

Liaison, élision et enchaînement : le rôle de la phonologie et du lexique chez les enfants au début de l'école primaire

Ruvoletto, Samantha

Université Paris 8, UMR 7023 – Structures Formelles du langage
samantha.ruvoletto@gmail.com

1 Introduction

Dans l'acquisition du français, l'enfant reçoit comme input une langue orale où les frontières entre les mots sont masquées par des phénomènes consonantiques comme la liaison (les [le] + ours [uʁs] => [lezuʁs]), l'élision (le [lə] + avion [a.vjɔ̃] => [la.vjɔ̃]) et l'enchaînement (la [la] + petite [pə.tit] + orange [ɔʁɑ̃ʒ] => [lapetitɔʁɑ̃ʒ]). Entre 2 et 3 ans l'apparition d'erreurs qui reflètent une difficulté de segmentation a été décrite et analysée par plusieurs auteurs. Elles apparaissent surtout dans le contexte spécifique déterminant-nom en cas de liaison.

Les erreurs de liaison montrent que les enfants ont des problèmes à déterminer les frontières gauches de mots. Il a été proposé que le français est une langue « cursus » qui s'organise en structures à syllabe ouverte du type CV (Pulgram, 1970 ; Bechade, 1992). Les structures syllabiques masquent dans l'input adulte les frontières lexicales, en cas d'élision (p.ex. *l'ami* [la.mi]) ou de liaison (p.ex. *les amis* [le.za.mi]), et les enfants doivent pour chaque entrée lexicale trouver quelle est la forme de référence alors que la frontière déterminant-nom qui est ambiguë est difficile à établir. On constate en effet des cas de mauvaises segmentations en contexte de liaison chez les enfants entre 2 et 3 ans. Plusieurs études (Wauquier, 2013 ; Chevrot et al., 2013) ont montré des productions caractérisées par l'intégration de la consonne de liaison au début du nom (p.ex. *de(s)[n]éléphants* [de.ne.le.fã] pour *des[z]éléphants* [de.ze.le.fã] ou [papanuʁs] dans *papa ours*) ou, dans une moindre mesure, par la complète omission de la consonne de liaison (*u(n)[Ø]avion* [ɛ̃ a.vjɔ̃] sans réalisation de la CL [n] pour *un[n]avion* [ɛ.na.vjɔ̃]). Entre 4 et 5 ans ces erreurs de liaison diminuent, avant de disparaître à l'âge de 6 ans (Basset, 2000). Après 5 ans, seules les erreurs d'omission de la CL subsistent dans la production orale (Chevrot et al., 2013).

Ces problèmes de segmentation se manifestent aussi sur d'autres types de productions erronées chez les enfants. Dans une étude de 2006, Dugua remarque dans la production des séquences déterminant-nom des types « un latrompe » ou « un lalampe » qui montrent que l'article n'est pas détaché du nom et forme avec lui une seule unité. Dans ces travaux, Bassano (2010) remarque que les articles sont non seulement très présents dans l'input des enfants, mais qu'ils sont aussi produits plus tôt, à 2 ans et demi, si l'on compare avec les productions des enfants dans les langues germaniques comme l'allemand et le hollandais. Elle montre également qu'avant la production des articles dans la forme standard reconnue à l'âge adulte, il y a la production de « fillers » constitués de structures monosyllabiques placées systématiquement avant les noms. Plusieurs données transversales (Braud, 2003 ; Wauquier-Gravelines, 2004) et longitudinales (Wauquier-Gravelines, 2004 ; Demuth & Tremblay, 2008) ont confirmé cette tendance en montrant que les enfants entre 12 et 26 mois produisent le nom commun avec la trace du déterminant défini interprétable comme un « proto-déterminant », sous-spécifié morphologiquement mais spécifié phonologiquement et prosodiquement (p.ex. production comme [lətɛ] pour *le chien* [le ʃjɛ̃]).

La prosodie du français permet, en effet, d'expliquer l'apparition de ce proto-déterminant ou « filler ». Cette langue a un rythme syllabique et 53% des noms sont monosyllabiques (vs. par exemple une autre langue romane comme l'italien où il y a 4% de monosyllabes). A l'oral, la prosodie induit une perception

des suites, déterminant + nom comme une seule unité prosodique que les enfants adaptent en produisant une structure dissyllabique accentuée sur la deuxième syllabe.

De plus, l'article français, qui est clitique, aide à distinguer le genre et le nombre dans la plupart des cas où ils ne sont pas morphologiquement et phonologiquement exprimés sur le nom (p.ex. *les phoques* [lə fɔk] vs. *le phoque* [lə fɔk] ; *le vase* [lə vaz] vs. *la vase* [la vaz]) et peut quasiment être considéré comme un préfixe détaché. D'ailleurs, au niveau syntaxique, le lien entre déterminant-nom dans le syntagme nominal est également considéré comme particulièrement étroit dans la structure des langues romanes (Chierchia, 1998). Ceci est spécialement vrai pour le français (Zribi-Hertz, 2011).

La force du lien déterminant-nom, la précocité d'apparition de ce déterminant et l'existence de resyllabations systématiques sur la frontière gauche de mot (comme la liaison) vont compliquer sérieusement la tâche de segmentation de l'input parlé et constituent un obstacle pour la production infantile. Cependant, on constate que ces erreurs de segmentation régressent avec l'âge et que s'installe progressivement une parfaite séparation du nom et du déterminant qui reflète la complète acquisition des deux formes séparées. Deux explications ont été données pour expliquer ce pattern de développement : une explication lexicale (Chevrot & Fayol, 2001) reposant sur la fréquence des mots et une explication phonologique autosegmentale (Wauquier & Braud, 2005 ; Wauquier, 2009).

L'explication lexicale de Chevrot & Fayol (2001) s'appuie sur une modélisation constructionniste. Ce modèle propose que les représentations internes du locuteur sont constituées de « constructions » créées selon le rapport entre forme sonore et sens que l'usage a rendu conventionnel. Ces constructions ont des niveaux d'abstraction variés. Elles peuvent être des séquences lexicales mémorisées globalement à partir de l'input phonétique, être associées aux catégories syntaxiques plus abstraites, comme la séquence sujet-verbe-objet (S-V-O), ou avoir des caractéristiques mixtes alliant des catégories abstraites et des éléments lexicaux comme par exemple « c'est+X ». Les constructions ne sont pas isolées les unes des autres mais vont progressivement s'organiser en réseau. Grâce à cette mise en réseau, des catégories plus abstraites peuvent émerger. A partir de ce cadre théorique, Tomasello (2003) propose une modélisation de l'acquisition de la syntaxe où les enfants mémorisent des « morceaux concrets de langage » ou *chunks* qui peuvent être des séquences phonologiques auxquelles ils ajoutent l'intention communicative et les informations contextuelles ou des séquences de mots stockées globalement. Des exemples de *chunks* sont les séquences « regardé (le) livre », « regardé maman », « regardé (la) télé » où le verbe « regarder » est stocké chez les enfants en différentes séquences selon son usage. Les enfants organisent ces éléments par leur mise en relation au travers de schémas locaux. Cela permet d'établir des relations entre *chunks* du type « regardé+X », « mangé+X » et « fini+X ». L'étape suivante prévoit une abstraction par la création d'un réseau de schémas et l'émergence d'une séquence comme « base verbale+i/é+Nom ».

Le même modèle est appliqué par Chevrot et al. (2013) à l'acquisition de la liaison. Il est proposé que d'abord entre 2 et 4 ans, l'enfant français mémorise des *chunks* fréquents constitués par déterminant + nom (p.ex. *les-[z]-avions*, *un-[n]-avion* ou *les canards*). Puis ces *chunks* sont mis en relation les uns avec les autres. Du point de vue fonctionnel, l'enfant construit des schémas généraux déterminant-nom (p.ex. *les + X* ou *un + X*) ; du point de vue formel, les noms ayant une consonne initiale qui est une consonne de liaison sont organisés entre eux et comparés avec les autres noms à consonne initiale lexicale (na.vjɔ̃, za.vjɔ̃ vs. naviR). La présence des différentes formes lexicales concurrentes explique les erreurs de consonne erronée dans la production infantile. Ensuite, entre 4 et 5 ans, les schémas deviennent plus spécifiques et, grâce à cela, l'enfant arrive à comprendre quelle variante du nom, dans sa mémoire, est associée à quel déterminant particulier (p.ex. *les* est suivi par des noms de type /zX/ vs. *un* /nX/). La disparition progressive des erreurs dans la production de liaison chez les enfants est interprétée comme la preuve de la réorganisation lexicale permettant une association pertinente entre déterminant et nom. Selon les auteurs, après cette phase l'enfant va rendre ses schémas plus abstraits et il va réussir à s'approprier une « construction » [DETERMINANT-z-[voyelle]-NOM] pluriel] (Bybee, 2001) pour s'en servir dans la production. Ce cadre théorique prédit par conséquent que les généralisations sur les noms plus fréquents doivent être maîtrisées avant les autres puisque plusieurs exemplaires du mot en différents contextes seront disponibles dans le lexique de l'enfant. Par conséquent on suppose que les liens et l'organisation en schémas seront plus fixes et stables.

On peut remarquer d'emblée que, si cette hypothèse permet de rendre compte de l'acquisition de la liaison dans les contextes du syntagme nominal [un + X] et [les + X], elle ne permet pas d'expliquer les autres contextes où il y a des rapports syntaxiques différents de [déterminant + nom] ou un contexte adjectival de type *les gentils éléphants*. De plus elle ne pose pas la question de la liaison en termes syllabiques et positionnels.

Si la première hypothèse prévoit une construction du lexique et une fixation des schémas selon la fréquence du mot, la deuxième hypothèse, dite morpho-phonologique, est formulée dans le cadre lexical de la « whole-word hypothesis » (Francescato, 1970 ; Ferguson & Farwell 1975 ; Menn & Matthei). Dans cette hypothèse l'acquisition d'un mot passe par des généralisations phonologiques précoces sans aucune médiation des autres unités phonologiques (à différence des *chunks*) mais au travers un gabarit (Vihman 1996). Le gabarit, ou « template » est la base spécifique au travers laquelle l'enfant peut « généraliser ».

C'est dans ce cadre que l'acquisition de la liaison a été formalisée par Wauquier (2009), Wauquier & Braud (2005). Grâce à ce concept de gabarit, l'acquisition de la liaison se déroule en quatre stades. Le premier stade concerne les enfants de 12 et 13 mois. Les associations dans les gabarits sont créées par les segments et les syllabes sont en relation de manière strictement alignée avec la position squelettale. Toutes les unités sont associées les unes aux autres dans un rapport de bi-univocité où ne sont présentes ni lignes autonomes, ni attaque branchante, ni coda (Figure 1).



Figure 1. Schéma d'acquisition de la liaison : le premier stade (Wauquier & Braud, 2005)

Ceci explique l'absence d'erreurs de liaison et soutient la conception selon laquelle la plupart des substantifs sont produits avec un déterminant non segmenté. Dans la deuxième étape, à 2 ans, l'input est mieux analysé et les unités phonologiques sont connues plus finement. Dans le gabarit, le contenu segmental est désassocié et diversifié et on retrouve la première manifestation d'autonomie entre information syllabique et information segmentale. Les enfants s'appuient sur les unités phonologiques les moins marquées (Peters, 1985) et remplissent l'attaque vide pour garder des schémas CV à l'attaque des mots. En accord avec cela l'enfant va rattacher la position prosodique de la consonne de frontière sur l'attaque du nom, pour maximaliser l'attaque au détriment de la coda précédente. La position syllabique reste associée au point squelettal et rattachée à l'attaque du nom. De plus, le contenu segmental est détaché et devient « flottant » sur la ligne segmentale. L'attaque vide est remplie par l'enfant avec une consonne qu'il a rencontrée dans ce contexte, une CL (« n » puis « l », « j » et « w » en ordre de fréquence) et comme conséquence la CL devient la consonne initiale du mot (Figure 2)

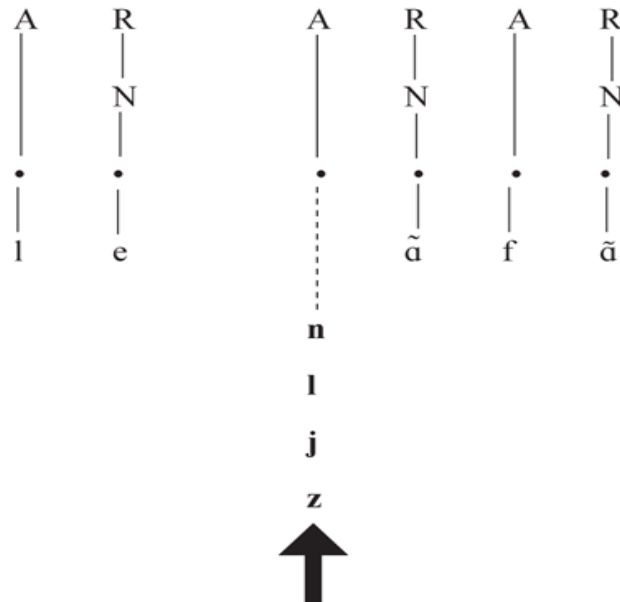


Figure 2. Schéma d'acquisition de la liaison : le deuxième stade (Wauquier & Braud, 2005)

C'est à ce stade que les erreurs systématiques de liaison sont enregistrées, dont spécialement l'utilisation de la consonne de liaison erronée et l'omission. À l'âge de 3 ans a lieu la troisième étape avec l'acquisition du double flottement sur le palier segmental et syllabique. Encore une fois, le gabarit change et l'enfant encode la liaison à la finale du déterminant mais sous la forme d'un segment flottant (Figure 3).

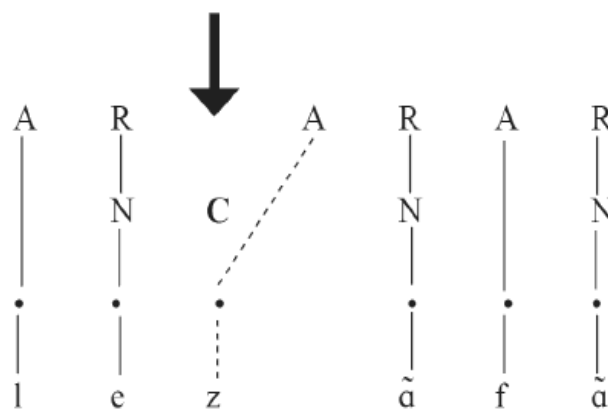


Figure 3. Schéma d'acquisition de la liaison : le troisième stade (Wauquier & Braud, 2005)

A ce stade un rôle central est joué par le *bootstrapping* morpho-phonologique : l'enfant utilise des informations morphologiques pour améliorer sa représentation phonologique. Par exemple, dans le cas de *les* dans la séquence *les enfants*, la CL [z] est encodée dans le déterminant comme morphème grammatical du pluriel qui porte un contenu segmental flottant. Ce contenu flottant peut remplir l'attaque

de la première syllabe du nom qui est vide s'il commence par une voyelle. De cette façon la consonne flottante est encodée dans la représentation lexicale sous-jacente et par conséquent les erreurs de liaison obligatoire vont peu à peu disparaître.

Pour les deux hypothèses présentées, les erreurs de segmentation en français continuent à baisser en nombre jusqu'à disparaître à l'âge de 7 ans. Cependant plusieurs études conduites chez des enfants de l'école primaire ont montré qu'ils continuent à avoir des problèmes à déterminer les frontières des mots et que cela émerge dans les premières formes d'écriture. Soum, (1997) en analysant les dictées des enfants du CE1 retrouve encore des erreurs de segmentation dans les syntagmes nominaux sujets à liaison. Si les enfants présentent un nombre non négligeable d'erreurs d'omission de la CL qui se resyllabe (p.ex. *le amis* pour *les[z]amis*) sur le mot à droite, la plupart des erreurs à l'écrit sont caractérisées par la présence d'une consonne qui va remplir l'attaque vide à droite (p.ex. *un lécureuil* pour *un[n]écureuil*) ou par la conservation de la CL simultanément dans le mot à gauche et le mot à droite (p. ex. *un gros sorage* pour *un gros[z]orage* ou *un nange* pour *un[n]ange*). La même problématique se manifeste pour l'enchaînement. Dans 50,6% des cas, les écrits des enfants du CE1 montrent un problème avec la frontière gauche (p.ex. *leu enfant* pour *leur enfant*). Si les erreurs entre déterminant et nom semblent disparaître avec la pratique de l'écriture, Sabio dans son étude (2000) remarque une persistance d'erreurs sur les frontières des mots en contextes différents dans le deuxième cycle de l'école primaire (*chez s'eu* pour *chez eux* ou *c'est qui se passent entre trelle* pour *entre elle*). Ces erreurs à l'écrit suggèrent que des problèmes de segmentation subsistent à des stades plus avancés que ceux que nous avons décrits plus haut. Ces problèmes se manifestent à l'écrit, mais peuvent découler d'une difficulté à maîtriser les phénomènes consonantiques oraux qui agissent entre les frontières des mots qui ne sont pas encore stabilisées. Le but de cet article consiste à étudier les types de fautes à ce moment charnière d'entrée dans l'écrit afin d'établir si les erreurs des enfants supposent plutôt des stratégies lexicales ou plutôt des stratégies morfo-phonologiques.

2 Hypothèses

Il a été également montré que les erreurs de segmentation vont disparaître à l'âge de 7 ans. Une raison peut en être l'entrée en école primaire, où le lexique et la phonologie deviennent plus stables avec l'apprentissage systématique d'une représentation orthographique des mots. Dans cet article, nous évaluons de manière détaillée l'importance relative de la phonologie et des connaissances lexicales des enfants à l'entrée du CP. Nous avons également voulu observer l'éventuel rôle du début de l'acquisition de l'orthographe dans la stabilisation des représentations lexicales dans une langue comme le français où les frontières de mots ne sont pas claires.

Nous avons testé des enfants francophones au début du CP afin d'observer d'une part s'ils ont tendance à utiliser une procédure lexicale ou, au contraire, une procédure phonologique pour segmenter les mots et d'autre part pour établir comment ils abordent les problèmes que leur posent les phénomènes consonantiques de liaison, d'élision et d'enchaînement.

3 Méthode

3.1 Participants

Nous avons testé 43 enfants français (âge moyen 6;3) qui seront suivis à partir du début du CP jusqu'au CE1. Les enfants font partie de deux classes d'une école primaire du Nord de la France. Pour recueillir des informations sociolinguistiques sur l'environnement dans lequel les enfants ont grandi et ont acquis la langue, nous avons distribué des questionnaires parentaux.

3.2 Procédure

Les deux classes du CP ont été testées en octobre 2013, au début de l'année scolaire. Nous avons proposé une tâche de dénomination guidée. Les enfants ont été testés un par un dans une salle silencieuse pendant 40 minutes. L'ordre de présentation du répertoire varie selon les enfants. Les performances audio ont été enregistrées par un enregistreur en format *.avi* avec une fréquence de 44000Hz.

Des phases d'entraînement précédaient la mise en forme de la tâche. Dans ces phases, l'expérimentateur expliquait à l'enfant la démarche et fournissait des exemples jusqu'à ce que la tâche soit comprise.

Tâche de dénomination guidée

Dix-huit couples de cartes avec des images colorées d'animaux et objets sont présentés un par un aux enfants. La moitié représente le test et l'autre moitié sont des *fillers*. Dans les 9 couples tests, 3 prévoient la production de liaisons, 3 d'élisions et 3 d'enchaînements. Dans chaque carte du couple, l'objet ou l'animal représenté est le même ; la seule différence est que dans l'un il y a un seul élément (p.ex. un lapin) et dans l'autre il y en a plusieurs (p.ex. des lapins), de telle sorte que l'enfant doit, soit produire le singulier à partir du pluriel, soit produire le pluriel à partir du singulier.

L'expérimentateur montre la première carte du couple et donne comme input le nom précédé par le déterminant, toujours l'article défini (p.ex. *le lapin*) et dans le cas d'enchaînement, avec l'adjectif antéposé (p.ex. *le petit chat*) ; ensuite il présente la deuxième carte du couple et demande à l'enfant de dire ce qui est dessiné en tenant compte de ce qu'il avait prononcé pour la première carte : la cible est donc la dénomination de la représentation dans la deuxième carte précédée par l'article défini ou aussi par l'adjectif associé (p.ex. *les lapins* ou *les petits chats*).

Dans les couples tests les inputs sont des mots en contextes de liaisons, élisions et enchaînements. Les enfants doivent déconstruire les phénomènes consonantiques pour isoler les mots. Ensuite, pour produire les cibles ils doivent ré-enchaîner ce qu'ils ont isolé. Les mots test donnés par l'expérimentateur et les réponses attendues par les enfants sont présentés dans le tableau 1.

Cette tâche est précédée d'un long entraînement de 9 couples pour faire comprendre à l'enfant qu'il doit nommer la deuxième carte en tenant compte de ce que l'expérimentateur a dit.

Input de l'expérimentateur : élision → Production enfant : liaison		
Contexte	Input : expérimentateur	Réponse attendue : enfant
Déterminant-nom	l'ours [lʊʁs]	les ours [lezʊʁs]
Déterminant-nom	l'armoire [laʁmwaʁ]	les armoires [lezɑʁ.mwaʁ]
Déterminant-nom	l'arrosoir [laʁɔzwaʁ]	les arrosoirs [laʁɔzwaʁ]
Input de l'expérimentateur : liaison → Production enfant : élision		
Contexte	Input : expérimentateur	Réponse attendue : enfant
Déterminant-nom	les oies [lezwa]	l'oie [lwa]
Déterminant-nom	les orages [lezɔʁaʒ]	l'orage [lɔʁaʒ]
Déterminant-nom	les oreillers [lezɔʁɛʒ]	l'oreiller [lɔʁɛʒ]

Input de l'expérimentateur : liaison → Production enfant : enchaînement		
Contexte	Input : expérimentateur	Réponse attendue : enfant
Adjectif -nom	les minuscules ogres [le minuskylzɔgʁ]	le minuscule ogre [le minuskylɔgʁ]
Adjectif-nom	les petites oranges [le petitɔʁɑ̃ʒ]	la petite orange [la petitɔʁɑ̃ʒ]
Adjectif-nom	les grosses araignées [le ɡʁos.zaʁɛne]	la grosse araignée [la ɡʁosazaʁne]

Tableau 1. Répertoire de la tâche de dénomination guidée

3.3 Matériel

Le répertoire de la tâche a été créé à partir de la consultation de la base de données du lexique des enfants MANULEX (Lété, Sprenger-Charolles & Colé, 2004). MANULEX fournit une liste de mots et de leurs fréquences à partir du CP jusqu'à la fin de l'école primaire grâce à l'analyse des occurrences de ceux-ci dans 54 manuels scolaires de lecture. Pour l'expérience, nous avons choisi des mots tests selon l'indice de fréquence standard (SFI) calculé à partir de la fréquence sur un million de mots¹. L'intervalle entre 50 et 65 représentent l'intervalle de la fréquence moyenne en MANULEX². La plupart des mots tests ont été sélectionnés dans cet intervalle. Quelques mots tests ont une SFI au-dessus de la moyenne (p.ex. *ours*). Malheureusement MANULEX ne donne pas la fréquence de séquences de mots mais seulement des mots isolés (p.ex. dans le couple *les amis* il y a la fréquence du mot *les*, du mot *amis* mais pas de l'ensemble) donc nous avons choisis uniquement les noms et les adjectifs tests de la tâche au travers cette base de donnée. Les autres catégories de mots utilisés et les contextes choisis pour la liaison ont été sélectionnés parmi les plus fréquents à partir de la classification proposée par Mallet (2008).

4 Résultats

De façon générale, la quantité d'erreurs en production (Tableau 2), montre que la tâche semble encore difficile à cet âge. Dans la tâche de dénomination 32,95 % d'erreurs ont été produites par les 43 enfants.

	% Erreurs	% Bonnes Réponses
Dénomination guidée	32,95	67,05

Tableau 2. Pourcentages d'erreurs et bonnes réponses dans la tâche de dénomination guidée

Dans le tableau 3 nous reportons en synthèse les erreurs de productions des enfants. Nous classifions les erreurs dans trois catégories : « Erreurs d'omissions » qui impliquent la non production du phénomène consonantique (p.ex. *le(s)[Ø]ours* [le uʁs] pour *les[z]ours* [lezuʁs], *la armoire* [la aʁ.mwaʁ] pour *l'armoire* [laʁ.mwaʁ] ou *la grosse[Ø]araignée* [la ɡʁos aʁne] pour [la ɡʁosazaʁne]); « Erreurs d'insertion d'une consonne erronée », où les enfants produisent une consonne différente de celle attendue dans les contexte de liaison, élision et enchaînement (p.ex. *le(s)[n]ours* [lenuʁs] pour *les[z]ours* [lezuʁs], *le [z]armoire* [lə.zaʁ.mwaʁ] pour *l'armoire* [laʁ.mwaʁ] ou *la grosse[z]araignée* [la ɡʁosazaʁne] pour [la ɡʁosazaʁne]); « Autres erreurs » qui sont plutôt des répétitions de l'input donné ou de mauvaise utilisations d'articles (p.ex. *le grosse araignée* [lə ɡʁosazaʁne] pour [la ɡʁosazaʁne]). Les pourcentages pour chaque type d'erreurs dans la production de chaque phénomène consonantique sont calculés sur le nombre total d'erreurs de production.

	Liaison	Élision	Enchaînement
Erreurs d'omissions	13,2 %	27,7 %	11,6 %
Erreurs d'insertion d'une consonne erronée	3,5 %	3 %	9,4 %
Autres erreurs	0 %	18,6 %	13 %
TOT. Erreurs	16,7 %	49,3 %	34 %

Tableau 3. Types d'erreurs et pourcentages dans la dénomination guidée pour les trois phénomènes

Si nous analysons les comportements des enfants pour le contexte déterminant-nom (donc pour élision et liaison), les principales erreurs sont l'omission de toutes les consonnes qui se resyllabifient sur la frontière gauche des mots commençant par une voyelle : pour la liaison (*le(s)[Ø]ours* [le uʁs] pour *les[z]ours* [lezuʁs]), pour l'élision (*la armoire* [la aʁ.mwaʁ] pour *l'armoire* [laʁ.mwaʁ]).

Si on regarde les erreurs totales sur les mots tests, la production d'élision est ce qui est le plus difficile avec 49.30 % des erreurs.

En regardant la production de liaisons, même si le pourcentage est inférieur aux autres contextes (16,7 % des erreurs totales) les erreurs continuent à se manifester. 13,2 % sont des erreurs d'omission, soit (*le(s)[Ø]arrosiers* [le a.ʁɔ.zwaʁ] pour *les[z]arrosiers* [le.za.ʁɔ.zwaʁ]) et 3,5 % sont des erreurs de consonne de liaison erronée (*le(s)[n]ours* [lenuʁs] pour *les[z]ours* [lezuʁs]). La consonne de liaison erronée insérée entre déterminant et nom est soit une possible consonne de liaison ([z] ou [n]), soit la consonne de l'article défini ([l]) ou une consonne arbitraire qu'on ne retrouve pas dans le nom, comme par exemple [d]. Une fois de plus, on constate que les erreurs d'omission prédominent sur les autres.

Pour l'élision la plupart des enfants font des erreurs d'omission, c'est-à-dire de hiatus (27,7 %). Donc s'ils ont à leur disposition une forme input comme *les[z]oreillers* [le.zɔ.ʁɛ.jɛ], ils produisent plutôt *le oreiller* [lə ɔ.ʁɛ.jɛ] que la cible *l'oreiller* [lɔ.ʁɛ.jɛ]. De plus beaucoup d'enfants répètent l'input (18,6 % dans « Autres erreurs »), sans produire de changement phonologique (p. ex. input de l'expérimentateur : *les oies* [le.zwa], réponse de l'enfant : *les oies* [le.zwa]). Enfin, les erreurs où la consonne de liaison est gardée comme initiale du mot dans la cible (p.ex. *la[z]oie* [la.zwa] ou *la[n]oie* [la.nwa] pour *l'oie* [lwa]) restent marginales (3 %).

De plus, le passage de l'occurrence en contexte de liaison (déterminant-nom au pluriel) au passage de l'occurrence en contexte d'élision (déterminant-nom au singulier) est plus difficile que l'inverse pour des questions de masquages de la frontière entre déterminant et nom dûs à la resyllabification impliquée par la liaison (49,3% des erreurs dans la production d'élision vs. 16,7 % des erreurs dans la production de liaison).

La production d'enchaînement en contexte adjectif-nom (p.ex. *la grosse araignée* [la ɡʁosɑʁɛne]) donne 49 erreurs, soit 34 % des erreurs totales. Les problèmes sont différents de ceux que pose la liaison. 13 % des erreurs sont classifiés comme « autres » et montrent la tendance à utiliser un article inapproprié comme un masculin pour un féminin (*le grosse araignée* pour *la grosse araignée*) ou un pluriel sur une image au singulier (*les grosse araignée*). Au contraire 9,4 % des erreurs totales sont encore des erreurs de frontières qui maintiennent la consonne de liaison en enchaînement (*la grosses[z]araignée* [la ɡʁo.za.ʁɛ.ne]). Les 11,6 % restantes sont des cas d'omission d'enchaînement, où ce phénomène consonantique n'est pas produit (*la grosse araignée* [la ɡʁos ɑʁɛne]).

5 Le rôle de la fréquence

Nous pouvons évaluer le rôle de la fréquence au travers la mise en relation des fréquences SFI des items test en MANULEX et les nombres d'erreurs dans la tâche de dénomination guidée (Tableau 4).

Nom au singulier	SFI au singulier	Nom au pluriel	SFI au pluriel	Nombre d'erreurs d'omission	Nombre d'erreurs de consonne initiale erronée		
					[z]	[n] ou [l]	Autres consonnes
orange	58,64	oranges	53,40	8	6	0	0
araignée	57,88	araignées	47,20	8	3	0	0
oie	57,59	oies	55,20	7	19	2	1
armoire	58,10	armoires	45,10	2	0	0	0
ogre	56,77	ogres	0	10	15	2	0
oreiller	57,88	oreillers	38,75	18	1	1	0
arrosoir	56,59	arrosiers	37,95	7	0	0	0
orage	58,28	orages	38,75	15	0	1	0

Tableau 4. Fréquence des mots tests et nombre d'erreurs dans la tâche de dénomination guidée

Le tableau 4 montre la fréquence des items tests, au singulier et au pluriel. Dans les items tests on ne considère pas le mot *ours*, car dans la base de données utilisée, sa fréquence n'est pas donnée pour les deux formes que nous analysons. Nous rappelons que les mots avec une SFI entre 50 et 65 sont traités comme des mots avec une fréquence moyenne. Les mots avec une SFI inférieure à 50 ont une basse fréquence et ceux avec une moyenne supérieure à 65 sont des mots de haute fréquence. Seulement deux mots (*orange* et *oie*) ont des formes singulières et des formes plurielles des fréquences moyennes. Au contraire, dans les autres couples, les items ont une fréquence moyenne au singulier et une fréquence basse au pluriel. Donc c'est seulement la fréquence au pluriel qui change (de basse à moyenne).

Nous avons ensuite classifié les erreurs selon qu'elles sont « omission » ou « consonne erronée ». Dans cette dernière catégorie, nous avons identifié 3 sous-groupes, selon le type de consonne erronée utilisée entre frontières des mots : la consonne erronée [z], plutôt associée au pluriel dans la resyllabation par liaison ou enchaînement; [n] et [l] qui apparaissent en général avec la forme singulière du nom en cas de liaison, élision et enchaînement; autres consonnes, présentes dans nos données mais qui ne peuvent pas être impliquées dans les phénomènes consonantiques testés (p.ex. [d]).

Si on considère l'explication lexicale décrite dans l'introduction, la fréquence de surface (donc la fréquence des deux formes séparément) influence les productions et la segmentation de l'enfant en cas de présence de phénomènes consonantiques. Donc on s'attendrait à ce que l'enfant produise davantage de [z] au début du nom lorsque le mot est plus fréquent au pluriel qu'au singulier, et des omissions ou mauvaises consonnes de liaison (comme [n] et [l]) si le mot est plus fréquent au singulier. Si nous analysons nos données il n'y a pas une corrélation entre les fréquences des formes au pluriel (et le nombre de productions avec la consonne erronées [z] ($r = .21, p > .60$).

Dans les couples où la fréquence au pluriel est basse et la fréquence au singulier moyenne, les erreurs d'omission et de consonne erronée [n] ou [l] sont supérieures aux erreurs de consonne erronée en [z].

Cependant, nous trouvons une exception avec le mot *ogre*. MANULEX n'indique aucune occurrence pour sa forme plurielle au cours du CP même si dans sa forme singulière il est présent en fréquence moyenne. Comme on peut voir dans le tableau, pour cet item, le nombre d'erreurs d'omission et de consonne erronée en [n] ou [l] est presque le même que le nombre d'erreurs de consonne erronée en [z]. Le cas du mot *ogre* montre donc une tendance différente des trois autres dont il faut tenir compte.

Le rapport entre fréquence et nombres d'erreurs pour la tâche de dénomination ne permet pas de considérer l'explication lexicale, basée sur le modèle exemplariste, comme la seule explication possible pour la détection des frontières des mots. Cependant d'autres tâches où la fréquence de formes singulières et plurielles des mots utilisés est plus contrôlée et où les nombres d'items tests est plus élevé doivent être conduits pour mieux évaluer le rôle de la fréquence dans les productions orales au CP.

6 Discussion / Conclusion

Les résultats montrent que les enfants au début du CP ont encore des problèmes avec le calcul des frontières de mots en français, surtout quand la tâche demandée repose sur une segmentation en mots.

La tâche de production proposée est une tâche complexe parce qu'elle suppose que l'enfant réalise une série d'opérations pour une bonne exécution : une première analyse de l'input, un détachement successif du nom et du déterminant et l'enchaînement du nom isolé avec un autre déterminant ou un adjectif entraînant une resyllabation différente. Une mauvaise production finale dans cette tâche est donc le possible reflet d'une mauvaise segmentation de départ. Nous constatons néanmoins que dans le doute, les enfants ont tendance à préserver la forme canonique du nom en acceptant les omissions.

La typologie des erreurs observées dans nos données aussi bien pour les omissions, que pour les élisions, les enchaînements et les liaisons, en production, montre que la chronologie proposée par Chevrot et al. 2013, Wauquier, 2009 pour les enfants plus jeunes peut être complétée et étendue au-delà de l'acquisition de l'oral. Nos résultats tendent à montrer l'existence d'un stade spécifique de l'acquisition des phénomènes consonantiques et des frontières de mot. En effet, comme on l'a vu dans la littérature sur la liaison, les erreurs chez les enfants entre 2 et 4 ans sont des erreurs de consonne de liaison erronée plus que des erreurs d'omission et que ces dernières sont les seules qui subsistent après cet âge. Nos résultats montrent effectivement que, dans le contexte déterminant-nom les erreurs de consonne erronée ont globalement disparu à l'âge de 7 ans. Mais ils montrent aussi en fait qu'il y a une persistance des erreurs liées à la difficulté que posent la segmentation et l'établissement des frontières lexicales après 6 ans et demie mais qu'elles sont d'une autre nature : les erreurs du type *le(s)[n]ours* [lenuks] deviennent essentiellement des erreurs d'omission du type *le(s)[Ø]ours* [le uks]. Le problème résiduel que nous avons remarqué dans la tâche de dénomination guidée au début du CP ne peut pas être limité au lexique, vu que la fréquence lexicale des items tests n'explique pas toutes les erreurs et qu'il n'est pas possible d'établir une corrélation entre la fréquence de mots en forme pluriel et le nombre d'erreurs d'insertion d'une consonne erronée [z].

Cette analyse s'applique pour l'instant au seul contexte déterminant-nom. Le contexte adjectif antéposé-nom fait émerger des erreurs de mauvais articles (p.ex. le petite orange) impliquant apparemment également la morphologie et la maîtrise du genre. Des données plus complètes doivent être apportées pour pouvoir généraliser cette analyse à d'autres contextes que le contexte déterminant-nom. Mais à ce stade, les résultats obtenus pour la tâche de dénomination d'image de notre étude longitudinale affinent la vision des étapes nécessaires à la résolution des problèmes que pose la segmentation des mots pour un enfant apprenant le français. Les erreurs attestées à l'écrit à l'école primaire (Soum, 1997 ; Sabio, 2000) et nos premières données sur l'oral au début du CP confirment l'importance de rendre compte du rôle que joue la phonologie et non pas exclusivement du lexique dans ces processus de régularisation. D'ailleurs ceci recoupe les conclusions de Sprenger-Charolles (1994) qui énonce que la première phase d'acquisition de la lecture repose essentiellement sur une analyse phonologique indépendante de la fréquence lexicale.

Références bibliographiques

- Bassano, D. (2010). L'acquisition du déterminant nominal en français : une construction progressive et interactive de la grammaire. *Cognitextes* [En ligne], 5, URL : <http://cognitextes.revues.org/315>
- Basset, B. (2000). *La liaison à 3, 7 et 11 ans : description et acquisition*. Mémoire de Maîtrise Sciences du Langage, Grenoble 3
- Bechade, H.-D. (1992). *Phonétique et morphologie du français moderne et contemporain*. Paris : Presses Universitaires de France
- Bosse, M.-L. (2005). De la relation entre acquisition de l'orthographe lexicale et traitement visuo-attentionnel chez les enfants. *L'orthographe lexicale*, 222, 9-30
- Braud, V., (2003). *Acquisition de la prosodie chez des enfants francophones : les phénomènes de troncations*. Thèse en Sciences du Langage, Université de Nantes
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Chevrot, J.-P., Dugua, C., Harnois-Delpiano, M., Siccardi, A., & Spinelli, E. (2013). Liaison acquisition: debates, critical issues, future research. *Language Sciences*, 39, 83-94.
- 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., Fayol, M. & Dugua, C. (2005). Liaison et formation des mots français : un scénario développemental. *Langages*, 158, 38-52
- Chevrot, J.-P., & Fayol, M. (2001). Acquisition of French Liaison and Related Child Errors. *Selected paper of the 8th congress for the study of child language*, San Sebastian, juillet 1999
- Chierchia, G. (1998). Reference to Kinds across Languages. *Natural Language Semantics*, 6(4), 339-405.
- Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determinants. *Journal of Child Language*. 35 (1), 99-127
- Dugua, C., Chevrot, J.-P. & Fayol, M. (2006). Liaison, segmentation des mots et schémas syntaxiques entre 2 et 6 ans : un scénario développemental, *Entretiens de Bichat : Orthophonie*, 230-244.
- Dugua, C., & Nardy, A. (2003). Acquisition des liaisons et rattachement lexical de la consonne de liaison dans la formation du lexique en français, in F. Mouret C. Raynal et M. Tellier (éd.). *Actes du 8^{ème} Atelier des Doctorants en Linguistique*, Paris, Denis Diderot Université Paris 7, 55-61
- Francescato, G. (1970). *Il linguaggio infantile. Strutturazione e apprendimento*. Torino : Einaudi
- Ferguson, C., & Farwell, C. (1975). Words and Sounds in Early Language Acquisition. *Language*, 51, 419-439
- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary-school readers. *Behavior Research Methods, Instruments, & Computers*, 36, 156-166
- Mallet, J.-M. (2008). *La liaison en français : descriptions et analyses dans le corpus PFC*. Thèse en Sciences du Langage, Paris Ouest Nanterre
- Menn, L., & Matthei, E. (1992). The "two-lexicon" model of child phonology: Looking back, looking ahead. in Ferguson, Menn, & Stoel-Gammon (eds.), *Phonological Development: Models, Research, Implications*, 211-247
- Nardy, A., & Dugua, C. (2011). Le rôle de l'usage sur le développement des constructions nominales chez les enfants pré-lecteurs. *Travaux de Linguistique*, 129-148
- Peters, A., M. (1985). *Language segmentation: Operating principles for the analysis and perception of language*, *The Crosslinguistic Study of Language Acquisition*, 2, 1029-1067
- Pulgram, E., (1970). *Syllable, Word, Nexus, Cursus*. Paris.Mouton : The Hague
- Sprenger-Charolles, L. (1994). L'acquisition de la lecture en français : étude longitudinale de la première à la seconde année du primaire. *L'Année Psychologique*, 94(4), 553-574.

- Sabio, F. (2000). Les difficultés de la notion de mot : l'exemple des liaisons graphiques dans les textes d'enfants, *Linx*, 42, 1-10
- Suom, C. (1997). L'apprentissage de l'écriture ; contraintes orthographiques ; contraintes orales. Thèse de doctorat, Université de Toulouse 2
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press
- Vihman, M. (1996). *Phonological Development: The origins of language in the child*. Oxford : Blackwell
- Wauquier, S. (2013). Convergence and divergence in the acquisition of French liaison by native and non-native speakers, a review of existing data and avenues for future research. *Language, Interaction and Acquisition* 4:2, 160-188
- Wauquier, S. (2010). Acquisition de la phonologie « du » français : des usages au système. *Revue Langue Française*, 1-20
- Wauquier, S. (2009). Acquisition de la liaison en L1 et L2 : stratégies phonologiques ou lexicales?. *Aile, Lia* 2, 93-130
- Wauquier-Gravelines, S., & Braud, V. (2005). Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages*, 158, 53-65
- Wauquier-Gravelines, S. (2004). Les liaisons dangereuses. Psycholinguistique et phonologie : une interface complexe, in Ferrand, L. & Grainger, J. (eds.) *Psycholinguistique cognitive, Essais en l'honneur de Juan Segui*, Bruxelles : De Boeck Université, Collection Neurosciences et Cognition
- Zribi-Hertz, A. (2011). Definite DPs without lexical nouns in French: clausal modifiers and Relativization. in J. Berns, H. Jacobs & T. Scheer (eds.), *Romance languages and linguistic theory, selected papers from Going Romance* 23, 363-390. Amsterdam : John Benjamins

¹ L'indice SFI donne le logarithme de la fréquence sur un million de mots sur la quelle on peut faire foi parce qu'il est en rapport avec l'estimation de la fréquence pour un million de mots (U) et au travers le logarithme il était statistiquement équilibré.

² En absence d'un modèle statistique précise et sans pouvoir tenir compte de médian ou écart-type, on a choisi de regarder la distribution de SFI pour trouver l'intervalle des fréquences moyennes. Si on regarde la distribution de SFI l'intervalle entre 50 et 65 correspond à la fréquence moyenne.