



**HAL**  
open science

# L’usage des liaisons lors de lectures partagées – Une étude exploratoire à partir du module “ Livres pour enfants ” d’ESLO

Céline Dugua, Flora Badin, Barbara Fallon, Olivier Baude

## ► To cite this version:

Céline Dugua, Flora Badin, Barbara Fallon, Olivier Baude. L’usage des liaisons lors de lectures partagées – Une étude exploratoire à partir du module “ Livres pour enfants ” d’ESLO. 8e Congrès Mondial de Linguistique Française (CMLF 2022), Jul 2022, Orleans, France. pp.09006, 10.1051/shsconf/202213809006 . hal-04106450

**HAL Id: hal-04106450**

<https://hal.parisnanterre.fr/hal-04106450v1>

Submitted on 25 May 2023

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# L'usage des liaisons lors de lectures partagées – Une étude exploratoire à partir du module « Livres pour enfants » d'ESLO

Céline Dugua<sup>1\*</sup>, Flora Badin<sup>1</sup>, Barbara Fallon<sup>1</sup> et Olivier Baude<sup>2</sup>

<sup>1</sup>LLL UMR7270, Université d'Orléans

<sup>2</sup>MODYCO UMR7114, Université de Nanterre

**Résumé.** Les lectures partagées constituent une forme particulière de discours adressé aux enfants. Nous inscrivons dans les approches basées sur l'usage, nous défendons l'idée de l'importance de cette pratique dans la construction du langage chez l'enfant et dans la perspective de son entrée dans la littératie. À partir du module « *Livres pour enfants* » du corpus ESLO, nous observerons l'usage des liaisons au regard des caractéristiques du lecteur, des enfants, des types d'albums lus et d'un ensemble de facteurs internes. Cette étude est exploratoire dans la mesure où la taille de notre corpus nous permet d'observer seulement certains de ces facteurs. Les premières tendances obtenues et les réflexions sur les aspects méthodologiques à creuser constituent un point d'étape important pour de futurs travaux autour de la lecture partagée et de l'usage des liaisons dans le domaine de l'acquisition du langage.

**Abstract.** *Liaisons in shared reading – An exploratory study of the ESLO module « Livres pour enfants ».* Shared reading is a particular form of child directed speech (CDS). In line with usage-based approaches, we defend the idea of the importance of this practice in the construction of language in children and in the perspective of their entry into literacy. Based on the module "*Livres pour enfants*" of the ESLO corpus, we will observe the use of liaisons with regard to the characteristics of the reader, the children, the types of books read and a set of internal factors. This study is exploratory in the sense that the size of our corpus allows us to observe only some of these factors. The first trends obtained and the insights on the methodological aspects to be explored constitute an important step for future work on shared reading and the use of liaisons in the field of language acquisition.

## 1 Introduction

### 1.1 La lecture partagée comme forme de discours adressé aux enfants

Dans le cadre de l'acquisition, de nombreux travaux, notamment ceux issus des approches basées sur l'usage, montrent l'importance de la parole entendue (input) sur la construction du langage chez l'enfant (Cameron-Faulkner et al., 2003; Chenu & Jisa, 2006; Hoff, 2002; Snow, 1995). L'input regroupe tout ce que l'enfant perçoit autour de lui et qui lui sert de

\* Corresponding author : [celine.dugua@univ-orleans.fr](mailto:celine.dugua@univ-orleans.fr)

socle pour développer son langage. Ces formes entendues sont déterminantes puisque c'est dans l'input que l'enfant va avoir accès à l'usage, comme le soulignent Ambridge et Lieven (2011, p. 2) « most constructivist approaches are input-based approaches in that they assume that characteristics of the input are a driving force in children acquisition ».

Le discours adressé à l'enfant (ou *Child Directed Speech* CDS) constitue l'une des manifestations de l'input. Il correspond à la façon particulière que les adultes adoptent pour s'adresser aux enfants (Snow, 1999). Les particularités de ce langage, par rapport au langage entre adultes, affectent tous les niveaux langagiers. D'un point de vue prosodique, la fréquence fondamentale est plus haute et ses variations au sein d'un énoncé importantes ; concernant le lexique, on trouve plus de redondance (Cameron-Faulkner & Hickey, 2011) et une richesse et une diversification qui s'élaborent progressivement (Snow, 1977) ; au niveau syntaxique, on relève par exemple des structures répétitives et beaucoup de questions (Cameron-Faulkner et al., 2003) ; l'ancrage dans le « ici et maintenant » (Snow, 1972), un nombre important de reprises, de reformulations des énoncés enfantins (Harris, 1990) sont parmi les caractéristiques pragmatiques du CDS. Enfin, on sait que le CDS est un usage qui évolue, notamment en fonction de l'âge de l'enfant (Foulkes et al., 2005).

Dans ce papier, nous nous intéressons à une forme particulière de CDS : celle qui résulte de la lecture partagée. Cette pratique consiste en une lecture à voix haute destinée à des enfants non lecteurs (Frier, 2006). Les ouvrages lus peuvent être des albums, des imagiers, des magazines par exemple. Lors des lectures partagées, le lecteur et l'enfant se trouvent dans une situation d'attention conjointe (Tomasello, 2008) que l'on sait être propice au développement langagier, et notamment lexical (Tomasello & Farrar, 1986). Une autre particularité qui distingue cette pratique des autres formes de CDS est la présence d'un support écrit, accompagné d'illustrations. La présence d'une forme écrite détermine le type de langage adressé à l'enfant, même si, en dehors de la lecture à proprement parler, des interactions ont souvent lieu durant ces moments partagés, et même si chaque lecteur aura sa propre façon de s'approprier l'album.

La lecture partagée est pratiquée dans de nombreuses familles (par des parents, par la fratrie : Bastide & Joigneaux, 2014), dans les structures de la petite enfance (crèche, assistant maternel : Agence Quand les livres reliait, 2018), à l'école maternelle (enseignant : Bonnéry, 2014 ; Cèbe & Sève, 2014, Boiron, 2015) et dans les bibliothèques (par des personnels formés : Alamichel, 2011). On trouve également nombre d'associations qui promeuvent cette pratique. D'un point de vue didactique, ce thème est bien documenté. Il porte souvent sur la lecture en maternelle et tend à montrer les différents styles de lecture (Mol et al., 2008), la façon dont le lecteur peut aider les enfants à comprendre le texte (Boiron, 2014), le rôle des illustrations (Melançon, 2015), les échanges qui peuvent se mettre en place autour d'un album (Reese et al., 2003). Ces aspects sont cruciaux puisqu'on sait que la participation régulière et fréquente à des lectures développe à la fois les capacités langagières orales des enfants mais participe aussi à l'entrée dans l'écrit (Kalb & van Ours, 2014) et aux compétences de narration (Sénéchal et al, 2008). Des sociologues se sont également emparés de cette question, à la fois sur les types d'albums favorisés, dans les cercles familiaux ou scolaires, sur les types de lectures plus ou moins proches du support écrit et sur la place de cette pratique à l'école maternelle (Bonnéry & Joigneaux, 2015).

En revanche, la lecture partagée est peu décrite avec les outils de la linguistique et dans une optique d'acquisition. Cette question est au cœur de notre propos et doit selon nous rendre compte de deux éléments cruciaux : d'une part, les caractéristiques linguistiques et littéraires des albums jeunesse – en quoi ils se distinguent linguistiquement du CDS – et d'autre part, la façon dont le lecteur se les approprie pour les oraliser devant l'enfant. Des recherches anglosaxonnes se sont penchées sur le premier point, le second fera l'objet

d'une section à part entière (voir section 1.3). Le travail de Cameron-Faulkner et Noble (2013) met en évidence les différences syntaxiques entre ce qui figure dans les albums et le CDS. À partir de l'étude de 20 albums destinés aux enfants de 2 ans (les « best-sellers » sur deux journées de l'année 2011), les auteurs observent que les formes Sujet-Prédicat et les structures complexes sont plus fréquentes dans les livres que dans le CDS. Dawson et al. (2021) relèvent quant à eux les particularités lexicales des albums, et trouvent par exemple davantage de « mots rares » dans les albums que dans le CDS, davantage de noms et adjectifs dans les albums et plus de pronoms dans le CDS. Ces auteurs « found that the words used in children's books are typically more diverse, more sophisticated, and lexically denser than those children hear via conversation » (Dawson et al., 2021, p. 22). Le langage auquel les enfants ont accès par les lectures d'albums se distingue donc du CDS à la fois par des propriétés lexicales et syntaxiques.

En français, peu de données de ce type sont disponibles. Signalons toutefois le travail de Pustka (2017) qui propose une classification de son corpus de livres lus (albums, contes) à partir de la présence ou de l'absence de traits liés à la distance ou à l'immédiateté : avec du côté de la distance l'usage du passé simple, de la double négation, du « nous », de l'inversion dans les interrogatives ; et du côté de l'immédiateté l'absence de passé simple, la simple négation, l'usage du « on » et des interrogatives basées sur l'intonation.

Nous proposons, à partir de notre corpus d'étude présenté plus loin, de rendre compte de la prégnance des contextes de liaison dans les albums en comparaison avec le CDS du corpus de Liégeois (2014). Il s'agit ici de voir si le nombre de contextes de liaison diffère entre les écrits pour enfants et le CDS. Précisons qu'ici nous ne nous intéressons pas à la question de la réalisation ou non des liaisons, mais seulement à la présence des contextes de liaison. Dans le corpus de Liégeois, correspondant à des interactions parents/enfants, dans trois familles, les passages de discours adressés à l'enfant et de discours entre adultes sont annotés. Liégeois propose un ratio entre le nombre de contextes de liaison et le nombre de mots produits, par les parents, dans le cadre du CDS. Il obtient, selon les familles, entre un contexte de liaison tous les 19 mots et un contexte de liaison tous les 22.9 mots. Dans notre corpus de lecture partagée, en appliquant le même ratio, nous trouvons un contexte de liaison tous les 20 mots (le nombre total de mots produits par l'ensemble des lecteurs est de 24.824, et le nombre de contextes de liaison potentielles chez ces mêmes lecteurs est de 1.239). Le fait que le nombre de contextes de liaison soit tout à fait comparable entre des situations d'interaction et de lectures partagées souligne que sur ce critère, modalités orale et écrite ne se distinguent pas.

## 1.2 Un objet linguistique variable : les liaisons

La liaison est un phénomène de sandhi externe largement étudié en français. Elle se caractérise par la production d'une consonne dite de liaison (/n/, /z/, /t/ pour les plus fréquentes) entre deux mots, le premier se terminant par une consonne graphique muette s'il est isolé (*grand, un*) et le second commençant par un son vocalique (*arbre, habit*). Par exemple, le /n/ de *un* apparaîtra dans le contexte *un enfant, un habit*, mais pas dans *j'en veux un* ou *un bureau*.

Précisons ici que le fonctionnement des liaisons s'inscrit dans l'histoire de la langue, à la fois l'évolution de sa prononciation et de son orthographe. Langlard (1928) distingue quatre étapes qui ponctuent ces changements :

- Période du Moyen-Âge : prononciation de toutes les consonnes finales. On ne trouve donc pas de liaison.
- Période du XIV-XV<sup>ème</sup> siècle : amuïssement de certaines consonnes finales. Il s'ensuit que, pour un mot donné, différentes prononciations étaient possibles selon

le contexte. Par exemple, l'amuissement affectait la consonne finale devant un mot à consonne initiale, en revanche, lorsque le mot suivant était à initiale vocalique, la consonne finale subsistait dans la prononciation et s'agglutinait à la voyelle dudit mot.

- Période du XVI-XVIIème siècle : aboutissement du processus de disparition des consonnes finales. Elles cessent pratiquement toutes d'être prononcées systématiquement, sauf dans des « locutions toutes faites » (Langlard, 1928, p. 22).
- A partir du XVIIème siècle : restitution des consonnes finales dans la graphie. C'est à cette époque que l'orthographe va commencer à se standardiser. On passe de ce que Catach (1973, p. 11) appelle une « orthographe pure [...] une sorte d'aide-mémoire plus ou moins instable » à une orthographe normalisée. Entre autres, les consonnes finales qui, à l'oral, poursuivent leur disparition, vont être peu à peu restituées dans la graphie. Au vu de ce processus, la graphie s'éloigne de la phonologie de la langue utilisée à cette époque. Les principaux arguments pour rétablir les consonnes finales dans l'orthographe étaient des arguments étymologiques et de maintien paradigmatique (Laks, 2005). Ainsi, l'orthographe a fixé la forme des mots avec consonne finale, c'est pourquoi l'usage des liaisons est tel qu'on le connaît aujourd'hui.

Ce point historique permet de comprendre certains éléments essentiels du fonctionnement de la liaison en français à savoir que cette dernière, connue pour être phonologique, a été influencée dans son histoire par la standardisation de l'orthographe. Aujourd'hui, l'orthographe joue encore un rôle essentiel dans le processus et la diffusion de la liaison, et c'est ce que nous pourrions observer dans cette étude en observant son usage en lecture.

Les contextes de liaison sont classiquement organisés en trois catégories. Nous nous référons aux travaux de De Jong (1988, 1994) et Booij et De Jong (1987) qui, à partir de l'étude des liaisons dans des entretiens issus du corpus d'Orléans, adoptent une approche descriptive de la classification des liaisons. En suivant ces auteurs, nous considérons comme liaisons obligatoires ou systématiques celles qui sont toujours réalisées, à savoir après déterminant (*un autre jour, des épices*), entre pronom et verbe (*on écoute, ils arrivent*), entre verbe et pronom (*dit-il, prends-en*) et dans certaines expressions figées (*tout à coup, de temps en temps*). Les autres contextes sont considérés comme variables ou facultatifs hormis quelques cas que nous pouvons considérer comme impossibles ou interdits (après un substantif au singulier, après la conjonction *et*, par exemple).

Les liaisons variables sont celles qui nous intéressent particulièrement ici, dans la mesure où elles constituent une variable sociolinguistique (Gadet, 1989), leur usage étant notamment déterminé par les caractéristiques du locuteur et de son rapport à son/ses interlocuteurs (variation diastatique) et par la situation en elle-même (variation diaphasique). Elles ont fait l'objet d'études de corpus nombreuses : dans le cadre du programme PFC (Durand & Lyche, 2008; Mallet, 2008), au sein du corpus d'Orléans (De Jong, 1988, 1994) puis du corpus ESLO (Baude & Dugua, 2011; Dugua & Baude, 2017), dans des enregistrements radiophoniques (Ågren, 1973), dans des corpus d'interactions parents/enfants (Chevrot et al., 2007; Liégeois, 2014), par des corpus plus contrôlés, notamment en acquisition (Chevrot et al., 2009; Nardy, 2008) et aussi dans des tâches de lecture réalisées par des enfants (Dugua & Baclesse, 2014). Nous observerons ici ce phénomène dans un contexte novateur : celui de la lecture partagée, par des lecteurs experts mais non professionnels.

### 1.3 L'appropriation des lectures adressées aux enfants

Concernant la façon dont les lecteurs s'approprient l'album, comme nous le soulignons précédemment, peu de travaux portent spécifiquement sur cette question. Nous pouvons toutefois nous appuyer sur la recherche menée par Pustka (2017) à partir d'un corpus de 32 livres lus (des enregistrements réalisés par des professionnels), corpus dans lequel l'auteur étudie l'usage des liaisons facultatives dans des contextes connus pour être fortement variables : après substantif, après des formes du verbe *être*, après certains adverbes et conjonctions. Après des substantifs au pluriel, 13% des liaisons sont réalisées dans ce corpus comparé au 3% chez Malécot (1975) (corpus de conversations naturelles par des locuteurs parisiens) ou 18% chez Ashby (1981) (corpus d'entretiens avec des locuteurs tourangeaux). Concernant les formes du verbe *être* au présent et à l'imparfait de l'indicatif, Pustka observe des taux relativement proches de ceux de Ågren (1973), dans un corpus radiophonique et globalement plus importants que dans le corpus d'entretiens de De Jong (1994) et dans le corpus PFC (Mallet, 2008). Précisons que Pustka affine ces résultats globaux en montrant que le lecteur fait varier son usage des liaisons selon le type de texte (Contes de fée vs Fables modernes ou Petit Nicolas) et selon le type de « discours » dans un même texte donné : récit vs discours direct par des adultes ou par des enfants. L'usage des liaisons dans des livres lus par des professionnels se distingue donc clairement des cas de conversations plus spontanées. Pustka souligne aussi que le genre « album » n'est pas homogène et qu'il est pertinent dans les analyses de distinguer les différents types de textes et de discours.

Notre volonté est de poursuivre le travail de Pustka (2017) en nous appuyant sur des lectures partagées, réalisées par des lecteurs non professionnels. Ce cadre nous permettra d'observer l'usage des liaisons dans un contexte original qui présente la particularité d'intégrer la dimension écrite par la lecture. Après avoir présenté notre méthodologie (corpus et méthode d'annotation), nous rendrons compte des premiers résultats et des pistes pour prolonger ce travail.

## 2 Méthodologie

### 2.1 Présentation du corpus

Le sous-corpus *Livres pour enfants* (LPE) est un module du corpus des *Enquêtes sociolinguistiques à Orléans* (ESLO). ESLO est un grand corpus oral à visée variationniste qui regroupe des données recueillies à deux périodes distinctes : dans les années 1970 et dans les années 2010 (Baude & Dugua, 2011). Le module en question a été recueilli en 2013 par des étudiants dans le cadre d'un cours de 3<sup>ème</sup> année de licence Sciences du Langage de l'Université d'Orléans sous la supervision d'Olivier Baude. Il s'agissait pour les groupes d'étudiants de se rendre dans des familles pour enregistrer un temps de lecture par un adulte à un ou plusieurs enfants. 28 enregistrements ont été réalisés, 30 histoires lues, d'une durée totale d'environ 2h50 (le plus court dure 1min40, le plus long 17min45), la durée moyenne est de 6 minutes environ. Ce corpus comprend 27.594 mots (dont 24.824 produits par les lecteurs, les autres provenant des enfants ou d'autres adultes assistant à la scène). Les informations disponibles sur les albums figurent dans le tableau 1, dans lequel nous faisons apparaître le nom simplifié du fichier (\_1251 signifie ESLO2\_LIVRENF\_1251), la durée de l'enregistrement en minutes, le nombre total de mots, le nombre de mots produits par le lecteur uniquement, le nombre de contextes de liaison dans la parole du lecteur uniquement, le titre de l'album et l'auteur et/ou la maison d'édition selon les informations disponibles (« nr » signifiant « non renseigné »).

Tableau 1. Description des enregistrements du corpus

Fichier	Durée (en min)	Nb mots Total	Nb mots Lecteurs	Nb contextes liaison Lecteurs	Titre	Auteur - Edition
1251	5	701	629	23	<i>Martin sur la lune</i>	Walt Disney
1302	2'30	296	222	10	<i>La petite bibliothèque des Barbapapas - La cuisine</i>	A. Tison et T. Taylor – Livre du dragon d'or
1303	3	376	323	15	<i>J'apprends à nouer mes lacets avec LES LAPINOS</i>	P. Couronne – Broché, Cert-Volant
1304	6'30	976	848	38	<i>Boucle d'or et les trois ours</i>	M. Morey – Pierre de soleil
1305	3	377	362	28	<i>Une fourmi pas comme les autres</i>	T. Corda (Italien), traduit par J. Duteil – Minedition
_1306	6'30	598	510	31	1) <i>Petit ours brun entend tout</i> 2) <i>Papoum</i>	1) M. Aubinats et D. Bour – Bayard Jeunesse 2) P. Teulade – Fleurus Presse
1307	5'45	1065	798	41	<i>Le Mange-Doudous</i>	J. Béziat – Pastel, l'école des loisirs (2013)
1308	9'30	1403	1402	54	<i>Clara et Béatrice</i>	C. Pistinier – Kaléidoscope
1309	10	1697	1592	121	<i>L'imagerie du Père Noël</i>	E. Beaumont – Fleurus Presse
1310	3'30	578	578	29	<i>Le petit chaperon rouge</i>	Ch. Perrault – <i>nr</i>
1311	17'45	2993	2750	155	<i>Winnie l'ourson et l'arbre à miel</i>	Disney – Hachette
_1312	8	1181	926	42	1) <i>Petit Lapin Blanc à la maternelle</i> 2) <i>Le vilain petit canard</i>	1) M.-France Floury et F. Boisnard – G. Langueureau 2) <i>nr</i>
1313	6	747	717	25	<i>La Fée Baguette et son Loudoudou</i>	F. Joly et M. Barçilon – Lito
1314	6'45	773	754	28	<i>La Fée Baguette et son Loudoudou</i>	F. Joly et M. Barçilon – Lito
1315	6	998	821	39	<i>La culotte du loup</i>	S. Servant et L. Le Saux – Didier jeunesse
1316	4'40	693	637	22	<i>Le plus féroces des loups</i>	S. Poillevé – Père castor Flammarion
1317	5'30	930	850	40	<i>Chasse aux trésors</i>	S. Ichikawa – L'école des loisirs
1318	3	480	469	17	<i>Mon histoire à colorier Cars</i>	Disney – Hachette Jeunesse
1319	8'15	1609	1564	61	<i>Le Petit Chaperon rouge</i>	<i>nr</i> – Père castor Flammarion
1320	5'10	854	806	42	<i>Belle la Coccinelle</i>	A. Krings – Gallimard Jeunesse
1321	7'15	1337	1264	46	<i>Donald détective privé</i>	Walt Disney – Hachette
1322	1'40	236	236	20	<i>Les histoires de Petit ours brun</i>	D. Bour – Bayard Jeunesse
1323	7	1151	1049	57	<i>Totoche et le poisson malheureux</i>	C. Valckx – Ecole des loisirs
1324	3	318	147	3	<i>Demain, je te mangerai</i>	C. Ramadier et V. Bourgeau – Ecole des loisirs
1325	7	1104	1098	75	<i>Arc-en-Ciel et le petit poisson perdu</i>	M. Pfister – Nord-Sud Eds
1326	9'30	1678	1515	115	<i>La fée sorcière</i>	B. Minne et C. Creut – Pastel
1327	9'20	1422	1184	30	<i>Cornébidouille</i>	M. Bonniol et P. Berthand – L'école des loisirs
1328	7'20	1023	773	32	<i>Melba l'Oursome</i>	G. Noël – Les belles histoires de Pomme d'Api

\* Corresponding author : celime.dugua@univ-orleans.fr

Vingt-six lecteurs différents (19 femmes, 7 hommes) ont lu une histoire à des enfants âgés entre 2;6 et 6 ans. Dans la majorité des cas, les lecteurs sont des membres de la famille de l'enfant (on trouve deux animatrices, et une amie de la famille). Les deux animatrices ont chacune lu une histoire à un groupe d'enfants (fichiers ESLO2\_LIVRENF\_1327 et ESLO2\_LIVRENF\_1328). Dans tous les autres cas, il y avait un ou deux enfants seulement qui assistai(en)t à la lecture.

L'ensemble des fichiers a été transcrit à l'aide du logiciel *Transcriber* en suivant la méthode initiée dans le projet ESLO (Dugua et al., 2019). Chaque fichier a été transcrit une première fois par les étudiants, puis relu à deux reprises par deux relecteurs différents pour aboutir à une version validée. Les versions validées ainsi que les enregistrements seront mis à disposition dans le courant de l'année 2022.

Dans cette étude exploratoire, les différents moments qui jalonnent les lectures partagées n'ont pas été annotés (lecture, commentaire, question/réponse). En d'autres termes, nous traiterons ici les contextes de liaison sans distinction entre les moments de lectures et ceux d'échanges, et à l'intérieur des moments de lecture, entre les dialogues et le récit. Nous reviendrons sur cette limite dans la discussion.

## 2.2 Méthode de repérage et de pré-annotation des liaisons

Les enregistrements ayant été transcrits avec le logiciel *Transcriber* (Barras et al., 2001), ce format offre la possibilité d'utiliser le logiciel *Jtrans* (Cerisara et al., 2009) d'alignement automatique mot/signal sonore. Ce choix a été fait dans le but de segmenter le mot1 sans avoir à le repérer sur le signal. Tous les duos enregistrements/transcriptions sont traités automatiquement de la même manière grâce à un script développé en Python et qui utilise des logiciels auxiliaires en ligne de commande, dont le logiciel *Jtrans*.

Une fois la transcription alignée au mot sur le signal sonore, elle se trouve au format du logiciel *Praat* (Boersma et Weenink, 1992-2021). L'avantage de ce logiciel est l'annotation multicouche. Le signal sonore est alors annoté au mot via *Jtrans* et en unités de parole issues de *Transcriber*. Si jamais des métadonnées existent au niveau du locuteur ou de l'enregistrement, il est possible de les ajouter lors de cette étape. Les annotations sont ensuite converties au format du logiciel *Elan* (Wittenburg et al., 2006) qui permet de garder l'annotation multicouche de *Praat* mais également de pouvoir travailler sur corpus en incluant certaines spécificités propres au logiciel : notion de type et de vocabulaire contrôlé.

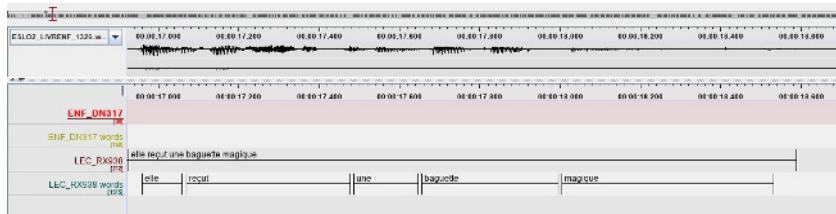


Fig. 1. Visualisation dans *Elan* de l'annotation au mot par *Jtrans*

L'enrichissement de l'annotation passe par l'étiquetage morphosyntaxique des mots et l'ajout de la ligne d'annotation "liaison" portée par le mot1. Pour les conversions et l'étiquetage, le script fait appel à l'outil *Teicorpo* (Parisse et al., 2020), l'étiquetage intégré étant *Treetagger* (Schmid, 1994). Les liaisons sont, elles, repérées automatiquement sur la ligne segmentée au mot avec le dernier caractère du mot1 et le premier caractère du mot2,

\* Corresponding author : [celine.dugua@univ-orleans.fr](mailto:celine.dugua@univ-orleans.fr)



via une liste préétablie et pouvant être mise à jour en fonction du corpus et intégrant la casse et l'accentuation des voyelles :

**Tableau 2.** Listes des premiers et derniers caractères des mots1 et mots2

Dernier caractère	t, s, p, x, r, n, z, d, w, m, -
Premier caractère	a, e, i, o, u, y, h, &, -

A partir de repérages dans différents modules d'ESLO, certains mots sont exclus en mot1 ou/et en mot2 en raison de leurs propriétés phonologique ou conversationnelle :

**Tableau 3.** Liste des mots ou expressions exclus des contextes de liaison

Exclus en tant que mot1	Exclus en tant que mot2	Exclus en tant que mot1 et mot2
*ar basket foot sept	euh	hein
*ir ben huit set	oui	hm
*or cet index sud		hum
*ur cher internet tennis		ouais
direct jadis test		
donc post-it zut		
et prout		

### 2.3 Annotation des liaisons

La ligne d'annotation de la liaison est ensuite corrigée manuellement dans *Elan* à l'écoute du signal sonore. La liste des liaisons possibles est intégrée en tant que vocabulaire contrôlé dans le logiciel et chaque ligne d'annotation comporte un type (spécificité du logiciel) différent de façon à pouvoir effectuer des recherches sur les lignes semblables par exemple : lemme, mots, unité de parole, liaison, ...

La liste des types de réalisation des liaisons possibles est la suivante :

- N, Z, T, R, P : pour les liaisons réalisées en /n/, /z/, /t/, /ʁ/, /p/. Ex : *très heureux* produit avec une liaison en /z/, *trop énérvé* produit avec une liaison en /p/ ;
- O : pour l'absence de liaison. Ex : *marcher ensemble* produit sans le /ʁ/ ;
- NSP : « ne sait pas », lorsque l'annotateur ne parvient pas à décider à l'écoute si la liaison est réalisée et/ou comment ;
- FRONT : pour les cas où, en raison d'une frontière syntaxique entre deux constituants, la liaison paraît impossible. Ex : *les gars ils partent* entre *gars* et *ils* ;
- PAUSE : s'il y a une pause entre les deux mots et que la liaison n'a pas été réalisée ;
- LN, LZ, LT, LR, LP, NZ, NR, NP, ZN, ZT, ZR, ZP : tout un ensemble de codages permettant de rendre compte de « remplacements » : ZT signifie que la liaison produite est en /z/ et qu'on attendait une liaison en /t/ comme par exemple dans *il prend z-un thé* où le /z/ remplace le /t/ attendu) ;
- Autre : pour tout autre cas où le contexte de liaison paraît impossible. Ex : devant une onomatopée ou une interjection.

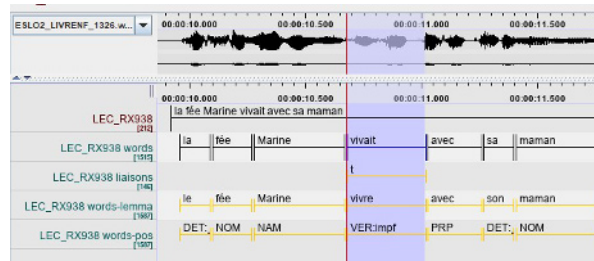


Fig. 2. Visualisation sous *Elan* de l'annotation des liaisons

### 3 Résultats

Après la présentation de résultats globaux, nous nous attacherons à rendre compte des données sur des lexèmes particuliers, ce type d'observation est pertinent pour la liaison dans une approche valorisant l'usage. Le travail précurseur de De Jong (1994) sur le corpus d'Orléans mettait en effet en évidence l'importance de s'appuyer sur des formes lexicales et non sur des catégories ou des lemmes, puisque par exemple, il observait une variabilité importante entre les formes du verbe « être » (69% après *est*, 46% après *sont*, 29.2% après *suis* par exemple), ou entre les prépositions (93.7% après *en*, 88.1% après *chez*, 16.7% après *pendant*). Cette variabilité devient invisible en observant uniquement les niveaux catégoriels tels que « après verbe », « après préposition », etc.

Afin de mieux cerner l'usage des liaisons en lecture partagée, nous comparerons nos résultats à des lectures faites par des professionnels (Pustka, 2017) et à des entretiens (Baude et Dugua, 2015) lorsque les données le permettront. Nous faisons l'hypothèse que des lectures faites par des membres de la famille se situent entre ces deux pôles. Nous nous attendons aussi, comme cela est le cas dans les entretiens (Baude et Dugua, 2015), à une variabilité interindividuelle forte. Plusieurs facteurs pourront être évoqués à titre d'hypothèses, parmi lesquelles : le lecteur en lui-même, son lien avec l'enfant, le type d'album lu.

#### 3.1 Résultats globaux

Les résultats présentés s'appuient uniquement sur les données des lecteurs (nous excluons les données des enfants et des interventions ponctuelles d'autres adultes).

Ce corpus regroupe 1.239 contextes de liaison, qui se répartissent en 442 contextes de liaison obligatoire (LO) et 797 contextes de liaison facultative (LF). Les LO sont réalisées à 100% par les lecteurs et les LF à 34.25% (14.21) en moyenne (entre 0% à 70% selon les différents lecteurs). Ce taux de LF réalisées est 13 points au-dessus de celui obtenu dans un corpus d'entretiens d'ESLO avec la même méthode (Dugua & Baude, 2017). La modalité de lecture semble donc générer plus de LF réalisées que la parole en entretien.

On retrouve par ailleurs le même ordre de fréquence des consonnes de liaison que Dugua et Baude (2017) : en LO : /z/ (249) > /n/ (129) > /t/ (64) et en LF : /t/ (147) > /z/ (115) > /n/ (11). Ce résultat découle des types de mots que l'on trouve en LO et LF. On constate que la modalité de lecture ou de parole spontanée n'affecte pas les types de mots permettant une liaison.

Ces données globales confortent les précédentes études sur la liaison : la systématisme des réalisations en LO, un taux de LF supérieur en lecture par rapport à une situation

d'entretien plus ou moins spontanée et une variabilité importante entre les locuteurs-lecteurs.

### 3.2 Liaisons et formes du verbe être

Regardons de façon plus fine les données sur des lexèmes particuliers.

Notre faible effectif de données nous oblige à limiter les lexèmes observables. Nous ne présenterons les résultats que si le nombre d'occurrences est égal ou supérieur à 15. Le graphique ci-dessous (Figure 3) rend compte des taux de réalisation des liaisons après deux formes du verbe *être* déjà observées dans d'autres corpus : le corpus de Pustka correspondant à des livres lus par des lecteurs professionnels, et les entretiens réalisés dans les années 2010 du module Diachronie d'ESLO (Baude & Dugua, 2015).

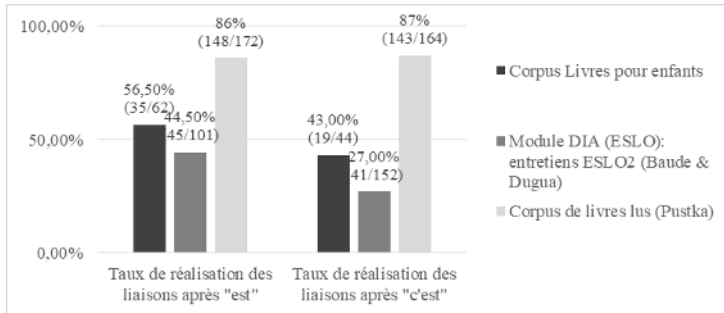


Fig. 3. Taux de réalisation des liaisons après "est" et "c'est" dans trois corpus différents

Le test Chi2 révèle des différences significatives avec le corpus Pustka après *est* (Chi2=21.714, p<0.001) et après *c'est* (Chi2=36.504, p>0.001), et non significatives avec les entretiens du module Diachronie d'ESLO (après *est* : Chi2=1.726, p=0.1889 ; après *c'est* : Chi2=3.492, p=0.062).

Sur ces deux formes du verbe *être*, la réalisation des LF en lecture partagée se situe davantage du côté des échanges de type « entretien » que des lectures faites par des professionnels. Ce résultat permet de pointer la spécificité des livres lus (Pustka, 2017) où le taux de LF est très important. Dans les lectures partagées, l'usage des liaisons se rapproche davantage de ce que les locuteurs pratiquent dans des situations d'entretiens.

### 3.3 Liaisons après les adverbes

Les adverbes observés sont *pas* (43 occurrences) et *très* (18 occurrences). On relève 27.9% de liaisons réalisées après *pas* (37% chez Pustka) et 100% après *très* (100% chez Pustka). Les différences sont moins marquées ici que sur les formes du verbe *être*.

Dans les données orales, la liaison après l'adverbe *très* peut être considérée comme systématique. C'est d'ailleurs ce qu'observait De Jong en 1994 avec 158 réalisations sur 159 contextes (99.4%).

L'adverbe *pas* est intéressant à plusieurs titres. Il constitue l'une des particules de la négation en français, celle qui subsiste seule dans la plupart des cas en français oral (dans l'usage oral le *ne* est souvent effacé (Riegel et al., 1994)). Or ici, la base étant écrite, on trouve davantage de formes complètes de négation (n=28) que de formes incomplètes (n=9). Si on regarde les taux de réalisation dans chacun des cas, on ne relève aucune

réalisation dans les formes incomplètes et 10/28 soit 35.7% dans les formes complètes. Il semble donc que la présence ou l'absence du *ne* joue un rôle sur la réalisation de la liaison après le *pas*, mais cela devra être confirmé par d'autres études avec davantage de données.

### 3.4 Liaisons et temps verbaux

Enfin, nous portons notre regard sur l'usage des liaisons après des formes verbales, et cela à travers l'entrée des temps verbaux (Figure 4). Nous nous appuyons sur l'annotation automatique en lemmes et parties du discours de *Treetagger* pour cette étude. Elle nous permet de faire des recherches rapides sur les événements linguistiques étudiés. Ainsi, les temps verbaux les plus fréquents dans notre corpus sont les suivants :

- L'infinifitif : 49 occurrences
- L'imparfait : 46 occurrences
- Le présent : 276 occurrences
- Le passé simple : 27 occurrences

La figure 4 rend compte du taux de réalisation des liaisons facultatives après chacun de ces temps verbaux.

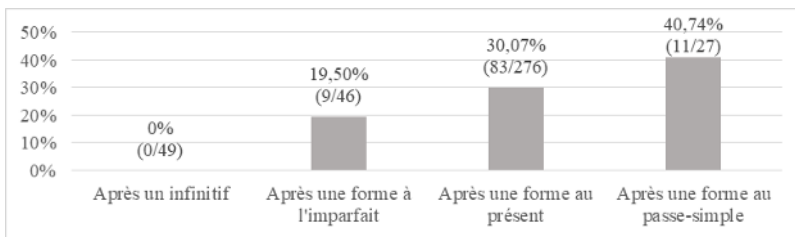


Fig. 4. Taux de réalisation des liaisons après quatre temps verbaux

Il s'avère qu'aucune liaison n'a été réalisée après un verbe à l'infinifitif. Il y a 34 lemmes différents dont les nombre d'occurrences se situent entre 1 et 4. Ce contexte génère des liaisons avec la consonne /ʒ/ qui représente généralement très peu d'occurrence. Dans le corpus de Dugua et Baude (2017), sur 23 formes à l'infinifitif en -er, aucune liaison réalisée n'avait été relevée.

A l'imparfait, la répartition est toute autre, autour de 16 lemmes différents, avec deux verbes *être* et *avoir* qui sont largement majoritaires : 13 occurrences chacun, soit 56% des lemmes. En moyenne, 19.5% des liaisons sont réalisées après une forme à l'imparfait (9/46). Parmi les 9 réalisations, 7 sont après une forme du verbe *être*. On ne relève aucune liaison réalisée après des formes du verbe *avoir* (De Jong (1988) trouvait des résultats similaires dans des entretiens), ni après les 13 autres lemmes. En bref, les liaisons réalisées après une forme à l'imparfait sont essentiellement après le verbe *être* et plus particulièrement la forme *était* (7/7).

Les deux autres temps verbaux sont ceux après lesquels les liaisons sont les plus souvent réalisées : après une forme verbale au présent 30.07% (83/276) et après une forme verbale au passé simple 40.74% (11/27). Le premier est le temps le plus fréquemment observé dans notre corpus, et aussi dans la parole spontanée (50 lemmes différents). Le taux de réalisation global est de 30.07% (83/276). Le taux le plus élevé est porté par le verbe *être* (52% de liaisons réalisées) : 88% sont des liaisons en /t/, 12% des liaisons en /z/. Le seul autre verbe qui dépasse les 15 occurrences (20 occurrences) est le verbe *avoir* qui a un taux de 0%.

Quant au passé simple, le temps du récit, il est fortement associé à l'écrit littéraire et par là, à une certaine forme de norme de l'écrit. Parmi les 19 lemmes, *mettre* prédomine (6 occurrences en tout) et engendre un taux important de réalisation de liaison puisqu'elle est systématiquement réalisée après les deux formes repérées : *se mit à* ou *se mirent à*. Les autres formes correspondent à une ou deux occurrences et sont la plupart du temps non réalisées (16 non réalisations contre 5 réalisations). Ces résultats méritent d'être affinés. Il s'avère que le taux global de 40.74% est essentiellement porté par une même lectrice (LEC\_RX938), une maman aide-soignante qui lit « Le fée sorcière » (5 réalisations sur les 11 du corpus pour cette lectrice).

### 3.5 Des lecteurs et des albums

Outre les paramètres internes à prendre en compte, on sait que la liaison facultative est une variable sociolinguistique qui à l'oral varie selon les locuteurs et les situations. Qu'en est-il en lecture ? L'approche que nous défendons ici consiste à partir d'un premier niveau d'analyse de résultats globaux fondés sur des grandes catégories, pour nous appuyer, dans un deuxième temps sur des observations fines qui prennent en compte les caractéristiques détaillées de chaque lecteur associées à celles de chaque situation.

Globalement, sur nos 26 lecteurs, le taux moyens de LF réalisées est de 34.25%, avec un écart type de 14.21 et une variabilité comprise entre 0% (0/12) et 70% (21/30). Les données dont nous disposons sur les lecteurs sont leur âge, leur sexe, leur lieu de naissance, leur date d'arrivée à Orléans, leur profession, leur niveau d'étude, et leur relation avec l'enfant. Nous pouvons constater que ces critères, pris individuellement, ne permettent pas d'expliquer les variations dans les résultats. On voit par exemple que, sans surprise, la lectrice qui réalise le plus de liaisons facultatives est avocate (70%) et le lecteur qui en fait le moins est ouvrier (0%), ce qui est compatible avec une analyse de co-variations des pratiques linguistiques et des catégories sociales. Mais on constate des variations internes aux catégories, ainsi une autre avocate a un taux bien plus faible (36%) que la première, tout comme on trouve un écart conséquent entre une professeure d'anglais à 55% et une professeure des écoles (24%) ou un professeur d'université (25%). Comme Mullineaux et Blanc (1982) l'avaient mis en évidence dans le corpus d'Orléans, ces simples critères de profession (ou plus largement de catégorie professionnelle), d'âge et de genre, ne sont pas suffisants pour rendre compte des variabilités. Les pratiques culturelles et ici en particulier, le rapport à la norme et à sa transmission sont à appréhender dans la dynamique des trajectoires de vie, et la dynamique de chaque interaction linguistique qu'il convient de pouvoir décrire comme éléments déterminants et enjeux des productions de chaque lecteur.

Quant au type d'album lu, contrairement au corpus de Pustka (2017) qui regroupait à la fois des contes, des fables, des œuvres littéraires longues et qui donc laissait apparaître des différences selon le genre, ici, les albums choisis n'ayant pas été contrôlés, nous ne retrouvons pas ces différentes catégories. Notre corpus permet toutefois à son échelle de montrer que le genre d'album à lui seul ne peut être la seule explication des variations. En effet, nous disposons d'une histoire *La fée baguette et son Loudoudou* qui a été lue par deux lecteurs différents, et pour laquelle nous relevons des taux de réalisations très différents (0% pour l'une, 38.5% pour l'autre sur 12 et 13 contextes).

Enfin, une tendance apparaît quant au critère de l'âge de l'enfant à qui l'histoire est lue. Les 5 lectures pour lesquelles les taux de LF sont les plus faibles (entre 0 et 25%) sont toutes faites à des enfants de 3 ans et les 5 lectures où les LF sont les plus réalisées (entre 47 et 70%) sont à destination d'enfants de 5 ou 6 ans. Cette tendance sera évidemment à

vérifier, mais le critère de l'âge de l'enfant et donc l'adaptabilité que le lecteur opère à l'idée qu'il se fait du langage de l'enfant semble jouer sur les lectures.

## Conclusion

En guise de conclusion, nous revenons sur les principaux éléments de l'article, à partir desquels nous évoquerons les limites de cette recherche exploratoire et les pistes que nous envisageons d'explorer désormais.

Observer l'usage des liaisons variables dans un corpus de lecture partagée permet d'aborder le lien entre l'usage oral et l'écrit, de pointer la nécessité de décrire les caractéristiques des albums jeunesse, de questionner la transmission entre parent et enfant. Cette étude exploratoire ne permet pas de cerner toutes ces questions mais elle ouvre la voie à de futures études, en mettant en place des protocoles plus contraignants et en augmentant le nombre de lectures.

Nous retiendrons ici que l'usage des liaisons d'un lecteur à l'autre est très variable et que l'âge des enfants à qui s'adresse la lecture semble être un paramètre pertinent mais pas unique. Il semblerait que les lecteurs réalisent davantage les liaisons avec des enfants plus grands (5-6 ans) – âge auquel ils entrent dans une relation scolaire avec l'écriture et la norme – qu'avec les plus jeunes (3 ans). Grâce aux métadonnées fournies par le corpus, nous disposons de certaines caractéristiques socio-démographiques des lecteurs ; elles sont probablement un facteur pertinent de variations structurées, mais le trop faible effectif ici ne permet pas de le vérifier de façon systématique. Par ailleurs, pour ce type de module sur la lecture, il serait intéressant de disposer d'informations sur la pratique de la lecture à haute voix. Entre les professionnels observés par Pustka (2017), les enseignants, les parents qui ont une pratique courante et d'autres lecteurs qui lisent pour la première fois, on peut sans risque penser que l'expérience aura une influence sur la fluence et donc sur l'usage des liaisons. A ce paramètre, s'ajoutent ceux de la trajectoire sociale, des pratiques culturelles, du rapport aux normes qui seraient nécessaires pour affiner les résultats généraux et comprendre de façon fine les variations. C'est l'approche adoptée par Dugua et Baude (2017) qui montrent sur des données orales l'intérêt de prendre en compte, un premier niveau d'analyse à partir de grandes catégories, puis des caractéristiques fines des locuteurs et de la situation afin d'atteindre un second niveau d'analyse des variations et *in fine* de rendre compte de l'évolution de leur usage de la liaison et de la dynamique du changement linguistique.

La question de l'influence du type d'album, qui a pu être mise en évidence par Pustka (2017) avec des lecteurs professionnels n'a pas pu être observée ici, car c'est un élément qui nécessiterait d'être contrôlé en amont. Nous pensons par exemple à une sélection d'albums selon des critères liés au style, à la période de parution (conte « classique » vs ouvrages récents), au statut des illustrations, etc. Ce contrôle est indispensable pour traiter l'album en tant que variable. Et par ailleurs, nous envisageons dans un travail à venir de sélectionner un album de référence qui serait lu par plusieurs lecteurs afin de neutraliser l'effet de l'album pour se centrer sur les paramètres des lecteurs et des enfants.

En lien avec cet élément, et comme nous l'avons déjà évoqué, nous devons distinguer par des annotations les différents temps de la lecture partagée : la lecture à proprement parler qui s'appuie sur un écrit et les échanges spontanés entre l'enfant et l'adulte ; et dans les albums, les temps de dialogue et les temps de récit. Ces annotations permettront d'observer si les usages des lecteurs sont influencés par la présence d'une forme écrite d'une part, et la modalité de lecture, d'autre part, par rapport à un oral spontané.

En ce qui concerne l'observation de l'usage des liaisons, nous défendons ici une approche qui s'appuie non pas sur des catégories grammaticales générales, mais sur des formes lexicales ou des constructions, approche qui nous semble beaucoup plus intéressante

pour rendre compte de la dynamique de la liaison. On sait ce phénomène très sensible à la fréquence d'usage, l'aborder par une voie lexicale est donc la plus pertinente. C'est ce que nous avons pu montrer avec deux formes du verbe être (*est* et *c'est*) et avec deux adjectifs particuliers (*pas* et *très*), chacun présentant des particularités syntaxiques et pragmatiques qui affectent l'usage des liaisons.

Prenant en compte la lecture partagée comme une forme de CDS, nous voyons que l'enfant sera confronté à des types de réalisation des liaisons différents dans son input. Cette transmission de l'usage des liaisons par le biais de la lecture peut être une façon d'entrer dans les spécificités de l'écrit, et dans la littératie. L'usage du passé simple dans les histoires est par exemple facilement repris par les enfants lorsqu'ils narrent à leur tour, et un usage particulier des liaisons variables en lecture partagée peut être un appui pour les enfants dans leur acquisition du phénomène.

## Bibliographie

- Agence Quand les livres relie (2018). *Les Tout-Petits, le Monde et les Albums*, Eres, coll. « 1001 BB », n° 154.
- Ågren, J. (1973). *Etude sur quelques liaisons facultatives dans le français de conversation radiophonique : Fréquences et facteurs*. Acta Universitatis Upsaliensis.
- Alamiche, D. (2011). *La bibliothécaire jeunesse : une intervenante culturelle : 60 animations pour les enfants de 18 mois à 11 ans*, Editions du Cercle de La Librairie.
- Ambridge, B., & Lieven, E. (2011). *Child Language Acquisition* (Cambridge University Press).
- Ashby, W. (1981). French liaison as a sociolinguistic phenomenon. In W. W. Cressy & D. J. Napoli (Éds.), *Linguistics Symposium on Romance Languages (9th)* (p. 46-57). Georgetown University Press.
- Barras, C., Geo, E., & Wu, Z. (2001). Transcriber: Development and Use of a Tool for Assisting Speech Corpora Production. *Speech Communication*, 33, 5–22.
- Bastide, I., & Joigneaux, C. (2014). Littératie précoce, albums et lectures partagées à l'école maternelle. *Spirale - Revue de recherches en éducation*, 53(1), 9-19.
- Baude, O., & Dugua, C. (2011). (Re)faire le corpus d'Orléans quarante ans après : Quoi de neuf linguistes? *Corpus*, 10, 99-118.
- Baude, O., & Dugua, C. (2015). Usage de la liaison dans le corpus des ESLOs : Vers de nouveaux (z) ouvrages de référence? In G. Dostie & P. Herdermann (Éds.), *La diavariation en français actuel* (p. 349-371). Peter Lang.
- Boersma, P. & Weenink, D. (1992–2021). Praat: doing phonetics by computer [Computer program]. Version 6.2, retrieved 15 November 2021 from <https://www.praat.org>
- Boiron, V. (2014). Raconter et lire des récits de fiction : effets comparés sur la compréhension d'élèves de maternelle, *Repères. Recherches en didactique du français langue maternelle*, 50 : 83-104.
- Boiron, V. (2015). Pratiques de lecture d'albums de littérature de jeunesse dans deux classes de petite et moyenne section de maternelle, *Spirale - Revue de recherches en éducation*, 55(1) : 31-42.
- Bonnéry, S. (2014). Les livres et les manières de lire à l'école et dans les familles : *Le français aujourd'hui*, 185(2), 47-57.
- Bonnéry, S., & Joigneaux, C. (2015). Des littératies familiales inégalement rentables scolairement. *Le français aujourd'hui*, 190(3), 23-34.
- Booij, G., & De Jong, D. (1987). The domain of liaison : Theories and data. *Linguistics*, 25, 1005-1025.
- Cameron-Faulkner, T., & Hickey, T. (2011). Form and function in Irish child directed speech. *Cognitive Linguistics*, 22(3), 569-594.

- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27, 843-873.
- Cameron-Faulkner, T., & Noble, C. (2013). A comparison of book text and Child Directed Speech. *First Language*, 33(3), 268-279.
- Catach, N. (1973). Notions actuelles d'histoire de l'orthographe. *Langue française*, 20, 11-18.
- Cèbe, S. & Sève, P. (dirs.) (2014). Lire en maternelle : la lecture avant que de savoir lire, *Repères. Recherches en didactique du français langue maternelle*, 50.
- Cerisara, C., Mella, O., & Fohr, D. (2009, September). JTrans, an open-source software for semi-automatic text-to-speech alignment. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association-Interspeech 2009*.
- Chenu, F., & Jisa, H. (2006). Caused motion constructions and semantic generality in early acquisition of French. In E. Clark & B. Kelly (Éds.), *Constructions in Acquisition*. CSLI.
- 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*, 17, 103-128.
- Chevrot, J.-P., Dugua, C., & Fayol, M. (2009). Liaison, word segmentation and construction in French : A usage-based account. *Journal of Child Language*, 36(3), 557-596.
- Dawson, N., Hsiao, Y., Tan, A. W. M., Banerji, N., & Nation, K. (2021). Features of lexical richness in children's books : Comparisons with child-directed speech. *Language Development Research*.
- De Jong, D. (1988). *Sociolinguistic aspects of French liaison*. Vrije Universiteit Amsterdam Editor.
- De Jong, D. (1994). La sociophonologie de la liaison orléanaise. In C. Lyche (Éd.), *French Generative Phonology : Retrospective and Perspectives* (p. 95-129). ESRI.
- Dugua, C., & Baclesse, M. (2014). *Incidence d'effets de fréquence sur l'usage de la liaison en lecture à haute voix et dans des jugements normatifs chez des enfants de CE2-CM1*. In C. Soum-Favaro, A. Coquillon & J.-P. Chevrot (Éds.). *La liaison: approches contemporaines*. Bern: Peter Lang: 117-139.
- Dugua, C., & Baude, O. (2017). La liaison à Orléans, corpus et changement linguistique : Une première étude exploratoire. *Journal of French Language Studies*, 27(1), 41-54.
- Dugua, C., Ganaye, J., Badin, Fl. & Baude, O. (2019). *Des liaisons et des corpus : apports d'une étude sur le changement linguistique en temps réel*, 10èmes journées de linguistique de corpus, Grenoble, 26-28 novembre 2019.
- Durand, J., & Lyche, C. (2008). French liaison in the light of corpus data. *Journal of French Language Studies*, 18(1), 33-66.
- Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 81, 177-206.
- Frier, C. (Éd.). (2006). *Passeurs de lecture : Lire ensemble à la maison et à l'école*. Retz.
- Gadet, F. (1989). *Le français ordinaire*. Armand Colin.
- Harris, J. (1990). *Early Language Development : Implications for Clinical and Educational Practice*. Routledge.
- Hoff, E. (2002). Causes and consequences of SES-related differences in parent-to-child speech. In M. H. Bornstein & R. H. Bradley (Éds.), *Socioeconomic status, parenting and child development* (p. 147-160). NJ: Lawrence Erlbaum Associates.
- Kalb, G. & van Ours, J. (2014). Reading to young children: A head-start in life? *Economics of Education Review*, 40: 1-24.
- Laks, B. (2005). La liaison et l'illusion. *Langages*, 158, 101-125.
- Langlard, H. (1928). *La liaison dans le français*. Librairie ancienne Edouard Champion.



- 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.
- Mélançon, J. (2015). Le potentiel des albums jeunesse pour soutenir la compréhension du monde de la pensée chez l'enfant d'âge préscolaire, *Revue de recherches en littérature médiatique multimodale*, vol. 2.
- Malécot, A. (1975). French liaison as a function of grammatical, phonetic and paralinguistic variables. *Phonetica*, 32, 161-179.
- Mallet, G. (2008). *La liaison en français : Descriptions et analyses dans le corpus PFC*. Thèse de doctorat. Université Paris Ouest Nanterre la Défense Éditeur.
- Mol, S., Bus, A., de Jong, M. & Smeets, D. (2008). Added Value of Dialogic Parent-Child Book Readings: A Meta-Analysis. *Early education and development*, 19: 7-26.
- Mullineaux, A. & Blanc, M. (1982). The problems of classifying the population sample in the socio-linguistic survey of Orléans (1969) in terms of socio-economic, social and educational categories. *Review of Applied Linguistics*, 55, 3-37.
- Nardy, A. (2008). *Acquisition des variables sociolinguistiques entre 2 et 6 ans : Facteurs sociologiques et influences des interactions au sein du réseau social*. Thèse de doctorat. Université Stendhal, Grenoble.
- Parisse, C., Etienne, C., & Liégeois, L. (2020). TEICORPO: a conversion tool for spoken language transcription with a pivot file in TEI. *Journal of the Text Encoding Initiative*.
- Pustka, E. (2017). L'écrit avant l'écriture : La liaison dans les livres audio pour enfants. *Journal of French Language Studies*, 27(2), 187-214.
- Reese, E., Cox, A., Harte, D. & McAnally, H. (2003). Diversity in Adults' Styles of Reading Books to Children. In van Kleeck, A., Stahl, S. & Bauer, E. (Eds). *On Reading Books to Children - Parents and Teachers*, 37-57.
- Riegel, M., Pellat, J.-C., & Rioul, R. (1994). *Grammaire méthodique du français* (1ère édition). PUF.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester, UK.
- Sénéchal, M., Pagan, S., Lever, R. & Ouellette, G. (2008). Relations Among the Frequency of Shared Reading and 4-Year-Old Children's Vocabulary, Morphological and Syntax Comprehension, and Narrative Skills. *Early Education and Development*, 19(1): 27-44.
- Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 43, 549-565.
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4(1), 1-22.
- Snow, C. E. (1995). Issues in the study of input : Finetuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher & B. MacWhinney (Éds.), *The Handbook of Child Language* (p. 180-193). Blackwell.
- Snow, C. E. (1999). Beginning from baby talk : Twenty years of research on input in interaction. In C. Gallaway & B. J. Richards (Éds.), *Input and interaction in language acquisition* (p. 1-12). Cambridge University Press.
- Tomasello, M. & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 57, (6), 1454-1463.
- Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In *Proceedings of the Fifth International conference on Language Resources and Evaluation* (pp. 1556-1559).