibilité de remplir la fiche.

Famille	Période 1 (P1)			Période 2 (P2)		
	Scène du quotidien	Participants	Durée	Scène du quotidien	Participants	Durée
FAM1 (Marise)	Playmobil	- Marise -Tante - Cousine - Mère	00:13:45	Devoirs	- Marise - Tante	00:19:01
	Devoirs	- Marise - Tante	00:17:43	Conversation	- Marise - Mère - Sœur - Père	00:01:15 00:08:22
	Film à la télévision	- Marise - Tante - Cousine	00:14:00 00:17:46	Coloriage	- Marise - Tante	00:11:19
	Cadeaux de Noël	- Marise - Tante	00:11:56	Matinée à l'école	- Marise - Elèves - Maîtresse	01:02:19 01:03:15
	Matinée à l'école	- Marise - Elèves - Maîtresse	01:23:46 01:08:31	Récréation à l'école	- Marise - Elèves - Maîtresse	00:23:08
	TOTAL P1	Nb de participants : 5+	03:47:27	TOTAL P2	Nb de participants : 6+	03:08:39
FAM2 (Fréïhr et Briséis)	Goûter	- Fréïhr - Sœur - Briséis - Mère	00:15:58	Repas en famille	- Fréïhr - Parents - Briséis - Grands- - Sœur parents	00:01:26 00:08:54
	Repas en famille	- Fréïhr - Parents - Briséis - Grands- - Sœur parents	00:27:00	Douche Fréïhr	- Fréïhr - Mère	00:09:58
	Devoirs Briséis	- Briséis - Mère	00:32:22	Repas et jeux	- Briséis - Ami - Sœur	00:21:41
	Jeu de société	- Briséis - Parents - Sœur	00:52:23			
	Déjeuner	- Fréïhr - Mère	00:42:21			
	Puzzle	- Fréïhr - Arrivée - Sœur Briséis	00:22:27			
	TOTAL P1	Nb de participants : 7	03:12:31	TOTAL P2	Nb de participants : 8	00:41:59
FAM3 (Roxane)	Repas en famille	- Roxane - Mère - Sœur - Père	00:26:14	Jeux	- Roxane - Mère - Sœur - Père	00:23:05
	Dessin sur tableau	- Roxane - Mère - Sœur	00:06:50	Dessin	- Roxane - Mère	00:13:11
	Repas	- Roxane - Mère - Sœur	00:06:06 00:17:53	Jeux	- Roxane - Mère - Sœur	00:11:26
	Jeux	- Roxane - Mère - Sœur - Père	00:21:58	Ordinateur	- Roxane - Mère - Sœur	00:33:14
	TOTAL P1	Nb de participants : 4	01:12:55	TOTAL P2	Nb de participants : 4	01:20:56
FAM4 (Lucile et Camille)	Repas en famille	- Lucile - Mère - Camille - Père	00:48:13	Repas	- Lucile - Mère - Camille - Père	1:11:41
	Joue seule	- Lucile - Camille joue seule - Mère	01:29:05 01:29:05	Douche	- Lucile - Mère - Camille	00:44:17
	Jeux	- Lucile - Mère - Camille - Père	01:11:08	Jeux	- Lucile - Mère	00:55:49
	Dessin spirographe	- Lucile - Mère - Camille	01:07:26	Jeux	- Lucile - Mère - Camille	01:04:10
				Jeux	- Lucile - Mère - Camille	00:31:47
	TOTAL P1	Nb de participants : 4	06:04:57	TOTAL P2	Nb de participants : 4	04:27:44
Durée totale de la période 1 = 14:17:50			Durée totale de la période 2 = 09:39:18			

Nous obtenons une première période de 14h17 et une deuxième période de 9h39. La différence de durée peut s'expliquer par la non-maîtrise du matériel de la part des familles. En

effet, deux familles ont eu des difficultés à se servir de l'enregistreur et/ou la caméra lors de la deuxième période. Nous obtenons alors des enregistrements plus courts que prévu.

Les types de situations recueillis sont relativement variés : des repas (goûter, repas de famille), différents types de jeux (solitaires, jeux de sociétés, dessins, puzzle), des temps scolaires, les devoirs, des bains/douches. Par ailleurs, d'une période à l'autre, nous recueillons des scènes relativement semblables ce qui va nous permettre une comparaison intéressante.

Les enregistrements du quotidien ne sont pas les seules données langagières que nous récoltons. Des tests de langage sont effectués pour situer les capacités langagières de notre échantillon mais aussi pour provoquer des contextes de liaisons similaires d'une période à l'autre.

1. Tests de langage

Nous effectuons deux séries de tests, la première composée de tests de langage pour situer l'enfant dans son acquisition langagière et une deuxième série pour évaluer la manière dont la liaison se met en place. Pour cette dernière série, nous présenterons des résultats préliminaires de trois enfants : Roxane, Lucile et Camille (les trois enfants pour lesquels nous avons toutes les données sur les trois périodes).

1. Tests de positionnement

A la suite de chaque période d'enregistrements, les enfants ont effectué des tests de langage issus de DyLNet (Nardy et al. 2016). Il s'agit de vérifier l'étendue du lexique de l'enfant, de contrôler sa mémoire des chiffres, de connaître ses compétences sociolinguistiques en perception et en production, de voir s'il parvient à détecter les variations stylistiques en fonction des différents interlocuteurs possibles et enfin de connaître la compréhension syntaxicosémantique. Nous réalisons ces tests avec les enfants de notre panel les plus petits (Fréïhr, Briséis, Roxane, Lucile et Marise) puisque ces tests ont été construits pour le niveau « maternelle ». Nos données pourront être comparées aux données DyLNet, recueillies sur toute une école maternelle de Grenoble.

2. Tests de liaison

a. Pour tous les enfants

La deuxième série de tests (empruntée à Dugua, 2006) porte sur la production de liaisons. En premier lieu, nous effectuons une tâche d'apostrophe où les enfants doivent

appeler, à deux reprises et les uns après les autres, des animaux en disant par exemple « ours ours viens jouer avec moi ! ».

Cette tâche a pour but de voir la forme lexicale la plus disponible pour les enfants : à voyelle initiale « ours » ([uʁs]) ou à consonne initiale « nous » ([nuʁs]) issu de [œnuʁs], « tours » ([tuʁs]) issu de [pətituʁs]. Les résultats laissent apparaître que les trois enfants aux périodes 2 et 3, produisent les formes à voyelle initiale. Pour la période 1 (voir Figure 3), Lucile dispose clairement de la forme à voyelle initiale. Pour Camille, la production avec déterminant correspond à la forme [pətitan] avec liaison, mais nous supposons qu'elle possède la forme à initiale vocalique puisqu'elle a produit [an] en isolation lors de la seconde production. Quant à Roxane, nous voyons que la forme à initiale vocalique n'est pas forcément acquise, les formes avec déterminant sont en l'+ « nom » ([lekyʁœj] et [lan]).

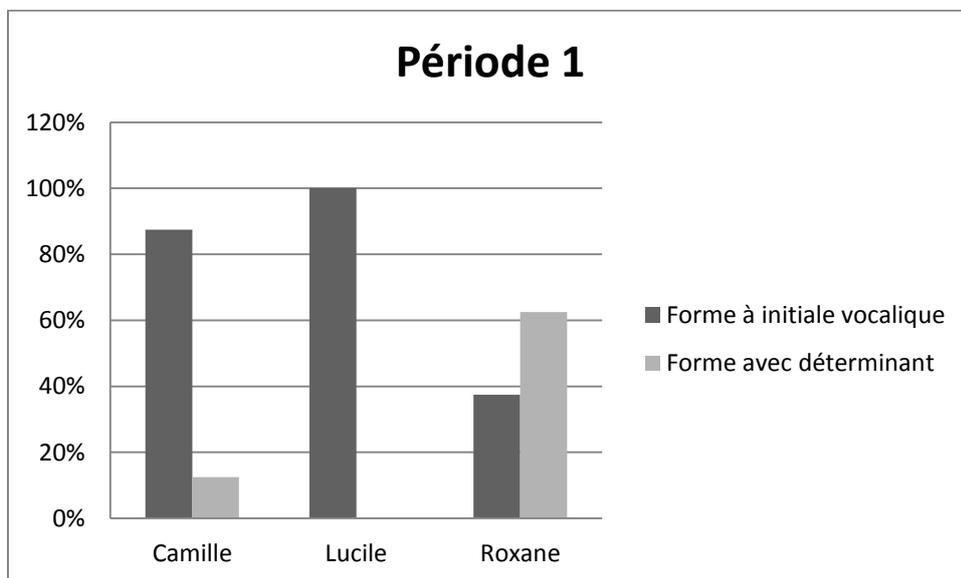


Figure 3: production de la tâche d'apostrophe (Période 1)

Dans un second temps, nous faisons une tâche avec des pseudo-noms à voyelle initiale (4 pseudo-noms), où nous présentons chaque pseudo-nom précédé du déterminant « un », par exemple [œnikat], l'enfant doit alors produire la forme « deux » + « pseudo-nom », par exemple [døzikat]. Nous relevons trois types de productions :

- Production de type maintien : nous disons [œnivak] l'enfant produit [dønivak]
- Production de type alternance : nous disons [œnivak] l'enfant produit [døzivak]
- Production de type omission : nous disons [œnivak] l'enfant produit [døivak]

Le graphique ci-dessous rend compte de la répartition des types de productions (4 productions au total) pour chaque enfant et à chaque période.

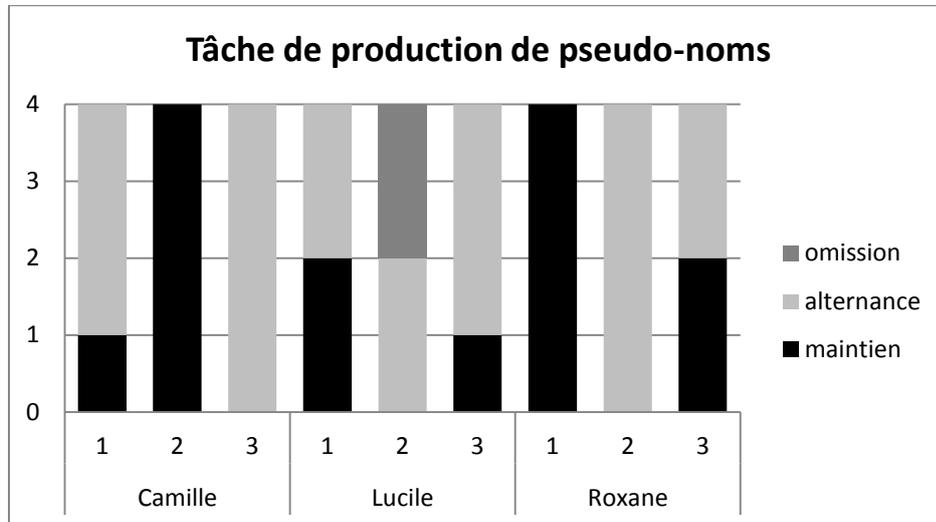


Figure 4 : Résultats de la tâche de pseudo-noms

Les trois enfants observées présentent des comportements variables sur les formes d'alternances et de maintien que nous ne pouvons interpréter pour l'instant. Seule Lucile produit des formes en omissions lors de la deuxième période.

Par la suite, nous effectuons une tâche de dénomination d'images en contexte de liaisons obligatoires après les Mots1 « un » et « deux », par exemple [œnuks]. Trois types de production sont attendus :

- Production de type liaison juste :
 - Après « un » : [œnuks]
 - Après « deux » : [døzuxs]
- Production de type remplacement :
 - Après « un » : [œzuks]
 - Après « deux » : [dønuks]
- Production de type omission :
 - Après « un » : [œuks]
 - Après « deux » : [døuks]

Le graphique ci-dessous montre les taux de chaque type de production pour chaque enfant et à chacune des trois périodes.

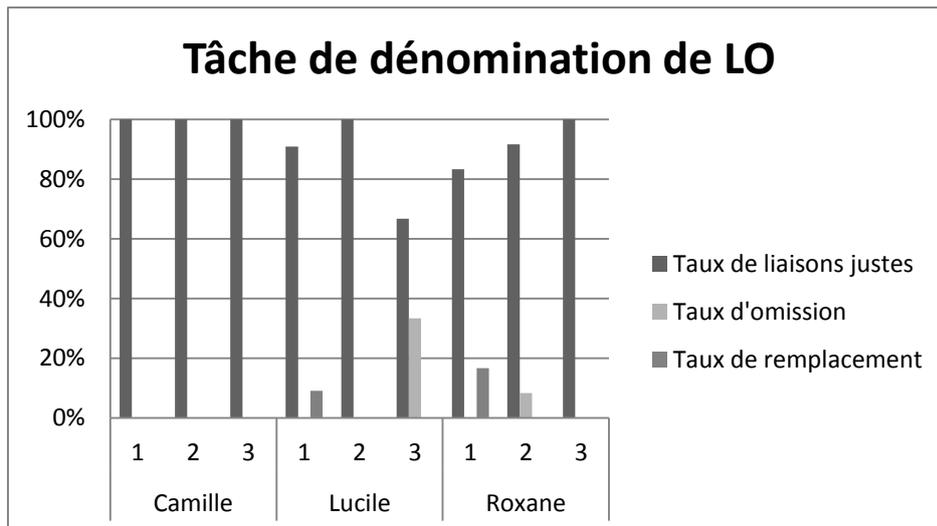


Figure 5 : Résultats de la tâche de dénomination de liaisons obligatoires

Pour Camille, qui est plus âgée, nous retrouvons des productions de liaisons justes systématiquement pour les trois périodes. Tandis que pour les plus petites, Lucile et Roxane, nous relevons des erreurs par remplacement lors de la première période, qui disparaissent par la suite. Des erreurs d'omission surviennent en période 2 et/ou 3, ce sont des erreurs qui peuvent subsister jusqu'à 6 ans (Dugua, 2006). Ces résultats sont en accord avec les données dont nous disposons sur les productions de liaisons.

Dans le même principe, la tâche suivante est également une dénomination d'images mais cette fois en contexte de liaison facultative après les Mots1 « petit » et « gros ». Trois types de productions sont attendus :

- Production de type liaison réalisée :
 - Après « petit » : [pətituks]
 - Après « gros » : [gʁozuks]
- Production de type non réalisée (après les LF les liaisons ne sont pas obligatoires, la non production de liaison n'est donc pas une erreur) :
 - Après « petit » : [pətiuks]
 - Après « gros » : [gʁouks]
- Production de type remplacement :
 - Après « un » : [pətizuks]
 - Après « deux » : [gʁonuks]

Dans le graphique ci-dessous, nous mettons en évidence les trois types de productions pour chaque enfant et à chacune des trois périodes.

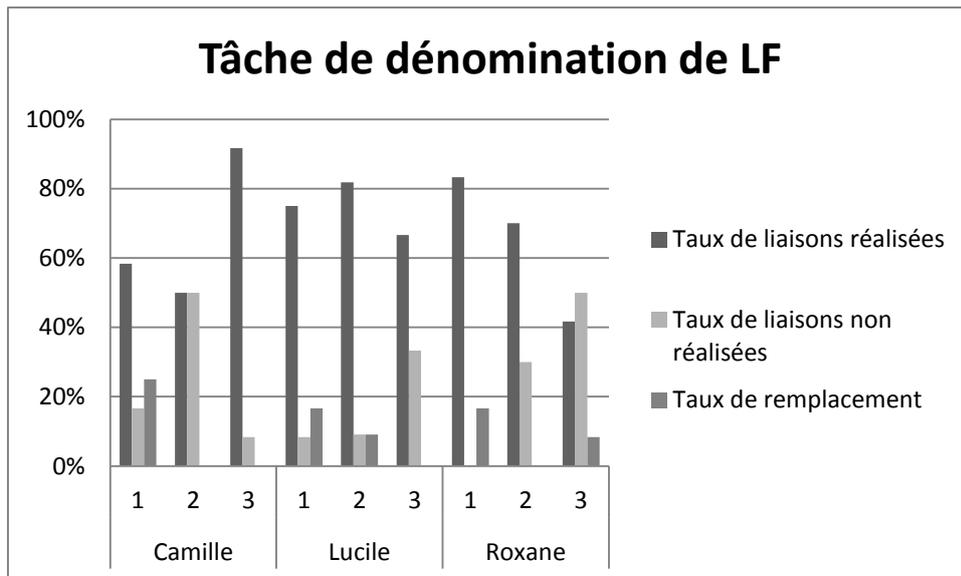


Figure 6 : Résultats de la tâche de dénomination de liaisons facultatives

D'après ces résultats, nous retrouvons globalement des taux de réalisations de LF nettement supérieurs à Dugua (2006), nous sommes ici en moyenne à 69% de liaisons réalisées contre 20,5% relevées chez Dugua (2006). Nous constatons des comportements bien différents chez les trois enfants : des taux de réalisation relativement stables et hauts pour Lucile, ces mêmes taux qui baissent pour Roxane et qui stagnent puis augmentent fortement pour Camille. L'analyse de ces résultats sera approfondie par la suite.

Enfin, la dernière tâche consiste à faire dénommer quatre noms à initiale consonantique (« nombril », « lavabo », « nuage », « zèbre ») après « un » et « deux ». Ces mots présentent la particularité d'avoir en initiale une consonne qui pourrait être une consonne liaison ou d'élision. Chez nos trois enfants, quand elles produisent les mots cibles, elles les réalisent correctement. Seules deux occurrences en période 1 présentent des variations intéressantes : [œnavabo] (Roxane) et [døzɔ̃bɔil] (Lucile). Dans ces erreurs, la consonne initiale produite correspond à la consonne de liaison attendue après les déterminant « un » et « deux ». Dugua (2006) interprète ce type de réalisation comme des surgénéralisations du phénomène de liaison. Dans ses données, elles apparaissent principalement vers 4-5 ans. Nous les retrouvons un an plus tôt chez Lucile et Roxane

D'une manière générale, les performances que nous obtenons chez ces trois enfants sont supérieures à celle de l'échantillon dans Dugua (2006). Ces résultats seront détaillés dans de prochains travaux.

b. Pour les enfants lecteurs

Lorsque les enfants sont en âge de lire (Marise, Briséis, Camille), ils effectuent deux tests supplémentaires : un test de lecture et un test d'histoire sans parole.

Pour le premier, les enfants doivent lire une série de 32 phrases avec des mots et contextes de liaisons choisis avec précautions. Pour le second test, la création d'une histoire sans parole favorisant les contextes de liaison s'est révélée un bon exercice pour faire parler les enfants à partir de dessins. Nous avons créé ce test pour que les enfants puissent être amenés à raconter une histoire comme elle pourrait être écrite en insérant des images favorisant les contextes de liaison. Ce démarche d'histoire sans parole peut être déjà connue des enfants, de même qu'ils ont déjà dû être amenés à raconter des histoires qui leur ont été lues par le passé.

Ces tests ont une grande importance car ils nous permettent de contrôler, dans des contextes similaires et ce, d'une période à l'autre, la réalisation de liaisons obligatoires et de liaisons facultatives et cela dans des modalités variées. Avec les données des tests recueillies, croisées avec les données « naturelles » que nous avons captées, nous serons en mesure de mieux comprendre comment s'acquiert la liaison dans des milieux et des situations variés.

Conclusion et perspectives

La création de notre corpus ESLO-Enfants permet d'élargir le périmètre du corpus ESLO2 tout en s'intégrant dans son architecture, mais permet aussi d'apporter de nouvelles données aux corpus enfantins déjà existants.

Outre les questions relatives à la mise en place du langage chez les jeunes enfants, ce corpus vise à fournir un cadre variationniste dans lequel nous pourrions décrire la diversité et la richesse des inputs dans lesquels les enfants évoluent. A travers l'étude de l'usage des liaisons à la fois obligatoires et variables, dans la parole environnante, comme dans la parole infantine, nous chercherons à mettre en évidence les liens entre l'usage des parents/fratrie/famille et l'usage des enfants, et la façon dont adultes et enfants adaptent leurs usages à la situation et aux participants. Des entretiens semi-directifs ont été réalisés (mais n'ont pas été présentés ici par manque de place) et permettront d'avoir un regard plus approfondi sur la vie au sein des familles, mais également d'avoir accès à l'usage langagier des parents dans un contexte plus formel que les scènes quotidiennes. La méthode que nous mettrons en place prendra deux orientations : (1) étant donné l'échantillon important recueilli, transcrit et annoté en liaisons, nous pourrions dégager des tendances statistiques (taux de

liaisons selon les situations, selon les participants, etc.). Mais, (2) dans le but de rendre compte des influences de l'input sur l'output, il nous semble indispensable d'observer les usages des liaisons avec une approche plus qualitative en repérant, par exemple, des constructions fréquentes dans une famille et en cherchant à voir comment elles sont transmises et reprises ou non par les enfants. Ces deux méthodes d'observation et d'analyse, croisées avec les métadonnées sur les familles d'une part (avec l'échelle SLCP), et sur le développement langagier des enfants d'autre part (avec les tests de langage) devraient nous permettre de mieux comprendre la façon dont le langage se construit dans la diversité des environnements et des situations que rencontre un enfant.

Bibliographie

Abouda L., & Baude O. (2005). « Constituer et exploiter un grand corpus oral: choix et enjeux théoriques. Le cas des ESLO. », *Corpus en Lettres et Sciences sociales, Des documents numériques à l'interprétation*, Albi, France.

Baude O. & Dugua C. (2011). « (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? » *Corpus n°10* : 99-118.

Baude O. & Dugua C. (2015). *Usage de la liaison dans le corpus des ESLOs : vers de nouveaux (z) ouvrages de référence ?* Dostie, Hedermann (Eds.). La dia-variation en français actuel. Bern : Peter Lang, 349-371.

Blanc, M. & Biggs, P. (1971). « L'enquête socio-linguistique sur le français parlé à Orléans » *Le français dans le monde*, n°85 : 16-25.

Bourdieu P. (1982). *Ce que parler veut dire : L'économie des échanges linguistiques*. éd. Fayard.

Bourdieu P. (1994). *Raisons pratiques : sur la théorie de l'action*. Paris : éd. du Seuil.

Bourdieu P. (1979). *La distinction. Critique sociale du jugement*. Paris : éd. de Minuit.

Bybee J. & Hopper P. (2001). *Introduction*. Bybee J. & Hopper P. (ed.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam / Philadelphia, Benjamins.

Cameron-Faulkner T., Lieven E. et Tomasello M. (2003). *A construction based analysis of child directed speech*. *Cognitive Science*, volume 27, n°6, 843-873.

Chevrot J.-P., Chabanal D., Dugua C. (2007). « Pour un modèle de l'acquisition des liaisons basé sur l'usage : trois études de cas » *Journal of French Language Studies* n°17, 103-128.

Chevrot J.-P., Dugua C., Fayol M. (2009). « Liaison Acquisition, Word Segmentation and Construction » *Fench : A Usage Based Account*. *Journal of Child Language* n°36, 557-596.

- Coquillon A., Durand J., Eychenne J. (2010). « La liaison dans PFC : les chiffres et leur interprétation » Atelier PFC. University Tulane, New Orleans., Jul 2010, Nouvelle Orléans, États-Unis.
- De Jong D. (1994). *La sociophonologie de la liaison orléanaise*. Lyche, C. (Ed.), *French Generative Phonology : Retrospective and Perspectives*, 95-129, Salford : ESRI.
- Dugua C. (2006). « Liaison, segmentation lexicale et schémas syntaxiques entre 2 et 6 ans - Un modèle développemental basé sur l'usage ». Thèse de doctorat, Université Stendhal, Grenoble.
- Dugua C. & Baude O. (2017). « La liaison à Orléans, corpus et changement linguistique: une première étude exploratoire », *Journal of French Language Studies* n°27, 41-54.
- Durand J. & Lyche C. (2008). « French liaison in the light of corpus data. » *French Language Studies* n°18, 33–66.
- Durand J., Laks B., Calderone B. & Tchobanov A. (2011). « Que savons-nous de la liaison aujourd'hui ». *Langue française* n°169, 103-135.
- Encrevé P. (1988). *La liaison avec et sans enchainement : phonologie tridimensionnelle et usages du français*. Paris : Seuil.
- Fijalkow J. (1997). « Entrer dans l'écrit : oui, mais par quelle porte ? ». *Repères, recherches en didactique du français langue maternelle* n°15, 113-129.
- Gadet F. (1989). *Le Français ordinaire*. Paris: Armand Colin.
- Gadet F. (1997). *Classe sociale*. M.L. Moreau (dir.). *Sociolinguistique. Concepts de base*. Liège : Mardaga.
- Gagnier N. (2011). « Le jugement des autres sur nos attitudes parentales ». La Presse. Mis en ligne le 22 mars 2011. Consulté le 27 juillet 2019.
- Goldin-Meadow S. (2010). « Rôle des gestes dans la création et l'acquisition du langage. » *Enfance*. Vol.62 n°3, 347-360.
- Institut National de la Statistique et des Etudes Economiques (2019). <https://www.insee.fr/fr/metadonnees/pcs2003/categorieSocioprofessionnelleAgregree/1>
- Kemmer S. & Barlow M. (2000). *Introduction : A usage-based conception of language*. Barlow, M. & Kemmer, S. (Ed.), *Usage-based models of language use*. Stanford Californie: CSLI Publications, 7-28.
- Labov W. (1976). *Sociolinguistique*. Paris : Editions de Minuit.
- Laks B. (2005). « La liaison et l'illusion. » *Langages* 39^e année n°158, 101-125.
- Liégeois L. (2014). « Usage des variables phonologiques dans un corpus d'interactions naturelles parents-enfant : impact du bain linguistique et dispositifs cognitifs d'apprentissage. » Thèse de doctorat, Université Blaise Pascal, Clermont 2.

Linard M. & Prax I. (1984). *Images vidéo, images de soi ou Narcisse au travail*. Paris : Dunod.

Mondada L. (2012). *The conversation analytic approach to data collection*. J. Sidnell & T. Stivers (Ed.) *Handbook of Conversation Analysis*. Oxford : Blackwell-Wiley.

Nardy A., Fleury É., Chevrot J.-P., Karsai M., Buson L., Bianco M., Rousset I., Dugua C., Liégeois L., Barbu S., Crespelle C., Busson A., Léo Y. & Bouchet H. (2016). DyLNet – Language Dynamics, Linguistic Learning, and Sociability at Preschool: Benefits of Wireless Proximity Sensors in Collecting Big Data (ANR-16-CE28-0013). <https://dylnet.univ-grenoble-alpes.fr/>

Neill A.S. (1966). *La Liberté, pas l'anarchie*. Hart Publishing (New York) 1966, Payot.

Queneau R. (1950). *L'Avenir du français*, Bâtons, chiffres et lettres. Paris : Gallimard.

Soum-Favaro C., Coquillon A., Chevrot J.-P. (2014). *La liaison : approches contemporaines*. Bern : Peter Lang.

Tomasello M. (2003). *Constructing a Language : A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Référence des corpus :

Corpus de Lyon : <http://childes.psy.cmu.edu/access/French/Lyon.html>

Laboratoire Ligérien de Linguistique - UMR 7270 (LLL) (2017). *ESLO [Corpus]*. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, <https://hdl.handle.net/11403/eslo/v1>.

Laboratoire de Recherche sur le Langage - EA 999 (LRL) (2017). *ALIFE (Acquisition de la Liaison et Interactions Parents Enfants)* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, <https://hdl.handle.net/11403/alife-000853/v1.1>.

Communication Langagière chez le Jeune Enfant (CoLaJE) : <http://colaje.scicog.fr/>

Morgenstern, A & Parrisé, C. (2012). The Paris Corpus. *French Language Studies* 22(1), 7-12, Cambridge University Press.

Parrisé C., Morgenstern A. (2010). Transcrire et analyser les corpus d'interactions adulte-enfant. Edy Veneziano, Anne Salazar Orvig, Josie Bernicot. *Acquisition du langage et interaction*, L'Harmattan, pp.201-222

Traitement des Corpus Oraux en Français (TCOF): <http://www.cnrtl.fr/corpus/tcof/>
Emmanuelle Canut, Virginie André, Bertrand Gaiffe. Mise à disposition de corpus oraux interactifs : le projet TCOF (Traitement des Corpus Oraux en Français). *Pratiques : théorie, pratique, pédagogie*, 2010, Interactions et Corpus Oraux, pp.147-148. <[hal-00523397](https://hal.archives-ouvertes.fr/hal-00523397)>