

Explorer des corpus oraux à l'aide de la base de données Démonette-2 : usage de mots construits dans des interactions adulte(s)-enfant(s)

Exploring oral corpora using the Démonette-2 database: use of constructed words in adult(s)-child(ren) interactions

Stéphanie Caët

Univ. Lille, CNRS, UMR 8163 – STL – Savoirs, Textes, Langage

stephanie.caet@univ-lille.fr

Caroline Masson

Université Sorbonne Nouvelle, EA 7345 CLESTHIA

caroline.masson@sorbonne-nouvelle.fr

Loïc Liegeois

Université Paris Cité, CNRS, EA 3967 CLILLAC-ARP et UMR 7110 LLF

loic.liegeois@u-paris.fr

Lucie Macchi

Univ. Lille, CNRS, UMR 8163 – STL – Savoirs, Textes, Langage

lucie.macchi@univ-lille.fr

Christine Da Silva-Genest

Université de Lorraine & EA 3450 DevAh

christine.da-silva-genest@univ-lorraine.fr

Nadejda Okinina

SONOS VOX France

n.okinina@yahoo.fr

Résumé

Cet article illustre l'usage que l'on peut faire de la base de données Démonette-2 (Namer et al., ce numéro) pour explorer des corpus oraux d'interactions adulte(s)-enfant(s). Pour ce faire, nous poserons deux questions susceptibles d'être soulevées en acquisition du langage : (i) quelle est la fréquence de mots construits dans les productions spontanées d'adultes et

d'enfants ? (ii) quelles sont les pratiques des adultes pour étayer la compréhension de mots construits ? Les corpus explorés, en français de France, transcrits au format CHAT de CLAN, contiennent des interactions parent(s)-enfant au développement typique et atypique, des interactions en crèche et des interactions orthophoniste-enfant, recueillies lors d'activités diverses (jeux, lecture, repas, activités métalinguistiques par exemple). Cet article intéressera des chercheurs travaillant sur corpus oraux ainsi que des professionnels du langage, notamment des orthophonistes.

Mots-clés : morphologie, dérivation, corpus, interactions, acquisition

Abstract

This paper illustrates how the Démonette-2 database (Namer et al., this issue) can be used to explore oral corpora of adult-child interactions. To do so, we will ask two questions that are likely to be asked in child language acquisition research: (i) what is the frequency of constructed words in the spontaneous productions of adults and children? (ii) what language practices do adults set up to support the understanding of constructed words? The French corpora explored, transcribed with CLAN in the CHAT format, contain interactions between children presenting a typical or atypical development of language and their parents, interactions in daycare centers or during speech and language therapy sessions. The data was collected during various activities (e.g., playtimes, readings, mealtimes, metalinguistic activities). This paper will be of interest to researchers as well as to professionals working on oral corpora, particularly speech and language therapists.

Keywords: morphology, derivation, corpus, interactions, acquisition

Financement : Cette recherche a été financée par l'Agence Nationale de la Recherche (France) dans le cadre du projet DEMONEXT [ID : ANR-17-CE23-0005].

1. Introduction

Dans le cadre du projet Demonext, une base de données morphologiques décrivant les propriétés dérivationnelles des mots de manière systématique a été créée la base Démonette-2 (Namer et al., ce numéro). De nombreuses informations sont associées aux lexèmes qui la composent : morphologiques, sémantiques, phonologiques. Ces informations linguistiques, complétées par des informations fréquentielles (en corpus), permettent ainsi une exploitation pédagogique ou orthophonique de la base (Cattini & Duboisindien, 2021). Dans cet article,

nous proposons d'illustrer la manière dont Démonette-2 peut être utilisée pour exporter les mots construits inventoriés dans la base et explorer des corpus oraux afin d'interroger leur usage dans ces corpus. Cette utilisation sera illustrée à partir de questionnements issus de recherches et de pratiques pédagogiques ou cliniques en lien avec le développement typique ou atypique du langage chez l'enfant.

Dans le cadre de l'étude du développement du langage de l'enfant, l'analyse de corpus d'interactions orales spontanées vise à rendre compte de la manière dont les enfants acquièrent à la fois la langue et son usage, en contexte. Si les perspectives théoriques adoptées peuvent différer (Ambridge & Lieven, 2011 ; Salazar Orvig, 2019), accordant une place plus ou moins centrale au rôle de la fréquence des constructions entendues par l'enfant ou au rôle de la dynamique interactionnelle et de la co-construction du sens, elles reposent sur l'idée fondamentale que le processus acquisitionnel est ancré dans l'interaction, notamment avec des locuteurs dits « experts ».

Le développement des compétences morphologiques dérivationnelles est décrit comme relativement précoce et graduel (Fejzo et al., 2018). Vers 2 ou 3 ans, les enfants commencent à employer des affixes (préfixes ou suffixes en français) de façon productive, comme en attestent les néologismes observés dans leurs productions orales spontanées (Clark, 1993). Dès l'école maternelle, ils possèdent des connaissances morphologiques implicites sur la formation des mots (Rassel, 2020). Ils développent ensuite progressivement une conscience morphologique leur permettant d'employer de façon comparable aux adultes les affixes et mots construits. Ce développement se prolonge tout au long de leur scolarité (Casalis & Louis-Alexandre, 2000 ; Colé et al., 2004).

Si la plupart des études soulignent l'importance de la conscience morphologique pour le développement des compétences en langage écrit (Carlisle et al., 2010 ; Fejzo et al., 2018 ; Rassel, 2020), nous disposons de peu d'informations quant à la manière dont les mots construits sont produits et progressivement acquis par les enfants en interaction. Nous proposons donc ici d'illustrer l'utilisation de Démonette-2 pour explorer des corpus d'interactions adulte(s)-enfant(s), en considérant : i) la fréquence de mots construits dans les productions d'adultes et d'enfants ii) la manière dont des adultes étayent la compréhension de mots construits par les enfants.

L'objectif principal de cet article est de présenter les possibilités méthodologiques et techniques liées à l'utilisation de la base Démonette-2 pour l'exploration de corpus d'acquisition. Il s'agira donc d'une première entrée dans ces questions visant à ouvrir des pistes de recherche future.

La méthodologie présentée ci-après a été appliquée à cinq corpus d'interactions spontanées (Tableau 1) :

- les corpus CoLaJe (Morgenstern & Parisse, 2012) et ALIPE (Liégeois et al., 2014), corpus longitudinaux d’interactions parent(s)-enfant disponibles sur ORTOLANG¹, explorés pour l’illustration d’analyses quantitatives (1 million de tokens environ) ;
- trois corpus explorés pour l’illustration d’analyses qualitatives (200000 tokens environ) : le corpus RaProChe (Masson, 2022) constitué d’interactions en crèche ; le corpus APPEL IOE (Caët et al., 2021) composé d’interactions en contexte orthophonique ; le corpus EVALANG (da Silva-Genest & Le Mené, 2018) constitué d’interactions en situation de jeu symbolique et de récits d’expériences personnelles entre un adulte et un enfant avec ou sans trouble développemental du langage (TDL – Bishop et al., 2017).

Corpus	Âge des enfants	Nombre d’enfants	Développement langagier	Interlocuteur principal	Volume total du corpus	Nombre de tokens analysés
CoLaJE	0-8;0	6	typique	parent(s)	180 h	845 742
ALIPE	2;4-5;4	3	typique	parent(s)	30 h	174 209
EVALANG	4;6-8;5	35	typique	parent ou	14 h	114 608
		11	TDL	expérimentateur	4 h	28 460
RaProChe	0;4-3;6	260	typique	professionnelle petite enfance	79 h	33 952
APPEL IOE	4;0-5;0	2	atypique	orthophoniste	4 h	17 129

Tableau 1. Informations développementales, contextuelles et matérielles relatives aux corpus traités.

2. Méthodologie

L’un des outils de transcription et d’analyse de corpus souvent utilisé dans les recherches sur les interactions verbales avec des enfants est le programme CLAN (MacWhinney, 2000). La méthodologie introduite ici s’applique donc à des corpus transcrits à l’aide de ce programme et structurés au format CHAT (Codes for the Human Analysis of Transcripts). Le programme CLAN permet notamment d’étiqueter les données transcrites au niveau morphosyntaxique, la commande « mor » réalisant la lemmatisation et la catégorisation morphosyntaxique des tokens en s’appuyant sur des grammaires développées à cet effet². Appliqué au français, cet étiquetage ne permet cependant pas de repérer aisément dans les corpus un ensemble de mots construits, en particulier suffixés. La Figure 1 ci-dessous illustre l’étiquetage réalisé pour un

¹ <https://www.ortolang.fr>

² disponibles sur le site du projet CHILDES (<https://talkbank.org/morgrams/>) ou sur <https://ct3.ortolang.fr/tools/> pour la grammaire du français.

énoncé contenant le mot *fermier* : celui-ci est identifié comme un nom masculin (n|fermier&m), mais il n'est pas identifié comme un nom suffixé et rien ne permet de le distinguer d'un nom non suffixé se terminant par *ier*, comme *dernier* par exemple (étiqueté n|dernier&m).

*CHI : *c'est le jardin du fermier.*
 %mor : pro:dem|ce\$v:aux|être&PRES&3s det:art|le&m&sg n|jardin&m prep|de&le
 n|fermier&m.

Figure 1. Exemple d'étiquetage morphosyntaxique réalisé par la commande « mor » de CLAN.

Si la base de données Démonette-2 répertorie quant à elle un ensemble de lexèmes construits, elle n'a pas été conçue pour permettre l'exploration de corpus. Le programme CLAN et la base de données Démonette-2 ayant été développés de façon totalement indépendante, la procédure présentée ci-après constitue une première tentative de mise en relation de ces ressources, dont les limites et les perspectives seront soulignées dans la partie discussion de cet article.

2.1. Extraction des mots de la base

Pour cet article, nous nous sommes focalisés sur les suffixes. 13 suffixes, présents dans la base Démonette-2, ont été explorés :

- *-euse* et *-rice* ainsi que *-age* et *-ion*, qui ont fait l'objet d'une attention particulière dans le cadre du projet Demonext ;
- *-iste*, *-ure* et *-ier* qui, avec *-age*, font partie des lexèmes construits les plus fréquemment utilisés par les orthophonistes selon l'enquête de Duboisdindien et Dal (2021) ;
- *-ette*, et *-eur*, deux des suffixes fréquemment utilisés par l'une des co-auteurs de cet article dans le cadre de sa pratique orthophonique avec des enfants présentant un trouble développemental du langage oral ;
- ainsi que *-et*, *-ière* et *-oire*.

Les listes de lexèmes suffixés ont été obtenues grâce à l'outil « Dérivations suffixales » présent dans la section « Outils spécifiques » de la base Démonette-2³. Après avoir saisi le suffixe ciblé dans l'espace « suffixe est » (Figure 2), l'ensemble des lexèmes concernés est généré.

³ <https://demonette.fr/demonext/vues/suffixes.php>

dérivé	base	suffixe
dérivé contient	base contient	suffixe contient
dérivé est	base est	suffixe est
dérivé commence avec	base commence avec	suffixe commence avec
dérivé finit avec	base finit avec	suffixe finit avec
asinién	Adj	Npx
asinién	Nim	Npx
aariite	Nf	Npx
abactérien	Adj	Nf
abaissable	Adj	V
abaissement	Nm	V
abaisseur	Nm	V

Figure 2. Page de consultation des dérivations suffixales de la base Démonette-2 sur le site du projet Demonext.

Tout ou partie du résultat peut être exporté dans différents formats : .txt, .csv, .xlsx, .pdf. Pour permettre l'exploration de corpus avec CLAN, un export au format .cut, spécifique à l'outil CLAN, a été introduit. Dans le format d'export proposé, les lexèmes extraits sont précédés et suivis d'un codage (Figure 3) qui rend possible la recherche de ces lexèmes sur la ligne comportant l'étiquetage morphosyntaxique (ligne %mor) des fichiers structurés au format CHAT et étiquetés au niveau morphosyntaxique.

```
m;abricotière,o-%
m;aigrière,o-%
m;aigrière,o-%
m;aisselière,o-%
m;alevinière,o-%
```

Figure 3. Cinq premières lignes du fichier .cut pour le suffixe *-ière*.

2.2. Exploration(s) des fichiers .cha

Une fois les fichiers .cut générés (un fichier par dérivation suffixale), ceux-ci sont placés dans le sous-dossier « lib » du dossier « Clan ». Les fichiers .cut peuvent alors être appelés par des commandes CLAN. Pour obtenir la fréquence des lexèmes qui nous intéressent, nous utilisons la commande « freq ». Par exemple, afin d'obtenir la fréquence des lexèmes issus d'une dérivation en *-ière* (ex. *fermière*, *crêpière*), nous utilisons la commande suivante :

```
freq + t*CHI + t*MOT + t*FAT + s@Xiere.cut + d2 *.cha
```

Cette commande permet d'obtenir la fréquence d'usage (freq) des lexèmes contenus dans le fichier « Xiere.cut » (+s@Xiere.cut) pour l'enfant-cible (+t*CHI), sa mère (+t*MOT) et son père (+t*FAT). Cette commande s'applique à chacun des fichiers CHAT présents dans le

dossier de travail (*.cha) et les résultats sont générés dans un format tableur (+d2).⁴ Afin d'afficher les lexèmes en contexte dans le but, par exemple, de mener des analyses qualitatives, la commande « kwal » peut être utilisée. Cette commande suit une syntaxe similaire à celle de la commande « freq » et permet également d'appeler un ensemble de lexèmes contenus dans un fichier .cut :

```
kwal + t*CHI + t*MOT + t*FAT + s@Xiere.cut *.cha
```

3. Analyses

Pour étudier le processus d'acquisition du langage, les corpus d'interactions spontanées peuvent donc être explorés de différentes manières : quantitativement et qualitativement. La section 3.1 ci-dessous présente les résultats d'une analyse de la fréquence des suffixes explorés dans les productions des parents et des enfants ; la section 3.2 introduit des analyses qualitatives de séquences dans lesquelles des mots dérivés ont été produits.

3.1. Analyse de la fréquence de mots suffixés dans les corpus

Au sein des corpus CoLaJE et ALIPE, nous avons sélectionné les fichiers de transcription dans lesquels des données pour les enfants et pour (au moins un de) leurs parents étaient disponibles, soit 170 et 70 fichiers respectivement, et près d'un million de tokens. Chaque fichier correspond à l'enregistrement d'une dyade à un âge donné (enfants âgés de 0 à 8 ans) ; ces corpus étant longitudinaux, les mêmes dyades ont été enregistrées plusieurs fois.

Dans les productions des parents, nous avons relevé 3 327 mots construits avec l'un des 13 suffixes. Parmi ceux-ci, les plus représentés sont *-ette* (23 % des mots suffixés étudiés ; $n = 769$) et *-ier* (21 % des mots suffixés étudiés ; $n = 712$). Certains suffixes sont réalisés dans une grande diversité de constructions, comme le suffixe *-ier* qui, au sein d'un enregistrement, pourra parfois être observé dans la construction de 9 mots différents, tandis que d'autres ne sont observés que dans une seule construction par enregistrement (c'est le cas de *-rice*) voire dans une seule construction sur l'ensemble du corpus (comme *-oire* observé uniquement dans *obligatoire*).

Le Tableau 2 renseigne quant aux proportions de mots suffixés relevés pour chacun des 13 suffixes étudiés, par rapport à l'ensemble des mots suffixés observés, dans les productions des parents d'enfants de 1 à 5 ans. Le nombre de fichiers par tranche d'âge varie entre 22 et 95, et le nombre de dyades par tranche d'âge varie entre 5 et 8.

Suffixes	1 an	2 ans	3 ans	4 ans	5 ans
----------	------	-------	-------	-------	-------

⁴ Afin de lancer les 13 commandes en une seule fois, nous avons utilisé la possibilité offerte par CLAN de réunir ces dernières dans un fichier .bat, que nous avons nommé « morpho_suffixes.bat ». Une seule commande suffit alors pour générer les 13 fichiers de résultat : batch morpho_suffixes.bat.

<i>-ière</i>	3 %	3 %	2 %	3 %	0 %
<i>-ier</i>	14 %	19 %	23 %	39 %	27 %
<i>-ure</i>	3 %	4 %	4 %	5 %	4 %
<i>-iste</i>	1 %	1 %	2 %	2 %	2 %
<i>-age</i>	6 %	8 %	5 %	6 %	7 %
<i>-oire</i>	0 %	0 %	0 %	0 %	0 %
<i>-euse</i>	1 %	1 %	0 %	2 %	1 %
<i>-eur</i>	12 %	11 %	11 %	9 %	16 %
<i>-et</i>	10 %	16 %	14 %	12 %	8 %
<i>-ette</i>	29 %	24 %	23 %	16 %	14 %
<i>-ment</i>	4 %	6 %	6 %	4 %	6 %
<i>-ion</i>	15 %	9 %	8 %	4 %	15 %
<i>-rice</i>	0%	0%	0%	0%	1%
Total de mots suffixés correspondants	810	1178	643	333	352

Tableau 2. Proportions de mots suffixés observés pour les 13 suffixes étudiés dans les productions des parents, en fonction de l'âge des enfants (1 à 5 ans).

La proportion des différents suffixes est relativement stable, avec toutefois une plus grande proportion de mots construits avec *-ier* quand les enfants sont âgés de quatre ans. Dans une perspective d'analyse future, les spécificités dyadiques ainsi que le rôle des activités réalisées et du lexique qu'elles mobilisent pourraient être interrogés.

Dans l'ensemble des productions des enfants de l'étude, nous avons relevé 1 501 mots suffixés avec l'un des 13 suffixes. Parmi ceux-ci, les premiers observés sont *-ier* (dans *escalier*), *-ette* (dans *assiette*) et *-eur* (dans *tracteur*) dans la tranche d'âge 1;6-1;8. On observe très vite ensuite un plus grand nombre et une plus grande diversité de suffixes, et aux précédents s'ajoutent rapidement *-age*, *-et*, *-ment* et *-ion*. Parmi les 13 suffixes et quel que soit l'âge des enfants, les plus représentés sont, comme chez les parents, *-ier* (26 % des mots suffixés observés ; $n = 385$) et *-ette* (26 % des mots suffixés observés ; $n = 392$). Le suffixe *-oire* n'est pas observé dans les productions des enfants.

Le Tableau 3 donne des indications plus précises quant aux proportions de mots suffixés pour chacun des suffixes observés par rapport à l'ensemble des mots suffixés relevés dans les productions des enfants âgés de 1 à 5 ans. La proportion du suffixe *-ette*, très fortement représenté chez les enfants d'un an, diminue progressivement dans les corpus, tandis que la proportion de suffixes déjà très présents (comme *-ier*) ou émergents (comme *-et* ou *-age*) augmente.

Suffixes	1 an	2 ans	3 ans	4 ans	5 ans
<i>-iere</i>	0 %	3 %	3 %	2 %	0 %

<i>-ier</i>	24 %	20 %	21 %	30 %	37 %
<i>-ure</i>	2 %	6 %	6 %	13 %	8 %
<i>-iste</i>	0 %	0 %	0 %	0 %	2 %
<i>-age</i>	2 %	9 %	5 %	6 %	5 %
<i>-oire</i>	0 %	0 %	0 %	0 %	0 %
<i>-euse</i>	0 %	0 %	0 %	1 %	1 %
<i>-eur</i>	13 %	11 %	10 %	7 %	12 %
<i>-et</i>	3 %	13 %	19 %	18 %	7 %
<i>-ette</i>	47 %	29 %	31 %	19 %	20 %
<i>-ment</i>	5 %	2 %	4 %	3 %	4 %
<i>-ion</i>	5 %	7 %	1 %	0 %	4 %
<i>-rice</i>	0%	0%	0%	0%	0%
Total de mots suffixés correspondants	62	545	349	284	178

Tableau 3. Proportions de mots suffixés observés pour les 13 suffixes étudiés dans les productions des enfants, en fonction de l'âge des enfants (1 à 5 ans).

Plusieurs facteurs peuvent être interrogés pour comprendre l'usage par les enfants de tel ou tel suffixe, parmi lesquels des facteurs sémantiques ou formels. On peut également interroger le lien entre les productions des enfants et celles de leur(s) parent(s). À cette fin, pour chacun des enfants des corpus étudiés ($N = 9$), nous avons sélectionné le fichier contenant le plus d'énoncés ($m = 786$ énoncés ; $ET = 272$), afin de disposer des données les plus sensibles possible. Une corrélation de Spearman entre le nombre de mots suffixés produits par ces enfants (âge moyen = 3;7 ; $ET = 1;4$) et le nombre de mots construits avec ce même suffixe par leur(s) parent(s) est forte et significative pour le suffixe *-eur* ($\rho = .769$; $p = .016$). Une corrélation moyenne à forte existe, à titre de tendance, pour le suffixe *-ion* ($\rho = .65$; $p = .056$). La corrélation n'est pas significative pour *-ière*, *-ier*, *-ure*, *-iste*, *-age*, *-et*, *-ette*, *-ment*. Les analyses corrélationnelles ne sont pas réalisables pour *-oire*, *-euse*, *-rice* (un des partenaires ou les deux ne produisaient pas ces suffixes dans les 9 fichiers analysés). Une analyse incluant davantage de données et des outils statistiques adaptés comme des modèles mixtes (Gries, 2015), une étude morphologique fine des productions des enfants et de celles des parents, ainsi qu'une analyse de la dynamique des interactions permettraient d'approfondir l'analyse de ces résultats.

3.2. Analyses qualitatives d'échanges autour de mots suffixés

Si la fréquence des unités, des constructions, peut avoir un impact sur le processus d'acquisition du langage, la manière dont elles sont inscrites dans le dialogue peut également favoriser leur appropriation par l'enfant. Dans les interactions quotidiennes, adultes et enfants co-construisent du sens dans le cadre d'activités diverses, à travers la production d'actions et

de constructions langagières qui à la fois s'appuient sur le déjà-connu et le déjà-partagé (le terrain commun ou « common ground » – Clark, 2015) et en même temps introduisent progressivement de nouveaux éléments d'information. La présentation par l'adulte d'unités nouvelles, en relation avec des unités déjà connues, dans des énoncés qui font sens pour l'enfant et pour la réalisation de l'activité, pourra contribuer à leur appropriation par celui-ci (Salazar Orvig et al., 2021).

Les adultes peuvent déployer différentes stratégies pour introduire des mots nouveaux et étayer leur compréhension par l'enfant. En ce qui concerne les mots construits, ils peuvent notamment avoir recours à des mots de la même famille morphologique. Dans l'extrait 1, l'orthophoniste et l'enfant racontent et réinventent le récit de *Boucle d'Or et les Trois Ours* à partir d'un album interactif où des figurines peuvent être mises en action. Dans cet extrait, l'enfant propose de cacher les personnages et associe le verbe « cacher » à l'action correspondante (« j(e) vais les cacher là »). L'orthophoniste, réagissant à cette proposition, va introduire « se cacher » et proposera immédiatement une autre unité de la même famille morphologique : le nom dérivé « cachette ». Poursuivant la construction du récit, l'orthophoniste et l'enfant vont tour à tour réutiliser le verbe « (se) cacher », dans des structures syntaxiques très proches, sans pour autant être tout à fait identiques (« elle se cache », « elle va se cacher », « qui est-ce qui se cache ») et l'orthophoniste présentera à nouveau le mot construit « cachette » à l'enfant.

*ENF: j(e) vais les **catcher** là xxx.

*ENF: Boucle + d'or e(lle) va ici.

*ORT: non elle arrive pas à se **catcher**.

*ORT: c'est la **cachette** des souris.

*ENF: non toi elle va là.

*ORT: elle se **cache** dans le + ...

*ORT: xxx tout + à + l'heure j(e) crois qu'elle va se **catcher** plutôt dans le lit.

*ORT: ce sera mieux.

*ENF: mais non elle va **catcher** dans l(e) frigo.

[...]

*ORT: ah et qui est + ce qui s(e) **cache** sous l(e) tapis ?

*ORT: ah c'est toujours nos amies les souris !

*ENF: alors lui il va **catcher** + ...

[...]

*ORT: voilà sous l(e) tapis.

*ORT: bon c'est une drôle de **cachette** ça sous l(e) tapis hein.

Extrait 1. Corpus APPEL IOE, Lilou, 4;05⁵

⁵ Les conventions de transcription sont adaptées du format CHAT de CLAN.

Dans ce type de séquence, l'enfant expérimente l'usage de mots construits et de mots de la même famille morphologique. Il peut découvrir leur forme, leur sens et leur relation grâce à leur association dans l'interaction verbale et dans l'action. Ces associations ne sont pas rares, en particulier dans les situations dites spontanées telles que le jeu symbolique. Elles peuvent également être combinées à des formes d'explicitation du discours. C'est le cas dans l'extrait 2 où la mère décrit la scène réalisée avec un personnage (« on la voit plus, elle est à l'école ») puis la met en mots en utilisant la forme « cachée ».

*MER : Léonie est à l'école.

*MER : regarde **on la voit plus** elle est à l'école fut@i **cachée**.

Extrait 2. Corpus EVALANG, Léonie, 6;08

Parfois, le langage peut constituer une activité et/ou un objet de discours en soi. C'est particulièrement le cas dans des interactions orthophoniste-enfant. Dans ce type de situation, des activités structurées, dirigées vers la compréhension ou la production de certaines unités ou constructions langagières peuvent être proposées à l'enfant, comme dans l'extrait 3 issu d'une activité de dénomination. Dans cet extrait, l'enfant découvre une carte représentant une calculatrice et doit nommer l'objet. Pour rectifier l'interprétation de l'enfant (« téléphone » au lieu de « calculatrice »), l'orthophoniste commence par lui présenter la fonction de l'objet (« c'est pour calculer »), introduisant ainsi le verbe « calculer », à partir duquel elle introduit ensuite le mot suffixé « calculatrice ».

*ENF: c'est a@f téléphone !

%pho: se a teefɔn

*OST : ah non.

*OST : c'est pour **calculer**.

*OST : c'est une **calculatrice**.

*OST : mais tu as raison ça ressemble.

Extrait 3. Corpus APPEL IOE, Maéva, 4;04

L'enfant peut ainsi accéder au sens du mot construit, à partir d'un mot de la même famille morphologique qu'elle connaît probablement déjà.

4. Discussion

Cet article avait pour objectif de mettre en avant le potentiel de la base de données morphologiques Démonette-2 pour l'exploration de corpus oraux d'interactions adulte(s)-enfant(s). Les fonctionnalités d'export de résultats dans la section « Outils spécifiques » de la base, qui rendaient possible la création de listes de lexèmes sous forme de fichiers textes

génériques (.txt), permettent également à présent l'export dans un format spécifique, le format .cut, utilisable par le programme d'analyse de corpus linguistiques CLAN.

La méthodologie présentée permet l'analyse à la fois quantitative et qualitative des corpus. Ici, les usages de 13 suffixes ont été explorés dans les productions des enfants et de leurs parents. D'autres pourraient être ajoutés. Toutes les familles de dérivation suffixale ne sont pas encore présentes dans la base, en particulier des familles contenant des mots construits relativement fréquents dans les corpus utilisés (dérivations en *-oir*, par exemple, comme *bavoir*). Certaines familles morphologiques étant en cours d'introduction dans la base, ces besoins pourront progressivement être comblés.

Si nous avons mis en évidence l'utilisation de fichiers .cut avec l'outil CLAN, d'autres exploitations sont imaginables. Par exemple, il serait possible d'utiliser une requête CQL (Corpus Query Language) dans le logiciel TXM (Heiden et al., 2010) en listant dans cette requête, au moyen d'expressions régulières, l'ensemble des suffixes à rechercher. La mise en relation de ressources différentes soulève toutefois d'importantes questions au niveau théorique. Dans notre cas, nous avons notamment été confrontés à des questions de compatibilité entre la lemmatisation réalisée avec le programme CLAN et le traitement morphologique des lexèmes construits dans la base de données Démonette-2. Le lexème *cafetière* est par exemple étiqueté comme le féminin de *cafetier* dans CLAN (n|cafetier&f) alors qu'il est considéré comme un dérivé de *café* dans la base Démonette-2. De ce fait, *cafetière* n'est pas repéré en corpus comme une dérivation en *-ière* (contenue dans le fichier Xiere.cut) mais comme une dérivation en *-ier* (contenue dans le fichier Xier.cut).

L'exploration de corpus d'interactions adulte(s)-enfant(s) collectés dans le cadre de projets sur le développement du langage, avec un outil développé quant à lui dans le cadre d'une recherche en linguistique avec une approche relationnelle et paradigmatique de la morphologie dérivationnelle (Hathout & Namer, 2022), ne peut qu'inviter à poursuivre les échanges interdisciplinaires : autour de la notion de « mot construit » d'une part, de ce qu'elle recouvre au niveau psycholinguistique, notamment pour le jeune enfant ; autour des notions de familles et de réseaux morphologiques d'autre part, de leur rôle dans la structuration du lexique de la langue, des individus et en particulier des apprenants (Dal Maos et al., 2017), voire dans la structuration des interactions en contexte d'acquisition ou orientées vers l'apprentissage.

5. Conclusion

Dans cet article, nous avons mis en évidence un usage possible de la base de données Démonette-2 et de ses outils, à savoir l'exploration de corpus oraux dans le cadre d'interactions adulte(s)-enfant(s). Si l'objectif de cet article était principalement méthodologique et technique, il permet aussi :

- d'ouvrir des pistes de recherche en acquisition du langage, car le développement de la morphologie dérivationnelle dans les productions spontanées de jeunes enfants et dans les interactions adulte(s)-enfant(s) reste encore peu exploré ;
- d'apporter, dans le cadre de cette étude exploratoire, des éléments d'information sur l'usage de mots suffixés dans les productions d'enfants au développement typique ou atypique et dans celles de leurs interlocuteurs, informations qui peuvent être utiles pour des chercheurs en acquisition et pour des professionnels notamment des orthophonistes ;
- d'identifier des besoins au niveau technique et pour le développement de la base ;
- de créer un pont entre différents domaines des sciences du langage, en particulier celui de l'étude linguistique de la morphologie dérivationnelle et celui de l'acquisition du langage, dont les questionnements, les approches théoriques et les méthodes diffèrent, mais peuvent s'enrichir mutuellement.

Bien qu'appliqué aux interactions adulte(s)-enfant(s) et à des corpus transcrits avec CLAN, cet article ouvre également des pistes d'usage de la base Démonette-2 pour des chercheurs ou praticiens qui travaillent sur des corpus issus de contextes variés ou avec d'autres outils.

Bibliographie

- Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition*. Cambridge University Press.
- Berthiaume, R., Daigle, D., & Desrochers, A. (Eds.). (2018). *Morphological processing and literacy development : Current issues and research*. Routledge, Taylor & Francis Group.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Catalise-2 Consortium, (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58(10), 1068-1080. <https://doi.org/10.1111/jcpp.12721>
- Caët, S., da Silva-Genest, C., Macchi, L., Patin, C., Gibaru, I., & Dumoulin, I. (2021). Co-analyse des modes d'intervention orthophoniques : Mise en place d'une collaboration orthophonistes-linguistes. *TRANEL (Travaux neuchâtelois de linguistique)*, 74, 47-65. <https://doi.org/10.26034/tranel.2021.2912>
- Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of Instruction in Morphological Awareness on Literacy Achievement: An Integrative Review. *Reading Research Quarterly*, 45(4), 464-487. <https://doi.org/10.1598/RRQ.45.4.5>
- Casalis, S., & Louis-Alexandre, M.-F. (2000). Morphological analysis, phonological analysis and learning to read French: A longitudinal study. *Reading and Writing*, 12(3/4), 303-335. <https://doi.org/10.1023/A:1008177205648>
- Cattini, J., & Duboisdindien, G. (2021). La conscience morphologique comme moyen d'intervention : Illustration clinique. *Les Cahiers de l'ASELF*, 18(2), 27-34.

- Clark, E. V. (1993). *The Lexicon in Acquisition (1re éd.)*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511554377>
- Clark, E. V. (2015). Common Ground. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 328-353). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118346136.ch15>
- Colé, P., Royer, C., Leuwers, C., & Casalis, S. (2004). Les connaissances morphologiques dérivationnelles et l'apprentissage de la lecture chez l'apprenti-lecteur français du CP au CE2. *L'année psychologique*, 104(4), 701-750. <https://doi.org/10.3406/psy.2004.29686>
- Dal Maso S., Voga, M., & Giraud, H. (2017). Du traitement de la morphologie à l'apprentissage du lexique L2 : Innover dans une perspective psycholinguistique paradigmatique. *Méthodal*, 1, 85-99.
- da Silva-Genest, C., & Le Mené, M. (2018). Analyser un discours semi-spontané en vue d'évaluer les compétences morpho-syntaxiques et discursives d'enfants à développement typique et atypique : étude exploratoire autour des récits d'expériences personnelles. *Journée d'étude BDEVALANG « Evaluation des pratiques langagières : des contextes familiaux aux contextes cliniques »*, Université de Lorraine et MSH Lorraine.
- Duboisindien, G., & Dal, G. (2021). Critical analysis of clinical resources and tools for derivational morphology used in francophone speech and language therapy. *Third International Workshop on Resources and Tools for Derivational Morphology - DeriMo2021*, 52-60.
- Fejzo, A., Desrochers, A., & Deacon, S. H. (2018). The acquisition of derivational morphology in children. In R. Berthiaume, D. Daigle, & A. Desrochers (Eds.), *Morphological processing and literacy development: Current issues and research* (pp. 135-156). Routledge, Taylor&Francis Group.
- Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95-125. <https://doi.org/10.3366/cor.2015.0068>
- Hathout, N., & Namer, F. (2022). ParaDis: A family and paradigm model. *Morphology*, 32(2), 153-195. <https://doi.org/10.1007/s11525-021-09390-w>
- Heiden, S., Magué, J.-P., & Pincemin, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. *JADT 2010: 10th International Conference on the Statistical Analysis of Textual Data*, 1021-1032. <https://shs.hal.science/halshs-00549779>
- Liégeois, L., Chanier, T. & Chabanal, D. (2014). *Corpus globaux ALIPE : Interactions parents-enfant annotées pour l'étude de la liaison*. Nancy, Ortolang, <http://hdl.handle.net/11041/alipe-000853/>
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (Vol. 1-2). Lawrence Erlbaum.

- Masson, C. (2022). Parler à et avec les enfants en crèche : conduites langagières et dialogiques des adultes et des enfants en situation de lecture partagée et de jeu libre. In *SHS Web of Conferences* (Vol. 138). 8^e Congrès Mondial de Linguistique Française. <https://doi.org/10.1051/shsconf/202213809008>
- Morgenstern, A., & Parisse, C. (2012). The Paris Corpus. *Journal of French Language Studies*, 22, 7-12. [doi:10.1017/S095926951100055X](https://doi.org/10.1017/S095926951100055X)
- Rassel, A. (2020). *Contribution de la conscience morphologique à l'apprentissage de la lecture : Études en milieu défavorisé*. Thèse de doctorat non publiée. Université de Lille.
- Salazar Orvig, A. (2019). Approches théoriques actuelles de l'acquisition du langage. In S. Kern (Éd.), *Le développement du langage chez le jeune enfant* (pp. 13-51). De Boeck Supérieur. <https://hal.archives-ouvertes.fr/hal-03562359>
- Salazar Orvig, A., Weck, G. de, Hassan, R., & Rialland, A. (2021). *The Acquisition of Referring Expressions: A dialogical approach*. John Benjamins Publishing Company.