

Learning to shift: How and when L1 and early L2 kindergarteners acquire the variability between Colloquial and Standard French

Katerina Palasis

Université Côte d'Azur, CNRS, BCL, France

Mar Cordero-Rull

Université Côte d'Azur, CoBTeK, France

Abstract: Much research has highlighted broad differences between Standard French (SF) and Colloquial French (CF). This contribution investigates three variable forms in adult L1, child L1 and early L2 (eL2) between 2;6 and 5;11, compares the distributions of the variants in each group, and examines the contexts for the emergence of the SF variants in the children. The results strongly suggest that L1 and eL2 kindergarteners predominantly utter the CF variants first: doubled rather than singleton noun phrases (NPs), elided rather than full *il* clitics in preconsonantal contexts, and simple rather than combined negation. SF emerges second, unevenly and non-linearly as from 2;9 in L1. The singleton NPs appear first, complex negation is often second, and full preconsonantal *il* clitics mainly come later. L1 children seem more spontaneous and advanced than eL2 children in integrating SF variants into their daily speech. Rote-learned nursery rhymes and songs may represent the crucial initial trigger to SF in L1 but not in eL2. Shared reading seems central in eL2 and takes over from non-spontaneous contexts in L1. L1 but not eL2 children utter SF variants in their personal narratives. In line with previous work on diglossia, we claim that CF and SF emerge as two distinct grammars, but the results point to broad individual differences regarding the mastery of SF. We hope that these findings will be useful for linguistic, educational and clinical purposes to give attention to the children who are less at ease with early oral SF before they start learning written SF in primary school.

Keywords: kindergarten; French; morpho-syntax; developmental sociolinguistics; diglossia

Résumé: De nombreuses recherches ont mis en avant des différences importantes entre le français standard (FS) et le français informel (FI). Cette contribution examine trois phénomènes (formes du sujet, élision du clitique *il* devant consonne et négation) chez des adultes L1, des enfants L1 et L2 précoce (L2P) entre 2;6 et 5;11. Les distributions des variantes FS et FI sont comparées pour chaque groupe et les contextes d'émergence des variantes FS sont étudiés chez les enfants. Les résultats suggèrent que les enfants de maternelle L1 et L2P produisent en premier et majoritairement les variantes FI : sujets nominaux (SN) doublés par un pronom avant SN non doublés, clitiques élidés devant consonne avant clitiques non élidés, et négation simple avant négation complexe. Les variantes FS émergent en second, de façon inégale et non-linéaire à partir de 2;9 en L1. Les SN non doublés apparaissent d'abord, souvent suivis par la négation complexe puis les clitiques non élidés. Les enfants L1 semblent intégrer les variantes FS dans leurs productions quotidiennes de façon plus spontanée et précoce que les enfants L2P. Les comptines et chansons représentent le premier contexte déclencheur de FS en L1 mais pas en L2P. Le contexte de lecture partagée est central en L2P et prend le relais des comptines et chansons en L1. Les enfants L1, mais pas L2P, produisent des variantes FS dans le contexte de narrations libres. Dans le cadre de l'hypothèse de la diglossie, nous proposons que FI et FS émergent en tant que deux grammaires distinctes, mais l'étude met en avant des différences individuelles importantes dans la maîtrise de FS. Nous espérons que ces résultats auront des implications en sociolinguistique développementale, enseignement et orthophonie, notamment pour

aider les enfants en difficulté avec la maîtrise du FS oral avant qu'ils ne soient confrontés au FS écrit à l'école élémentaire.

Mots clés: maternelle; français; morphosyntaxe; sociolinguistique développementale; diglossie

Corresponding author: Katerina Palasis, BCL, UMR CNRS 7320, Maison des Sciences de l'Homme et de la Société Sud-Est, Campus universitaire St Jean d'Angély, 5 rue du 22ème BCA, 06300 Nice, France. Email: Katerina.Palasis@univ-cotedazur.fr.

ORCID ID: Katerina Palasis <https://orcid.org/0000-0002-9115-654X>; Mar Cordero-Rull <https://orcid.org/0000-0003-1424-2847>

Citation: Palasis, K., & Cordero-Rull, M. (2026). Learning to shift: How and when L1 and early L2 kindergarteners acquire the variability between Colloquial and Standard French. *Language Development Research*, 6(1), 208-258. <https://doi.org/10.34842/ldr2026-966>

Introduction

Acquiring a language consists in developing multidimensional skills, which include the ability to adapt to variable linguistic forms, contexts and/or addressees (e.g., Clark, 2004; Taguchi, 2019). Among many other capacities, the ability to switch between formal and informal registers within a language attests of such skills, and one perfect language to study variation is contemporary Hexagonal French, which is the variety of French spoken in the European part of France (also known as the ‘Hexagon’). Indeed, much fieldwork in sociolinguistics has highlighted broad phonological and morpho-syntactic differences between formal Standard French (SF) and informal Colloquial French (CF). Table 1 illustrates three main discrepancies between the two varieties: optional doubling of the noun subject with a clitic pronoun (e.g., Lambrecht, 1984; Blanche-Benveniste, 1994), the variable suppression of [l] in the clitic *il* ‘he/it’ before a consonant (e.g., Ashby, 1984; Morin, 1979), and the variable omission of the preverbal negative marker *ne* (e.g., Ashby, 1976; Coveney, 2002).

Table 1. Phonological and morpho-syntactic variation between Colloquial and Standard French

	Colloquial French	Standard French	
Subjects	<i>le chien il aboie</i>	<i>le chien aboie</i>	‘the dog (he) barks’
Clitics	<i>i dort</i>	<i>il dort</i>	‘he sleeps’
Negation	<i>i dort pas</i>	<i>il ne dort pas</i>	‘he (neg) sleeps not’

Native (L1) adult speakers switch from one variety to the other according to socio-pragmatic factors (i.e., literacy, context and/or addressee, Armstrong & Smith, 2002; Blanche-Benveniste, 2003). Research in second language (L2) acquisition has shown that L2 adults also generally acquire the linguistic variants and the socio-pragmatic factors that condition the three alternations presented in Table 1 (e.g., Nagy et al., 2003; Howard et al., 2006; Donaldson, 2017, respectively). The questions we address in this study deal with kindergarten children and their acquisition of these three alternations in L1 and early L2 (eL2) development. L1 development refers to children who grow up with one single spoken language in their infancy and early childhood (i.e., from birth to about age 6), whereas eL2 refers to children who grow up monolingually but start to regularly hear a second language (L2) in late infancy and/or early childhood through daycare and/or kindergarten (De Houwer, 1990).

L1 children mostly hear Colloquial French daily, through child-directed speech (CDS), adult-directed speech (ADS), and speech by other children. There are exceptions though, at home with television and shared reading, and in school where teachers’ CDS can be more formal than their speech to peers (Buson et al., 2022). L1 kindergarten children are thus exposed to a mixed environment in terms of

sociolinguistic variants, both at home and in school. Research has shown that children, sometimes as early as age 2, show sensitivity to phonological and morpho-syntactic variation, although the results vary according to language, age, linguistic component, task, and other factors (e.g., Johnson & White, 2020; Samara et al., 2017). As far as French is concerned, whereas L1 children's sensitivity to phonological variables and early morpho-syntax is documented (e.g., Nardy et al., 2013; Parisse & Le Normand, 2000, respectively), L1 kindergartener's sensitivity to morpho-syntactic variation seems less studied. Early L2 kindergarten children are exposed to the same mixed environment as L1 children in school, whereas their input at home with television and shared reading inevitably varies according to the language(s) used in each household. Less is known about this configuration: the acquisition of variability in eL2 acquisition altogether is described as an 'underdeveloped' area of investigation (Taguchi, 2019), and studies on linguistic variation across registers in L2 acquisition in comparison with L1 speakers are described as still very limited and inconclusive (Gan, 2024).

Furthermore, the theoretical problematization of linguistic variation represents a long-standing and still lively debate between two main approaches in sociolinguistics. On the one hand, diglossia hypothesizes that the alternating forms belong to two distinct –nevertheless cognate– grammars. Ferguson (1959) initially described the relationship between the standard and colloquial varieties of Arabic, Modern Greek, Swiss German and Haitian Creole. Among others, this hypothesis was recently extended to Cypriot Greek, African American English and Hexagonal French (e.g., Grohmann et al., 2016; Washington & Seidenberg, 2022; Massot & Rowlett, 2013, respectively). Generalizing from his adult dataset, Massot (2010, p. 90) further suggested the very strict constraint that a speaker would never switch from one grammar to the other within an utterance, hence suggesting a very strong pattern of co-occurrence between the variants of each grammar. A sentence with a SF subject and a CF negation such as **Le chien aboie pas* 'the dog does not bark' is therefore ruled out within this approach. Although strong, this generalization presents the advantage of being testable against new datasets. On the other hand, variationism hypothesizes that the alternating forms belong to the same grammar and that variation depends on socio-stylistic factors (Labov, 1963). Among others, Coveney (2011) applied this framework to Hexagonal French and described the variability of subject types and negation in adult speech in terms of frequencies and co-occurrences between variants that both change from one speaker to another. In this approach, SF and CF variants can emerge within the same utterance according to a combination of linguistic, stylistic and pragmatic factors (Coveney, 2011, p. 75). The extent to which speakers select and cluster variants has also been addressed in this framework, and the covariation of variables is described as weak (Guy, 2013; Guy & Hinskens, 2016). In sum, diglossia posits that speakers manage two separate grammars with strong intra-grammar coherence, whereas variationism argues for one grammar with

multidimensional sociolectal cohesion.

As far as acquisition is concerned, Ferguson (1959) and Labov (1964) both addressed the question of when children acquire variation. In the diglossic approach, Ferguson (1959, p. 331) suggested that children acquire the child-directed, colloquial grammar first, at home and by implicit transmission, whereas the standard grammar is learnt later, mainly at school and with more explicit information. In this framework, the within-grammar consistency suggested for adults by Massot (2010) could be envisaged as a constraint that alleviates the learnability problem caused by the optionality of two or more variants in a language (e.g., Sorace, 2003). In the variationist approach, Labov (1964, p. 91-93) suggested that children are initially monostylistic and that they introduce socially conditioned, stylistic variation in their speech only when they reach early adolescence. Although very different, these foundational approaches thus both suggested a sequential development of variation. More recently, much research in the field of developmental sociolinguistics has provided evidence in favor of more parallel views for the acquisition of variation. Children produce alternating forms as early as age 3, differentiate between colloquial and standard variants as from age 5, and tend to produce standard variants more frequently as from age 6 (e.g., Chevrot, forthcoming; Ghimenton et al., 2021; Johnson & White, 2020). Age 6 also represents the moment when differences in sociolinguistic patterns start to emerge between boys and girls (review in Chevrot, forthcoming).

The current study aims to contribute to the growing field of developmental sociolinguistics by examining and comparing adult L1, kindergarten L1 and eL2 datasets between 2;6 and 5;11 in order to document the acquisitional shift between CF and SF in kindergarteners and the patterns of co-occurrence, if any, between the variants reported in Table 1. More specifically, the phonological and morpho-syntactic variation will be investigated in the three groups to shed light on the distribution of the CF and SF variants in each group and examine the contexts for the emergence of the SF variants in child speech. In this comparative study, the Method section begins by detailing the target structures introduced in Table 1 and then presents the adult (ADS, CDS) and child (L1, eL2) corpora under scrutiny. The Statistical Analyses section presents the Generalized Linear Mixed-effect Models (GLMMs) chosen in this contribution. The Results section then examines and compares the distribution of the CF and SF variants for each phenomenon and further investigates the contexts for the emergence of the SF variants in child L1 and eL2 speech. It then reports on the individual acquisition paths for the three SF variants and the distribution of the contexts that seem to trigger SF throughout the kindergarten years. The Results section ends with a study of the co-occurrence of the SF variants. The results lead us to discuss French diglossia and some implications for educational purposes.

Method

This section will first detail and illustrate the target structures introduced in Table 1 and then present the adult and child datasets of oral Hexagonal French collated for this investigation.

Target Structures

Subject Types

In all the corpora examined in this study, we distributed the verbal sentences according to three subject types (a, b, c), as illustrated in (1) with adult-directed speech (ADS): (a-a") personal, impersonal and deictic singleton subject clitics, (b-b') subject clitics with co-occurring noun phrases (NPs) and strong pronouns, and (c) singleton NPs.

(1) Subject types (Blanche-Benveniste, 1994, p. 82-94):

- a. Je suis bien contente.
I am well happy
'I am very happy.'
- a'. Il faut être raisonnable.
one must be reasonable
'One must be reasonable.'
- a". C' est fait exprès.
that is done intentionally
'That is done intentionally.'
- b. Le médecin il est bon.
the doctor he is good
'The doctor is good.'
- b'. Nous on apprend les médicaments.
We one learns the medicine
'We learn how to prepare medicine.'
- c. Mais le problème est le même.
but the problem is the same
'But the problem is the same.'

The crucial condition that distinguishes CF and SF is (1c), because the absence of a subject clitic uncontroversially belongs to SF only, whereas the use of clitics seems less conclusive. Nevertheless, the co-occurrences in (1b) have been extensively studied and debated over the last four decades, because the co-occurrence of a clitic pronoun and an NP questions the morpho-syntactic status of the clitic in CF versus SF. Indeed, whereas the clitic is uncontroversially analyzed as a syntactic subject in

SF (and the NP as a dislocation), there is a debate in CF as to whether the clitic is a *syntactic* constituent or a *morphological*, preverbal agreement marker (e.g., Auger, 1994; Cabredo Hofherr, 2004; Culbertson, 2010; De Cat, 2005; Legendre et al., 2010; Liang et al., 2024; Palasis, 2015). The latter configuration is labelled doubling and hence contrasts with dislocation in SF.

Clitic Elision

The elision of the clitic *il* ‘he/it’ before a consonant is another well-studied marker that distinguishes CF and SF (Ashby, 1984; Morin, 1979). The elision is attested as early as the 12th Century (Pope, 1934), and the reintroduction of the [l] before a consonant is attributed to normative pressure and schooling in the 19th Century (Chervel, 2011). A full *il* before a consonant, as illustrated with child-directed speech (CDS) in (2a), is therefore nowadays considered as a marker of SF in contrast with the elided form *i* in CF (2b). Consequently and extending the description of the contrast between dislocation and doubling reported in Subject Types, when a third-person singular clitic co-occurs with an NP, if the clitic is full in a preconsonantal context, as in (2d), the co-occurrence belongs to SF (and is dislocation), whereas an elided clitic, as in (2e), signals that the co-occurrence belongs to CF (in which case it is doubling).

(2) Distribution of *il* clitics:

- a. *il/_C*: Il nage avec les petits poissons. (CDS2, Palasis, 2012a)
he swims with the small fish
‘He swims with the small fish.’
- b. *i/_C*: Oui i dort. (CDS2, Palasis, 2012a)
yes he sleeps
‘Yes he is asleep.’
- c. *il/_V*: Il est trop gros ! (CDS2, Palasis, 2012a)
he is too big
‘He is too big!’
- d. NP+*il*: Antonio, il me l’ apprend. (ADS, Ashby, 1988, p. 204)
Antonio he to.me it teaches
‘Antonio teaches it to me.’
- e. NP+*i*: Mat i te donne les jetons. (CDS2, Palasis, 2012a)
Mat he to.you gives the tokens
‘Mat gives you the tokens.’

Datasets unanimously show that prevocalic *il* clitics, as in (2c), never display [l] elision. This study will therefore consider preconsonantal contexts only and will focus on the third singular *il* clitics in these contexts.

Negation

Simple, post-verbal negation, as illustrated with ADS in (3a) with the negative adverb *pas* ‘not’, is also well documented as a marker of CF, described as gaining ground in oral French particularly since the 19th Century (Martineau & Mougeon, 2003). ADS data throughout the 20th Century and the beginning of the 21st Century illustrate the evolution by reporting (very) low rates of *ne* retention (e.g., Armstrong & Smith, 2002: 1.1%; Ashby, 1982: 16%; Coveney, 2002: 18.8%). However, the studies also show variability, and Ashby (1981) suggested that familiarity with the addressee may be a factor of variation. The simple, post-verbal negation is opposed to the combined, pre- and postverbal markers *ne... pas*, as in (3b), which are reported to signal SF (Ashby, 1976, 1981, 2001; Armstrong & Smith, 2002; Coveney, 2002; Hansen & Malderez, 2004).

(3) Simple and combined negation (Ashby, 1981, p. 647-648):

- a. C’ est pas mon ordinateur.
that is not my computer
‘That’s not my computer.’
- b. Parce que France Telecom ne s’ occupe pas [...].
Because France Telecom neg itself deals not
‘Because France Telecom does not deal with [...].’

This study will therefore examine the omission and the retention of the preverbal marker *ne* in negative verbal sentences.

Adult Data

The adult counts for the three linguistic phenomena reported in this study come from L1 adult-directed speech (ADS) and child-directed speech (CDS) corpora.

ADS

The ADS stems from sociolinguistic research on the three variables shown in Table 1. The occurrence counts are gathered in Table 2.

The dataset on subject types comes from ten minutes of an oral conversation between two friends analyzed in Blanche-Benveniste (1994). It amounts to 203 verbal sentences. In the original study, the sentences were distributed according to six different subject types, which still appear in (1) but were subsumed under three types only (a, b, and c) for the purpose of this study. The ratios calculated on clitic elision come from fieldwork from Ashby (1984). The author recorded 26 adults of different ages and socio-economic categories and analyzed all their clitic pronouns ($n = 4,033$). Their third singular *il* clitics only were retained for this study ($n = 2,629$). Finally, the

ADS ratios on negation stem from oral interviews of 103 adults of different ages and socio-economic categories, who produced 2,818 negative verbal utterances (Ashby, 1981).

Our ADS dataset for the three linguistic phenomena under investigation therefore amounts to 5,650 utterances, as reported in Table 2.

CDS

Research has also well documented differences between ADS and CDS (since Ferguson, 1964; Newport et al., 1977; Snow, 1972). Among many characteristics, CDS presents shorter utterances, more questions and requests, and fewer third-person pronouns than ADS (Veneziano, 2000). The CDS examined in this contribution stems from two datasets.

Our first dataset is the one reported in Culbertson (2010). The author examined the ‘Lyon Corpus’ (106,000 interactions between 4 L1 children aged 0;11-4;6 and their mothers; Demuth & Tremblay, 2008). We used the counts reported in Culbertson (2010) on subject types and negation, and we labelled this dataset CDS1 in our study. The data from one adult interacting with a kindergarten class with L1 and eL2 children were added to complement the first dataset, notably because CDS1 lacked data on clitic elision. The second CDS dataset is labelled CDS2 (27,077 utterances; Palasis, 2012a). The CDS subsets calculated for this contribution for all three linguistic phenomena are reported in Table 2. CDS data for subject types stem from Culbertson (2010; $n = 5,253$) and Palasis (2012a; $n = 17,983$). Elision data come from Palasis (2012a, $n = 1,888$). Negative sentences come from Culbertson (2010; $n = 5,990$) and Palasis (2012a; $n = 1,890$). Both datasets are available on the CHILDES platform (MacWhinney, 2000a).

Table 2. Adult datasets

	ADS	CDS1	CDS2	Total
Subject types	203	5,253	17,983	23,439
Clitic elision	2,629	n.a.	1,888	4,517
Negation	2,818	5,990	1,890	10,698
Total	5,650	11,243	21,761	38,654

In sum, the overall dataset for adult speech amounts to 38,654 utterances, as shown in Table 2. Following Parisse and Le Normand (2000), we assumed that the joint ADS and CDS datasets would give us reliable measures of the morpho-syntactic variability young children are exposed to daily and hence could serve as points of comparison for the study of the child data.

Child Data

The L1 and early L2 (eL2) child phonological and morpho-syntactic variation was examined in a three-year longitudinal kindergarten database (Y1: 2;5-4;0, $M = 3;3$; Y2: 3;6-4;11, $M = 4;3$; Y3: 4;5-5;11, $M = 5;2$) with 16 L1 and 2 eL2 children (see Appendix 1 for each child's age range and gender). The first two years of this naturalistic dataset are available on the CHILDES platform (MacWhinney, 2000a; Palasis, 2012a). Groups of two to four children were recorded while interacting with either the researcher (the CDS2 reported above), L1 or eL2 peers during 20 to 25-minute sessions. The 33 sessions throughout the three-year period (at an average interval of three to four weeks except for holidays) enabled the researcher to collect 65 hours of child-child and child-researcher interactions. The activities and speech were informal, and different contexts occurred: narratives based on personal events, game playing, narratives based on shared reading, nursery rhymes/songs, and actual reading. The 16 L1 children (9 female) had French-speaking environments at home, whereas the two eL2 children (2 female) had Russian and European Portuguese L1 environments, respectively. The dataset for this study was extracted using CLAN commands (MacWhinney, 2000b) and was then manually checked. The overall amount of child occurrences investigated in this contribution is 31,757 (29,603 for L1 and 2,154 for eL2). Tables 3 and 4 distribute the occurrences per group, kindergarten year, and linguistic phenomenon. All the utterances and morpho-syntactic codes per year and child are available on the Open Science Foundation website (OSF, <https://osf.io/29twx>).

Table 3. L1 kindergarten datasets

	Y1	Y2	Y3	Total
Age range	2;5-4;0	3;6-4;11	4;5-5;11	2;5-5;11
Subject types	7,621	8,948	7,455	24,024
Clitic elision	868	1,293	677	2,838
Negation	856	951	934	2,741
Total	9,345	11,192	9,066	29,603

Table 4. eL2 kindergarten datasets

	Y1	Y2	Y3	Total
Age range	2;7-3;7	3;7-4;6	4;7-5;6	2;7-5;6
Subject types	388	762	672	1,822
Clitic elision	26	78	59	163
Negation	30	66	73	169
Total	444	906	804	2,154

Subject Types

In order to examine the subject types in the child database, the verbal sentences were categorized along the same three conditions as for the adult data: (a) singleton subject clitics, (b) subject clitics with co-occurring NPs and strong pronouns, and (c) singleton NPs, as illustrated in (4) with child data. Null-subject sentences, which represent a well-documented developmental phenomenon (e.g., Palasis, 2012b), and forms that did not undoubtedly exhibit a clitic, such as [ja] (for *il y a* ‘there is’), were left aside to match the three adult categories, as in (1a-c).

A total of 24,024 occurrences met these criteria for the 16 L1 children and 1,822 occurrences for the two eL2 children, as detailed in Tables 3 and 4. The variation between CF and SF was examined by investigating the ratios and contexts of emergence of the singleton NPs, as in (4c), considered as the marker of SF (see Appendix 2 for the full list of 160 child occurrences of singleton NPs).

(4) Subject types:

- a. Là on dirait qu’ elle marche. (CAR, 5;2)
there one says.cond that she walks
‘There it looks like she is walking.’
- b. Papi et Mamie i(ls) sont venus à la maison. (ELE, 5;4)
Grandpa and Grandma they are come to the house
‘Grandpa and Grandma came to my place.’
- c. Là son copain à Juliette tape des pieds et des mains. (NIN, 5;4)
there her friend to Juliet taps the feet and the hands
‘There Juliet’s friend taps his feet and hands.’

Clitic Elision

Since the elided [l] consonants were transcribed within parentheses in the child database, the two variants of *il* ‘he/it’ (i.e., full and elided) were extracted using the Kwal +R Option in CLAN (MacWhinney, 2000a, 2000b), and the items were further examined according to their following phonological context (i.e., vocalic vs.

consonantal). Like adults, L1 children always produced full clitics before a vowel (5a), and one of the two eL2 children produced only one prevocalic elision in the entire dataset. This context was therefore excluded from our study. The two remaining contexts are exemplified in (5): (b) preconsonantal elided *i*, and (c) preconsonantal full *il*. The L1 and eL2 children uttered a total of 2,838 and 163 preconsonantal *i(l)* clitics, respectively (see Tables 3 and 4).

(5) Distribution of *il* clitics:

- a. *il/_V*: Il est à moi. (WIL, 2;9)
 he is to me
 'It is mine.'
- b. *i/_C*: L' escargot i dort. (LUS, 2;9)
 the snail he sleeps
 'The snail is asleep.'
- c. *il/_C*: Il cherche son trésor. (MAT, 4;11)
 he searches his treasure
 'He is looking for his treasure.'

Empirical studies on the variable phonological form of *i(l)* 'he/it' in adult speech all show that *il* is overwhelmingly elided in preconsonantal contexts in CF, and that this variability does not apply to SF (see Table 1). Clitic elision is thus conditioned not only by phonological but also by socio-pragmatic factors. Thus, the crucial condition that illustrates the presence of SF is (5c) in contrast to (5b), regardless of whether the clitic is alone or in co-occurrence with an NP. The ratios of elision and non-elision of [l] were compared in ADS, CDS, L1 and eL2 speech and the contexts of non-elision in preconsonantal contexts in child data were examined to ascertain whether L1 and eL2 children are sensitive to the same socio-pragmatic cues as adults. One of these factors could be the presence of the preverbal negative marker *ne*, which displays two apparently conflicting characteristics. On the one hand, its presence after an *il* clitic sets the consonantal context that is expected to trigger the elision of the clitic (i.e., *i ne*). On the other hand, its presence is strongly associated with SF, in which the elision is not expected (i.e., *il ne*, as shown in Table 1). The behavior of *il* in the *ne* context will therefore be examined specifically (see Appendix 3 for the full list of 59 occurrences of full *il* clitics in preconsonantal contexts in the child data).

Negation

Negative verbal utterances were extracted by searching all the negative adverbs and pronouns in the database (i.e., *pas* 'not', *plus* 'not anymore', *jamais* 'never', *que* 'only', and *rien* 'nothing'). The variation between CF and SF was examined by comparing the ratios of these negative adverbs/pronouns uttered alone and those complemented with the preverbal negative marker *ne*. Condition (6a) represents CF, whereas (6b)

reveals the presence of SF. Tables 3 and 4 show that the L1 and eL2 children uttered a total of 2,741 and 169 negative sentences, respectively (see Appendix 4 for the full list of 47 child occurrences of combined negation).

(6) Simple and combined negation:

- a. J' ai pas fait l(e) papillon. (NIN, 5;2)
I have not done the butterfly
'I have not drawn the butterfly.'
- b. Ah non je n' ai pas. (CAR, 5;8)
ah no I neg have not
'Ah no I don't have (it).'

This contribution will investigate the distribution of the variants for the three linguistic phenomena presented in Table 1, in ADS, CDS, L1 and eL2 child data. The results will then be compared to establish if child L1 data are different from adult L1 data, and if L1 and eL2 children differ. Finally, we will examine the contexts of emergence of the variants pertaining to SF in child speech to determine if and how variation is conditioned in L1 and eL2 child data.

Statistical Analyses

Data from eL2 were excluded from subsequent statistical models due to the small sample size and scarce data points. Data from CDS2 were also excluded as they originated from a single speaker, which prevents the estimation of between-speaker variability. To enable developmental comparisons across the three years, data from ADS and CDS1 corpora were also excluded from the models. The statistical analyses presented in the Results section thus focused exclusively on the production counts of the 16 L1 children.

The goal of the analyses presented in the Results section is to evaluate differences in the use of the three linguistic phenomena and their developmental trajectories. For that, Generalized Linear Mixed-effect Models (GLMMs) were chosen due to their robustness against violations of homogeneity of variance and sphericity (Quené & Van den Bergh, 2004). While Poisson regression models are typically appropriate for count data (Bentz & Winter, 2013; Winter, 2019), the data did not meet the assumptions of Poisson regression, as the sample mean and the sample variance were not identical, and data inspections revealed severe overdispersion (i.e., the sample variance exceeds the mean) in all three phenomena. To account for this overdispersion, we fitted Negative Binomial GLMMs, which incorporate an additional parameter to estimate dispersion (Bentz & Winter, 2013; Winter et al., 2018).

Model selection procedure was conducted wherein all relevant model configurations were fitted and compared for each linguistic phenomenon, separately. All models were fitted using the *glmmTMB* package (McGillcuddy et al., 2025) in R, which accommodates both overdispersed and zero-inflated distributions. Categorical predictors in all models were coded using treatment contrasts. The most parsimonious model was selected (Bates et al., 2015), evaluating fit through residual inspection, comparisons of AIC and BIC values using the *anova()* function, and the computation of adjusted R-squared using values via the *MuMIn* package (Bartoń, 2025).

Significant effects were computed through the *joint_test()* function from the *emmeans* R package (Lenth et al., 2023). Post hoc pairwise comparisons were conducted using the *glht()* function from the *multcomp* package (Bretz et al., 2011) with Tukey-adjusted p-values. All statistical analyses were performed in R (v. 4.2.3; R Core Team, 2023). Code scripts, model estimates of the selected models, and comprehensive pairwise comparison tables are provided in the OSF repository (<https://osf.io/29twx>).

Results

Subject Types

The adult and child L1 and eL2 distributions of subject types are reported in Tables 5 and 6, respectively.

Table 5. Subject types (adult and kindergarten L1)

% (n)	Adults			Children		
	ADS	CDS1	CDS2	Y1	Y2	Y3
Clitics alone	81.3 (165)	96.9 (5,092)	75.3 (13,536)	66.9 (5,095)	72.1 (6,452)	76.3 (5,691)
Clitics + NPs	12.8 (26)	2.6 (137)	24.4 (4,393)	32.7 (2,492)	27.4 (2,450)	22.7 (1,692)
NPs alone	5.9 (12)	0.5 (24)	0.3 (54)	0.4 (34)	0.5 (46)	1.0 (72)
Total	100 (203)	100 (5,253)	100 (17,983)	100 (7,621)	100 (8,948)	100 (7,455)
Clitics/All subjects	94.1 (191)	99.5 (5,229)	99.7 (17,929)	99.6 (7,587)	99.5 (8,902)	99.0 (7,383)
Doubled/All NPs	68.4 (26)	85.1 (137)	98.8 (4,393)	98.7 (2,492)	98.2 (2,450)	95.9 (1,692)

The first three lines display the three types of subjects under scrutiny (i.e., Clitics alone, Clitics + NPs, and NPs alone), as exemplified in (1a-c) and (4a-c). The Clitics/All

subjects ratios calculate the frequencies of subject clitics (Clitics alone and Clitics + NPs) against the total number of subjects. The Doubled/All NPs ratios calculate the frequencies of doubled NPs against the total number of NPs.

Table 6. Subject types (kindergarten eL2)

% (n)	ELA (L1: Russian)			RIA (L1: Portuguese)		
	Y1	Y2	Y3	Y1	Y2	Y3
Clitics alone	65.6 (147)	73.3 (349)	82.2 (295)	68.9 (113)	64.4 (186)	63.6 (199)
Clitics + NPs	34.4 (77)	26.7 (127)	16.7 (60)	31.1 (51)	34.6 (100)	35.1 (110)
NPs alone	0.0 (0)	0.0 (0)	1.1 (4)	0.0 (0)	0.0 (0)	1.3 (4)
Total	100 (224)	100 (476)	100 (359)	100 (164)	100 (286)	100 (313)
Clitics/All subjects	100.0 (224)	100.0 (476)	98.9 (355)	100.0 (164)	100.0 (286)	98.2 (309)
Doubled/All NPs	100.0 (77)	100.0 (127)	93.8 (60)	100.0 (51)	100.0 (100)	96.5 (110)

Group Results

Tables 5 and 6 show that the three subject types rank similarly across the four sets of oral data (i.e., ADS, CDS, child L1 and eL2). Clitics alone always represent the most frequent type of subject (ranging from 63.6% in eL2 to 96.9% in CDS1), singleton NPs are always the rarest type (from 0% in eL2 to 5.9% in ADS), and Clitics + NPs present intermediary ratios (from 2.6% in CDS1 to 35.1% in eL2). Second, the Clitics/All subjects ratios, which add Clitics alone and Clitics + NPs, highlight that clitics are pervasive in the four corpora (from 94.1% in ADS to 100.0% in eL2). Finally, the Doubled/All NPs ratios, which calculate the ratios of doubled NPs against the total number of NPs, represent an additional way of measuring how NPs appear in oral French. The ratios, which range from 68.4% in ADS to 100% in eL2, confirm that NPs have a strong tendency to co-occur with a clitic (as in 1b and 4b) rather than alone (as in 1c and 4c) in oral data.

To evaluate differences in the use of subject types in the 16 L1 children across the three years, production counts were fitted into the models as the dependent variable. The data exhibited substantial overdispersion (dispersion parameter = 17.21; Pearson $\chi^2(1569) = 26998.09$, $p < .001$). For the selected model, Subject Type (three levels: Clitics alone [reference], Clitics + NPs, NPs alone) and Year (three levels: Year 1 [reference], Year 2, Year 3) were submitted to the model as fixed factors, including

their interaction. Model inspection indicated that including participants with random intercepts only provided the most parsimonious model.

Post hoc comparisons indicated that Clitics alone were produced significantly more often than both Clitics + NPs and NPs alone across all three years ($ps < .03$). L1 children used significantly more often Clitics + NPs than NPs alone across the three years ($ps < .001$). Results revealed that L1 children significantly increased their use of Clitics alone from Year 1 to both Year 2 ($\beta = -0.709$, $SE = 0.21$, $z\text{-ratio} = -3.339$, $p = .024$) and Year 3 ($\beta = -0.736$, $SE = 0.22$, $z\text{-ratio} = -3.382$, $p = .021$). However, no significant differences were found between Year 2 and 3 ($p = 1$). Children's use of Clitics + NPs across the three years revealed no significant differences ($ps > .6$). In the case of NPs alone, L1 children used them more often in Year 3 than in Year 1 ($\beta = -1.296$, $SE = 0.31$, $z\text{-ratio} = -4.186$, $p = .001$), but no significant differences were found between Year 1 and 2 or Year 2 and 3 ($ps > .2$).

Individual Results

The child datasets further showed that, although rare (from mean 0.4% to 1% in L1 and from 0% to 1.3% in eL2), the singleton NPs were produced by all the children at some point during the kindergarten years and that the category slightly increased between Year 1 and Year 3. However, the increase was far from large, and the overall L1 ratios masked very different patterns of use. First, the L1 children presented individual differences (the individual tables are provided at <https://osf.io/29twx>). Indeed, the individual child ratios varied from 0% to 1.2% in Year 1 (with one child producing 55.9% of all singleton NPs), from 0% to 3.5% in Year 2, and from 0% to 2.7% in Year 3. Nine children produced no singleton NPs in Year 1, and this number decreased drastically to one child in Year 2 and one different child in Year 3. Furthermore, the Doubled-NP category always emerged before the NP-alone category in the production of the L1 children (data was inconclusive for one child only, who produced both categories at 2;11 during the first recording session). Overall, the individual ratios of singleton NPs thus increased throughout the period, showing that this type of subject, that signals SF, developed very gradually, unevenly, and non-linearly between 2;6 and 5;11 in L1 children. Moreover, in the individual development of each child, the NP-alone category always emerged after the CF variant.

Second, the two eL2 children showed a different developmental pattern from the L1 group. Indeed, the two children spent their first two years with no singleton NPs at all, like nine L1 children in Year 1 but like one L1 child only in Year 2. Then, their third-year development also diverged from the L1 children with higher ratios in Year 3 (i.e., ELA: 1.1% and RIA: 1.3%, higher than 11 and 12 out of 16 L1 children in Year 3, respectively, whose mean ratio was 1% in Year 3). Overall, the eL2 children produced singleton NPs later than their L1 peers, but once the category had entered their

system, it was used more frequently.

In sum, the NP-alone category, which signals the presence of SF in the dataset, remained rare in kindergarten speech despite a slight increase throughout the three years. These output ratios in L1 children were nevertheless in line with the ratios in their input, particularly in CDS, which were also very low (mean 0.4% from CDS1 and CDS2 in Table 5). Our new datasets (CDS2, L1 and eL2 children) thus comply with previous studies that had shown the predominance of clitics and the scarcity of singleton NPs in CF compared to SF (e.g., Culbertson, 2010).

As noted in the Introduction, research has shown that L1 adult speakers switch from CF to SF according to socio-pragmatic factors, notably literacy, context and/or addressee, and that L2 adult speakers also manage this point of morpho-syntactic variation, even when it is not explicitly taught in the classroom. The open question then deals with the ability of kindergarteners, who also develop their early morpho-syntactic skills from implicit exposure, to connect the different morpho-syntactic variants with the relevant contexts and/or addressees. The Contexts section will aim to fill this gap by examining the contexts in which the L1 and eL2 children uttered their singleton NPs, in order to uncover what triggered their emergence throughout the three kindergarten years.

Contexts

The 18 kindergarten children produced a total of 160 singleton NP subjects (L1: $n = 152$; eL2: $n = 8$; see Tables 5 and 6, and full list of occurrences in Appendix 2). The kindergarten dataset comprised all the utterances produced by the 16 L1 and the two eL2 children during the recording sessions. The production types, therefore, were varied and ranged from fully spontaneous child-child speech to rote-learned lyrics and emergent genuine reading. The emergence of the NP-alone category, which is assumed to belong to SF, was therefore investigated in the dataset as a function of these varied production types.

First, the utterances were split into two types according to 'spontaneity'. An utterance was considered spontaneous, except when it was an immediate repetition, a nursery rhyme/song, formulaic speech or actual reading (as in 7a). This first distribution showed that 18.1% of the NP-alone subjects ($n = 29/160$) belonged to non-spontaneous speech, and that the L1 and eL2 rates for this category were different (17.8% and 25%, respectively). Second, the contexts of the spontaneous utterances were examined ($n = 125$ and $n = 6$, in L1 and eL2 respectively), and a second distribution was established according to the activity of 'shared reading', whether the utterance was a narrative based on a book (as in 7b) or a narrative based on personal events (as in 7e). The former accounted for 48% of the L1 utterances ($n = 60/125$) and 100% of the eL2

utterances ($n = 6/6$). These results show that the NP-alone subject type may have a connection with shared reading, which is one of the contexts where a child is exposed to SF when an adult reads a story. This connection is large for L1 children (48% of spontaneous utterances, $n = 60/125$) and categorical for eL2 children (100%, $n = 6/6$) in our dataset. These results thus also show differences between L1 and eL2 children. The eL2 children seem to utter NP-alone subjects in very limited contexts, totally tied to either the shared reading context only (ELA: $n = 4$) or the shared reading context and non-spontaneous utterances (RIA: $n = 2$ and 2 , respectively). The L1 NP-alone subjects distributed over these two contexts (as illustrated in 7a and 7b), but were also found in free, personal narratives (52% of the spontaneous utterances, $n = 65/125$, as in 7e), which was not the case for the eL2 children.

(7) L1 contexts:

- a. Nursery rhymes/songs:
Et ma petite cuillère marche marche. (LUS, 3;4)
and my little spoon walks walks
'And my little spoon walks walks.'
- b. Shared reading:
Juliette est en vacances. (ELE, 4;9)
Juliette is in holidays
'Juliette is on holiday.'
- c. Preverbal negation:
Mais lui n'est pas d'accord. (KEL, 4;8)
but him neg is not in.agreement
'But he doesn't agree.'
- d. Indefinite/quantified NPs:
Et après tout le monde dort. (CAR, 5;2)
and afterwards all the people sleep
'And afterwards everybody sleeps.'
- e. Other personal narratives:
Jean avait trop peur. (WIL, 4;9)
Jean had too fear
'Jean was too scared.'

These personal narratives were further examined to uncover other possible triggers to the NP-alone subject type in L1. Two characteristics that are documented as co-occurring with NP-alone subjects rather than subject doubling in adult French were noticed: preverbal negation (as in 7c) and indefinite/quantified NPs (as in 7d). These two contexts accounted for nearly half of these free narratives (4.6% plus 44.6%, respectively). The remaining 50.8% ($n = 33$) were further investigated, and individual variation appeared as an additional factor in the emergence of the NP-alone category. Three L1 children (ELE, CAR, LIN) had no NP-alone subjects in this category, and six

children (LOU, ENZ, LIZ, MAT, NIN, MAS) uttered only one utterance each. Free narratives with an NP-alone subject thus stemmed from seven children only, who produced 81.8% of this category, but which amounted to 3 to 5 utterances maximum per child throughout the three-year period in this dataset. Finally, it was hypothesized that the addressee of these free narratives might also have been a trigger to this SF variant. Indeed, these narratives were addressed to the researcher collecting the data and not to another child.

This section examined the ratios and contexts of emergence of the singleton NPs to document the emergence of SF in kindergarteners as well as their ability to connect the NP-alone subject type with relevant contexts and/or addressees. The results showed that this type of subject was rare but not randomly uttered by either L1 or eL2 children. To the contrary, singleton NPs seemed noticeably connected to shared reading, which represents one of the moments when children are exposed to SF (when an adult reads a story to the child), which in turn, thus also corresponds to one of the moments when children tend to produce SF. L1 and eL2 children did not behave strictly identically though. The data showed that singleton NPs emerged slightly later in eL2 acquisition compared to L1, and that the eL2 children seemed to utter singleton NPs in very limited contexts (non-spontaneous and shared reading) compared to some L1 kindergarteners who started to produce this SF type of subject in free narratives too. The next section will complement these findings by examining the form of the clitic *il* 'he/it' in the two other types of subjects, i.e. Clitics alone and Clitics + NPs.

Clitic Elision

Previous empirical studies on the variable phonological form of *il* 'he/it' in adult speech have shown that the clitic is overwhelmingly elided in preconsonantal contexts in CF, and that this categoricity does not apply to SF (see Table 1). As expected then, most of the L1 and eL2 occurrences alternated between full clitics before vowels (as in 5a) and elided clitics before consonants (as in 5b). The child dataset thus showed that the variation according to phonological context seems also very robust in the kindergarten stage. This study delved into a second, more fine-grained alternation. Tables 8 and 9 thus leave aside the non-variable, prevocalic *il* clitics and focus on the variability found in preconsonantal contexts between elision and non-elision of [l] (as in 5b and 5c, respectively), considered as a possible sign of alternation between CF and SF. Tables 7 and 8 detail the ratios of elision and non-elision of [l] in each adult group for which data was available (i.e., ADS and CDS2), in each kindergarten year for the L1 children as a group, and for the eL2 children individually.

Table 7. Forms of clitic *il* in preconsonantal contexts in L1

% (n)	Adults			Children		
	ADS	CDS1	CDS2	Y1	Y2	Y3
Elided <i>i</i>	79.6 (2,093)	n.a.	99.7 (1,882)	98.4 (855)	99.1 (1,281)	96.5 (653)
Full <i>il</i>	20.4 (536)		0.3 (4)	1.6 (13)	0.9 (12)	3.5 (24)

Table 8. Forms of clitic *il* in preconsonantal contexts in eL2

% (n)	ELA (L1: Russian)			RIA (L1: Portuguese)		
	Y1	Y2	Y3	Y1	Y2	Y3
Elided <i>i</i>	100 (13)	88.0 (44)	81.0 (17)	100 (13)	100 (28)	100 (38)
Full <i>il</i>	0 (0)	12.0 (6)	19.0 (4)	0 (0)	0 (0)	0 (0)

The datasets showed that elision is indeed the dominant option in preconsonantal contexts, whether in ADS (79.6%), CDS (99.7%) or child speech (81%-100%). Productions from L1 children were submitted to negative binomial models. The dispersion parameter was 4.56, significantly greater than 1 (Pearson $\chi^2(1044) = 4760.34$, $p < .001$), indicating overdispersion. The selected model included production counts of L1 children as the dependent variable. Elision Category (Elided *i* [reference], Full *il*) and Year were fitted as fixed factors with their interaction. Model inspection indicated that the Participants intercepts-only model provided the most parsimonious fit, while more complex random-slope models failed to improve model fit significantly. Detailed information about the selected model can be found on OSF (<https://osf.io/29twx>).

Post hoc comparisons revealed that L1 children uttered significantly more elided clitics than full clitics across the three years ($ps < .0001$). They appeared to significantly reduce their use of elided clitics in Year 3 (96.5%) compared to Year 2 (99.1%) ($\beta = 0.656$, $SE = 0.17$, z -ratio = 3.871, $p = .002$). In contrast, no significant differences were found for the use of full clitics across the three years ($ps > .2$).

Considering all four groups (i.e., ADS, CDS, L1, and eL2 children), we note that all the children except two (LIN, RIA) and all the adults uttered some full clitics in the preconsonantal contexts. First, the ratios in the two adult datasets were very different from each other (ADS: 20.4% vs. CDS2: 0.3%), much more divergent than the NP-alone ratios for the same groups (ADS: 5.9% vs. CDS2: 0.3%). The explanation is probably methodological insofar as the CDS2 data came from the same corpus for both phenomena, whereas the ADS for subjects and elision stemmed from two different corpora (Blanche-Benveniste, 1994 and Ashby, 1984, respectively). Second, in the children's datasets, the full clitics emerged in 15 L1 children out of 16 and in one of

the two eL2 children, and the timeline was the same in all children: the full clitics always emerged after the elided ones in preconsonantal contexts. Table 7 shows that the L1 ratio of full clitics doubled in Year 3 compared to Year 1 (3.5%-1.6%), and Table 8 shows that ELA displayed very high ratios in Years 2 and 3 (12%-19%) after having produced no full clitics at all in Year 1. The data thus show a clear developmental pattern in the emergence of the two variants, whether in L1 or eL2 data: the elided CF variant is present before the full SF variant, which is the same order as for the CF and SF variants of NPs (i.e., doubled first, alone second). However, the two acquisition paths do not seem correlated, because both LIN and RIA uttered no full clitics but produced a few singleton NPs (in very limited contexts though: non-spontaneous and book contexts).

Finally, the eL2 ratios were very different from each other and from the mean L1 ratios. Whereas RIA uttered no full *il* clitics in preconsonantal contexts, ELA showed very high ratios in Years 2 and 3 (12%-19%), much higher than any of the L1 children in Year 3, whose highest individual rate was 12.9% (KEL). We will elaborate on possible influences of their different L1s (ELA: Russian, RIA: European Portuguese) in the Learning to Shift section. Let us first examine the contexts in which these full *il* clitics emerged.

Contexts

The 18 kindergarten children produced a total of 59 full *il* clitics in preconsonantal contexts (L1: $n = 49$; eL2: $n = 10$; see full list in Appendix 3). The contexts of emergence for each of these 59 full *il* clitics were examined, using the same list of contexts as for the singleton NPs (i.e., spontaneity, shared reading, and preverbal negation).

The first distribution according to spontaneity shows that 18.6% of full clitics in preconsonantal contexts ($n = 11/59$) belonged to the non-spontaneous category, and that the L1 and eL2 rates in this category were very different, since the eL2 children uttered none in this context. Second, the contexts of the spontaneous utterances were examined in L1 and eL2 ($n = 38$ and $n = 10$, respectively), and a second distribution was established according to the activity of shared reading. This context accounted for 68.4% of the L1 occurrences ($n = 26/38$) and 90% of the eL2 occurrences ($n = 9/10$). These results show that the full version of the *il* clitic seems to be connected to this context, both for L1 and eL2 children (limited to ELA, since RIA uttered none). Moreover, the results show differences between L1 and eL2 children. ELA uttered the full clitics in very limited contexts, nearly exclusively tied to shared reading. Indeed, ELA uttered only one full clitic in a free narrative (representing 10% of her occurrences), whereas L1 children produced 31.6% of their full clitics ($n = 12$) in free narratives. Third, there were no full clitics in the context 'indefinite/quantified NPs', whether in L1 or eL2 children. Finally, preverbal negation appeared to be closely

related to shared reading in this part of the study since 83.3% of the full clitics with preverbal negation ($n = 15/18$, L1 children only) emerged in the shared reading context.

Comparing the results on clitic elision with subject types, we note that the distributions according to spontaneity and shared reading provide similar ratios in both linguistic phenomena. Indeed, 18.6% of full clitics in preconsonantal contexts and 18.1% of singleton NPs belonged to the non-spontaneous category, and shared reading accounted for 68.4% of the L1, 90% of the eL2 full clitics, and 48% of the L1 and 100% of the eL2 singleton NPs. Both distributions also show differences between L1 and eL2 children.

Negation

Table 9 displays the different forms of negation, with and without the preverbal negative marker *ne*, in the adult corpora. In his study of ADS, Ashby (1981) pointed to familiarity with the addressee as a function of variability between speakers, and the addressee was also found to influence the retention of *ne* in adult speech at school, depending on whether the teachers were addressing peers or children. Indeed, whereas teachers and teaching assistants showed low rates of *ne* retention in peer speech at school (6.1% and 1.5%, respectively), they exhibited higher rates of *ne* retention in CDS with a significant difference between teachers and teaching assistants (29.4% and 6.2%, respectively; Buson et al., 2022). Culbertson (2010) also found variability in the *ne* retention rate in the CDS1 dataset, but variation depended on the type of co-occurring subject. Thus, the 7.6% rate in Table 9 is an average of a particularly high retention rate with singleton NP subjects (83.3%), a high rate with imperatives and dropped impersonal subjects (14.1%), and much lower rates with singleton clitics and doubled NPs (6.3% and 6.6%, respectively). Finally, the CDS2 dataset shows that 96.4% of *ne* retention emerges in imperatives and orders given to children (e.g., *ne pousse pas* or *tu ne pousses pas* ‘(you) don’t push’). ADS and CDS, therefore, provide widely variable forms of simple and complex sentential negation to young children according to the speaker, the addressee and the type of subject.

Table 9. Forms of negation in adults

Corpora	Omission of <i>ne</i>	Retention of <i>ne</i>	Total <i>n</i>
ADS	63.0% (1,775)	37.0% (1,043)	2,818
CDS1	92.4% (5,534)	7.6% (456)	5,990
CDS2	92.6% (1,750)	7.4% (140)	1,890

The L1 and eL2 children produced a total of 2,910 negative utterances, mainly with the postverbal marker *pas* ‘not’ (88.8% and 91.1%, respectively). The other postverbal

markers were *plus* ‘not anymore’ (7.8% and 6.5%), *jamais* ‘never’ (1.8% and 0.6%), *que* ‘only’ (1.6% and 1.8%), and *rien* ‘nothing’ (0.1% and 0.0%). The timeline was the same for all children: simple postverbal negation always emerged before combined pre- and postverbal negation. Tables 10 and 11 report the ratios of *ne* omission and retention in L1 and eL2 for each kindergarten year (and see Appendix 4 for the full list of 47 utterances with *ne* retention). Tables 9 to 11 thus show that the dominant option in all oral corpora is single, postverbal negation, with differences though in *ne* retention rates between ADS (37%), CDS (mean 7.5%), L1 (1.2%-2.1%), and eL2 children (0%-3.4%).

Table 10. Forms of negation in L1 children

Years	Omission of <i>ne</i>	Retention of <i>ne</i>	Total <i>n</i>
1	98.2% (841)	1.8% (15)	856
2	97.9% (931)	2.1% (20)	951
3	98.8% (923)	1.2% (11)	934
Total <i>n</i>	2,695	46	2,741

Table 11. Forms of negation in eL2 children

Years	ELA		RIA		Total <i>n</i>
	Omission of <i>ne</i>	Retention of <i>ne</i>	Omission of <i>ne</i>	Retention of <i>ne</i>	
1	100% (18)	0% (0)	100% (12)	0% (0)	30
2	100% (43)	0% (0)	100% (23)	0% (0)	66
3	100% (44)	0% (0)	96.6% (28)	3.4% (1)	73
Total <i>n</i>	105	0	63	1	169

Differences in the use of preverbal negation of L1 children were assessed using negative binomial models. The dispersion parameter of the data was 5.78, significantly above 1 (Pearson $\chi^2(5271) = 30457.88$, $p < .001$). Production counts of L1 children were submitted as the dependent variable. For the selected model, Negation Category (absence of *ne* [reference], presence of *ne*) and Year were submitted as fixed factors including their interaction. Model inspection indicated that including random slopes for Negation Category by Participant provided the most parsimonious model. Detailed information about the selected model can be found on OSF (<https://osf.io/29twx>).

Post hoc comparisons revealed that the absence of preverbal negation occurred significantly more often than the presence of preverbal negation across the three years ($ps < .0001$). However, results revealed no significant differences in the production of utterances containing preverbal negation across the three years ($ps >$

.5). These results are very similar to those found in Buson et al. (2022), who also showed no increase in *ne* retention in kindergarten years.

We also converge on finding very important individual differences in kindergarteners (see individual counts at <https://osf.io/29twx>). In this study, four L1 children never uttered preverbal negation (ELE, LIN, WIL, MAS), whereas MAI uttered 23.9% ($n = 11$) of all the occurrences. In Year 1, four L1 children only produced scarce preverbal markers (VIC, LUS, NIN and MAI; total $n = 15$), and MAI produced 66.7% of the occurrences ($n = 10$). The number of L1 children rose to 10 in Year 2 with less inter-subject deviation (5-25% of the occurrences) but decreased to six in Year 3 (9.1%-27.3% of the occurrences). The two eL2 children had different patterns. ELA never produced *ne*, like four L1 children. RIA uttered only one *ne* in Year 3, which makes her rate of *ne* retention higher than 14 L1 children in Year 3.

Contexts

The contexts of emergence of the preverbal marker *ne* were examined for each of the 47 occurrences reported in Tables 10 and 11. First, 22 of these 47 utterances (46.8%) were already present in the list of utterances with either a singleton NP or a full *il* 'he/it' clitic, as illustrated in (8a) and (8b), respectively. The utterances that exhibit these concomitant variables mostly emerged in Years 1 and 2 (90.9%), and the context that may have triggered these combinations was overwhelmingly non-spontaneous in Year 1 (83.3%), as illustrated in (8a) and (8b) with nursery rhymes and songs.

(8) Contexts for *ne* retention:

- a. Singleton NP (in a nursery rhyme):
 Le facteur n' est pas passé. (VIC, 3;7)
 the postman neg is not been
 'The postman has not been.'
- b. Full *il* clitic (in a song):
 Il ne chante plus. (MAI, 3;2)
 he neg sings no.longer
 'He no longer sings.'
- c. Other clitics:
 Je n(e) sais pas. (ENZ, 4;2)
 I neg know not
 'I don't know.'
- d. Imperative verbal forms:
 Ne les touche pas Lou ! (LIZ, 5;5)
 neg them touch not Lou
 'Don't touch them Lou!'

The remaining 25 utterances with preverbal negation (53.2%) displayed either other clitics (e.g., *je* 'I', *on* 'one', *ce* 'this', $n = 17$, as in 8c) or imperative verbal forms ($n = 8$, as in 8d). The contexts for this subset of 25 utterances were mainly spontaneous: 60% ($n = 15$) were free narratives or game playing, and 28% ($n = 7$) were in a book context. Only 12% ($n = 3$) were non-spontaneous and corresponded to actual reading by the oldest child of the group in Year 3. The data thus seem to show a shift in the contexts that may have triggered the emergence of the preverbal marker *ne*, from non-spontaneous to more spontaneous activities. Closer examination of the latter, however, cast doubt on the full spontaneity of some of these utterances, and even in this category we suggest that the preverbal marker may have emerged because of the formulaic aspect of some of the sequences, notably with the verb *savoir* 'know' (as in 8c), which appeared in 7 L1 utterances, mainly in Year 2. The very specific contexts in which *ne* emerged is also illustrated in the single eL2 occurrence by RIA (4;7), who produced the complex negation while pretending she was reading (line 33 in 9), after an L1 child (MAS, 4;5) had produced an utterance with a singleton NP, also pretending he was reading (line 32 in 9). Line 29 in (9) further shows that RIA's preceding utterance did not display any obvious SF variants, contrary to line 33 after hearing MAS (line 32). We suggest a possible priming effect of MAS' SF syntax on RIA's utterance, because both utterances exhibit the same singleton NP subject. The hypothesis is in line with the previous observations that the two eL2 children are less spontaneous in uttering SF variants than the L1 children, and priming effects are well documented in language acquisition (see Kumarage et al., 2024 for a meta-analysis).

(9) *Ne* retention in eL2 (Y3, I3c, book context):

29 RIA (looking at a book):

Un jour Sanbarbe il est dans une pierre de...
 one day Sanbarbe he is in a rock of
 'One day Sanbarbe is (sitting) on a rock...'

30 MAS (looking at the same book):

Sanbarbe hi hi hi.
 Sanbarbe hi hi hi
 'Sanbarbe hi hi hi.'

31 RIA:

Dans une xxx.
 in a xxx
 'In a xxx.'

32 MAS (pretending he is reading):

Sanbarbe s'assoit pour jouer aux cartes.
 Sanbarbe himself sits to play at.the cards
 'Sanbarbe sits down to play cards.'

33 RIA (pretending she is reading):

Mais Sanbarbe ne veut pas que...

- but Sanbarbe neg wants not that
 ‘But Sanbarbe doesn’t want that...’
- 34 MAS (pretending he is reading):
 Sanbarbe est sur l’ eau tout mouillé.
 Sanbarbe is on the water all wet
 ‘Sanbarbe is in the water all wet.’

In sum, the preverbal marker *ne*, which signals the presence of SF, was scarce in kindergarten speech and lower than in CDS. It also presented wide interindividual variability and no increase in either L1 or eL2 throughout the kindergarten years. The emergence of *ne*, however, did not seem to be random, since it appeared to be limited to very specific activities, such as nursery rhymes, songs and reading, and very specific structures, such as imperatives and formulaic-like sequences.

Learning to Shift

Tables 12 and 13 recapitulate the results for the L1 and the eL2 children, respectively. They establish that the French-speaking kindergarteners in this study, whether L1 or eL2, predominantly uttered the three CF variants for the three phenomena under scrutiny: doubled rather than singleton NPs, elided rather than full *il* clitics in preconsonantal contexts, and simple rather than combined negation. The occurrences of the three SF variants were therefore much less frequent than the CF variants, and this holds for all three kindergarten years.

Table 12. Overview of L1 kindergarten French

% (n)	Subjects	Clitics	Negation	Total n
CF	99.4 (23,872)	98.3 (2,789)	98.3 (2,695)	29,356
SF	0.6 (152)	1.7 (49)	1.7 (46)	247
Total n	24,024	2,838	2,741	29,603

Table 13. Overview of eL2 kindergarten French

% (n)	Subjects	Clitics	Negation	Total n
CF	99.6 (1,814)	93.9 (153)	99.4 (168)	2,135
SF	0.4 (8)	6.1 (10)	0.6 (1)	19
Total n	1,822	163	169	2,154

Acquisition Path for SF Variants

Figure 1 collates the ages of emergence (in months) for the three SF variants in the 16 L1 and the two eL2 children (indicated with an *), hence providing an overview of the

development of the SF variants in these kindergarteners. The figure shows that the singleton NP category appears in all L1 and eL2 children and that it is usually the first SF variant to emerge in L1 (between 2;9 and 4;9). When it is not the case, its emergence in L1 is concomitant with one or the two other SF variants. Complex negation (12/16 L1 children) is often the second SF variant (between 2;10 and 5;5), and full preconsontantal *il* clitics (15/16 L1 children) mainly come later (between 3;0 and 5;6) but can emerge in children who do not yet utter preverbal negation (e.g., ELE). Figure 1 thus illustrates that the SF variants, which always develop after the CF counterparts in this dataset, emerge gradually, unevenly, and non-linearly between 2;9 and 5;6 in the 16 L1 children.

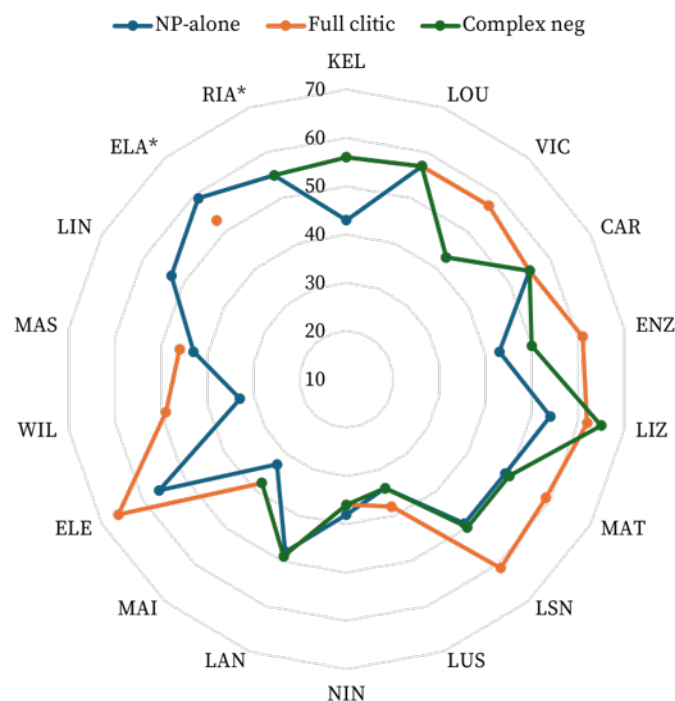


Figure 1. Observed emergence of the SF variants in all kindergarteners (ages in months).

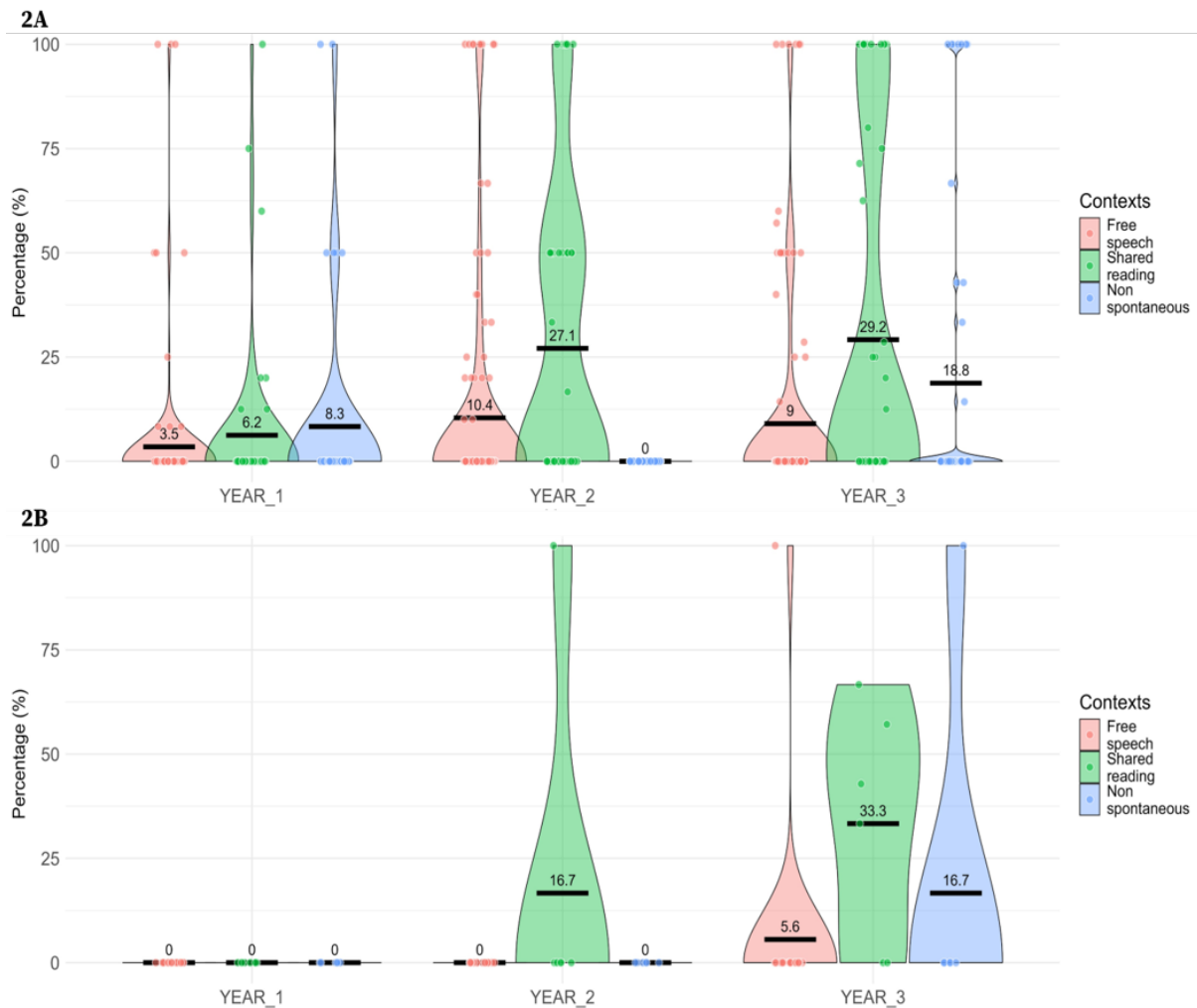
Furthermore, the two eL2 children differ not only from the L1 children but also from one another. ELA produces full clitics in the same age range as in L1 (at 4;5), but this variant is the first to emerge. Her singleton NPs appear later than in L1 (at 4;11) and she never produces complex negation. RIA, to the contrary, utters complex negation but no full clitics, and her two SF variants emerge in the same sentence at 4;7 (see Example 9, line 33). We suggest that their respective L1s could explain these divergences. In Russian, the nominative form of the masculine singular pronoun is

on ‘he’, pronouns are stressed and obligatory in most cases, and *on* is described as acquired around 2;9 in monolingual L1 Russian (Voeikova & Krasnoshchekova, 2020). The syllabic structure of *on* (identical to French vowel-consonant *il*) and the phonological status (non-clitic) of the former could then explain the emergence of full *il* in ELA’s utterances. To the contrary, pronouns in European Portuguese (EP) are clitics and optional in most cases, and their acquisition is described as starting to stabilize only around 4;0 in monolingual L1 EP, due to a complex placement pattern according to contexts, one of which being the preverbal negative marker *não* (Costa et al., 2015). Let us note, however, that despite these probable interferences with their L1s, both eL2 children seem to spot the variability between CF and SF in two of the three linguistic phenomena under investigation, just like three L1 children (ELE, WIL, MAS).

Distribution of SF Contexts

The emergence of the SF occurrences for the three linguistic phenomena (total $n = 266$) was examined according to context, by distributing the occurrences between non-spontaneous and spontaneous speech. The former covered immediate repetition, nursery rhymes/songs, formulaic speech (often obvious with specific intonation) and actual reading, whereas the latter contained shared reading (narratives according to books) and free speech (narratives according to personal events, playing games). Figures 2A and 2B display the distribution of these contexts throughout the three kindergarten years in L1 and eL2, respectively.

Figures 2A and 2B show two main divergences between the L1 and the eL2 children. First, the timeline of the developmental path of SF seems to be delayed in eL2 compared to L1 (Year 1: 62 utterances of SF in L1 vs. none in eL2). Second, the contexts that may trigger the emergence of the three SF variants are different between both groups. Non-spontaneous contexts (mainly represented by nursery rhymes and songs) appear as a crucial initial trigger to SF in L1 in Year 1 (50% of utterances, $n = 31/62$), whereas these contexts are hardly present in eL2 and appear in Year 3 only (15.4% of utterances, $n = 2/13$). Interestingly, there is a shift in L1 within the non-spontaneous category. Whereas non-spontaneous utterances are mainly nursery rhymes and songs in Year 1, they turn into actual reading in Year 3.



Figures 2A and 2B. Distribution of percentages of SF occurrences according to Group (L1 in 2A and eL2 in 2B), Year, and Context. Violin width reflects the density of observations; dots represent individual children; horizontal bars indicate mean values.

To assess how SF utterances were distributed across years and contexts, SF occurrence counts were submitted to negative binomial models as the dependent variable. The data exhibited overdispersion (dispersion parameter = 1.98; Pearson $\chi^2(714) = 1408.33, p < .001$). Context Category (free speech [reference], shared reading, non-spontaneous speech) and Year were submitted as fixed factors, with their interaction. Model inspection indicated that the Participants intercepts-only model provided the most parsimonious model. Detailed information about the selected model can be found on OSF (<https://osf.io/29twx>).

Post hoc comparisons revealed distinct patterns across years. In Year 1, significantly more utterances were produced in non-spontaneous contexts than in free speech ($\beta = -1.71$, $SE = 0.47$, $z\text{-ratio} = -3.639$, $p = .008$). No significant differences were observed between shared reading and non-spontaneous speech or between shared reading and free speech contexts ($ps > .7$). In Year 2, no significant differences were found among contexts ($ps > .1$). In contrast, in Year 3 significantly more utterances were produced in shared reading context than in free speech ($\beta = -1.365$, $SE = 0.4$, $z\text{-ratio} = -3.444$, $p = .02$).

Comparisons across years in L1 further revealed that more utterances were found in free speech contexts in Year 3 compared to Year 1, although the difference was only marginally significant ($\beta = -1.165$, $SE = 0.40$, $z\text{-ratio} = -2.916$, $p = .085$). Significantly more utterances were found in shared reading contexts in Year 3 than in Year 1 ($\beta = -1.707$, $SE = 0.54$, $z\text{-ratio} = -3.192$, $p = .038$). Taken together, these results support the evidence in our data for a shift in the use of SF utterances from non-spontaneous contexts in Year 1 to more spontaneous ones starting in Year 2.

As for eL2 in this dataset, Figure 2B shows that shared reading appears to be the crucial initial trigger to SF. It starts operating in Year 2 (100%, $n = 6/6$) and remains pervasive in Year 3 (76.9%, $n = 10/13$). Shared reading seems to represent a main trigger in L1 too, but contrary to eL2 it is present throughout the three kindergarten years (22.6%-48.7%-43.9%, respectively). Finally, free speech (narratives according to personal events and playing games) appears in all three years in L1 (27.4%-51.3%-37.4%) and is very scarce and delayed in eL2 (Year 3 only, 7.7%).

In sum: (i) in Year 1, rote-learned nursery rhymes and songs may represent the crucial initial trigger to SF in L1 but not in eL2, (ii) in Year 2, shared reading appears as pivotal in eL2 and seems to take over from non-spontaneous contexts in L1, (iii) throughout the period, L1 but not eL2 children (except for one utterance) utter SF variants in their free speech (but see the discussion on formulaic-like sequences in this category too, as illustrated in 8c). The results thus seem to show that L1 children are more spontaneous and advanced than eL2 children in integrating oral SF into their daily speech, which overwhelmingly remains CF in all children. Despite these differences, both groups appear to have spotted the differences between CF and SF and seem to switch to SF in very limited contexts, particularly in narratives according to books, which is precisely one of the contexts that exposes them to SF when an adult reads a story to them.

Co-occurrence of SF Variants

Co-occurrence of SF variants was examined in the sub-group of negative, non-imperative utterances with third-person masculine singular subjects ($n = 22/47$). Thus, the negative SF utterances excluded from the analysis were either imperative ($n = 8$)

or featured a different pronoun (e.g., *je* 'I'; $n = 17$). The results showed that 95.5% of these negative utterances ($n = 21/22$) exhibited a second SF variant: either a singleton NP ($n = 3$, as in 10a) or a full *il* clitic ($n = 17$, as in 10b). These full *il* clitics were mainly alone ($n = 13$, as in 10b) and sometimes doubled by either the strong pronoun *lui* 'him' ($n = 3$, as in 10c) or an NP (as in 10d). The remaining utterance ($n = 1/22$, in 10e) exhibited the strong pronoun *lui* alone, meaning that this utterance did not exhibit a subject type or a clitic form identified as a CF variant. Therefore, it seems that all the negative utterances that show the SF variant for negation in the child dataset, whether L1 or eL2, also show a subject that signals SF, as identified in the previous sections, namely either a singleton NP or a full *il* clitic. Example (10e) complements the list by showing that the NP can also be a strong pronoun, and Examples (10c-d) point to another additional possibility: full *il* clitics with a co-occurrent NP or a strong pronoun.

(10) Co-occurrent SF variants:

- a. Le facteur n' est pas passé. (VIC, 3;7)
the postman neg is not been
'The postman has not been.'
- b. Il ne sait pas nager. (CAR, 4;7)
he neg knows not swim
'He does not know how to swim.'
- c. Lui il n' est pas d'accord. (KEL, 4;8)
him he neg is not in.agreement
'HE disagrees.'
- d. Et le petit lapin youpi il ne pleure plus ! (NIN, 3;0)
and the little rabbit hooray he neg cries not.anymore
'Hooray, and the little rabbit isn't crying anymore!'
- e. Lui n' est pas d'accord. (KEL, 4;8)
him neg is not in.agreement
'HE disagrees.'

NIN's utterances further document these co-occurrences not only between SF variants but also between CF variants. Indeed, in the same narrative reported in (11), NIN (3;0) uttered 7 occurrences of the verb *pleurer* 'cry' with a third-person clitic, and the clitic was elided (lines 27, 44, 55, 65, 80), except when it emerged before the preverbal marker *ne* (line 69). Line 69 thus exhibits a combination of two SF characteristics (i.e., a full preconsonantal clitic and preverbal negation) among a series of utterances that displays a combination of two CF characteristics (i.e., an elided preconsonantal clitic and postverbal negation; lines 44, 55, and 80).

(11) Narrative in book context (NIN, Y1, VI26a):

- 27 le petit lapin i pleure.
the little rabbit he cries
'The little rabbit is crying.'
- 44 après le p(e)tit lapin i pleure plus avec ça.
after the little rabbit he cries not.anymore with that
'Afterwards the little rabbit doesn't cry anymore with that.'
- 55 i pleure pas.
he cries not
'He isn't crying.'
- 65 le petit lapin i pleure et le petit lapin i pleure encore.
the little rabbit he cries and the little rabbit he cries more
'The little rabbit is crying and the little rabbit is still crying.'
- 69 et le petit lapin youpi il ne pleure plus!
and the little rabbit hooray he neg cries not.anymore
'Hooray, and the little rabbit isn't crying anymore!'
- 80 parc(e) que i pleure plus le petit lapin xxx.
because he cries not.anymore the little rabbit xxx
'Because the little rabbit doesn't cry anymore xxx.'

A similar pattern seems to appear in LAN's (4;1) narrative reported in (12). Lines 210 and 212 exhibit postverbal negation and elided clitics, whereas line 225 shows concomitant preverbal negation and a full preconsonantal *il* clitic.

(12) Narrative in book context (LAN, Y2, VII26b):

- 210 pourquoi i va pas ?
why he goes not
'Why doesn't he go?'
- 212 pourquoi i va pas ici à ce chemin ?
why he goes not here at this path
'Why doesn't he choose this path?'
- 225 pourquoi il n' y va pas ?
why he neg there goes not
'Why doesn't he go there?'

Let us note that statistical investigations of adult corpora found similar associations between the retention of *ne* and the type of subject (e.g., Culbertson 2010; Liang et al., 2024, and see Example 3b). We will elaborate on these co-occurrences in the Discussion.

Discussion

The current study documents the acquisitional shift between CF and SF in kindergarteners by focusing on subject types, clitic elision and negation in 16 L1 and two eL2 children between 2;6 and 5;11. The results seem to confirm the sequential development of variation suggested in both diglossia and variationism (Ferguson, 1959; Labov, 1964, respectively) and complement the more recent developmental studies on the early acquisition of variation (e.g., Ghimenton et al., 2021). Thus, CF variants seem to always emerge before SF variants, and SF variants appear early, before 3;0 in some children (i.e., 2;9 for singleton NPs, 2;10 for complex negation, and 3;0 for full preconsonantal *il* clitics).

Diglossia or Variationism?

Our first main point of discussion deals with the theorization of linguistic variation. As noted in the Introduction, diglossia and variationism account for variation very differently, hypothesizing either two distinct grammars or only one, respectively. We address this matter by examining how our three SF variants do or do not co-occur in the kindergarteners' utterances.

The results pointed to a possible association between the emergence of the preverbal marker *ne* and the phonological form of the preceding *il* clitic, which is always full in this context. Syntacticians claim that the emergence of preverbal negation introduces a syntactic separation between the finite verb and the constituents to the left of the verb, mainly represented by subject clitics and NPs (e.g., Culbertson, 2010; De Cat, 2005). In line with this claim, we suggest that it is the emergence in syntax of the preverbal negative marker that triggers the full *il* clitic and not the contrary, because it has also been shown that *il* clitics elide before any other preverbal constituents with an initial [n] except *ne* (e.g., *i nous mangera pas* 'he will not eat us', Palasis, 2013). Other debates on the morpho-syntactic status of subject clitics (proper subjects or preverbal agreement markers) and the syntactic difference between dislocation and doubling in relation with the status of the subject clitics go beyond the scope of this contribution and are hence left aside here (but see Palasis, 2015 and Liang et al., 2024 on child and adult French, respectively). We will limit ourselves to the overarching question of how to formalize the acquisition of variation further to the covariation of SF variants observed in the child dataset.

In the diglossic approach, Massot (2010) suggested that variants strongly covariate within sentences, whereas Guy (2013) described covariation as weak in the variationist approach. Culbertson (2010) and Liang et al. (2024) both provided statistical support for the existence of two grammars in contemporary French and the latter further argued in favor of diglossia, entailing that French adult speakers

manage two grammars and switch from one to the other according to context and/or addressee. Regarding acquisition, we noted in the Introduction that diglossia makes a testable prediction about the absence of within-sentence switching between CF and SF. As long as we do not find any counterexamples to this strong constraint in child data, we suggest that within-grammar consistency may be extended to children. We further hypothesize that such consistency may serve as a frame that guides the children along the acquisition paths of CF and SF, hence alleviating the learnability problem caused by the optionality of two or more variants in a unique system.

Within the diglossic approach of French, we thus claim that L1 children first acquire CF (which is dominant in CDS) and subsequently develop SF, gradually, unevenly and non-linearly, as illustrated in this contribution. All French native speakers thus manage CF but may show discrepancies in SF, as early as the kindergarten stage and possibly all the way to adulthood. The two eL2 children in this dataset also showed early and initial acquisition of CF but seemed to exhibit a slight delay in the onset of SF, possibly due to no or fewer nursery rhymes and songs heard at home compared to L1 households.

Bridging the Gap between CF and SF

Tables 12 and 13 showed that the French-speaking kindergarteners in this study, whether L1 or eL2, did produce alternating forms for the three phenomena under scrutiny, and Figure 1 provided the detail of the acquisition paths per phenomenon and per child. Importantly, the results point to broad differences according to linguistic phenomenon, child, and child group. French-speaking children thus seem to leave kindergarten with very different levels of proficiency in *oral* SF, particularly regarding complex negation, whose child rates are much lower than the adult rates, even in CDS (L1-Y3: 1.2% vs. CDS2: 7.4%). These results thus seem to establish that French-speaking children do not all start primary school and the learning of *written* SF with the same mastery of its specificities in the oral modality.

A preliminary small review of subject types in five French grammar schoolbooks from first to third grade (between 6 and 8 years of age) shows that singleton NPs are pervasive, singleton clitics (always full) are present but rare, subject omission in coordination can be found (e.g., *ils vont et viennent* ‘they come and go’), and dislocation is totally absent (Palasis, 2026). The discovery of French grammar schoolbooks in primary school is then expected to be very different from one child to another, and more difficulties may be anticipated for children who do not (or very rarely) utter the SF variants by the end of kindergarten.

Setting these individual difficulties with SF aside, the L1 and eL2 kindergarteners altogether seemed to differentiate CF from SF. Figures 2A and 2B provided an

overview of the contexts associated with the SF occurrences according to child group and year and pointed to a possible distribution between spontaneous and non-spontaneous contexts in the emergence of SF variants. The investigation thus suggests that the children may have developed sufficient socio-pragmatic abilities to switch from one variety to the other according to context and/or addressee. These findings are in line with much previous work on the positive impact of book reading on language development (e.g., Le Normand, 2007; Sénéchal et al., 2008; Wasik & Bond, 2001) and additionally point to the early and crucial role of rote-learned nursery rhymes and songs in the early development of SF. Because their frequency inevitably differs from one family to another and between L1 and eL2 households, schooling may also play a central role in exposing all children to these contexts. We suggest that pedagogically oriented exposure to non-spontaneous contexts may help more kindergarteners familiarize themselves with SF variants before they reach primary school. The current study thus calls for further work on this matter in collaboration with teachers and in line with experimental work carried out on Standard Arabic, which is another language with hypothesized diglossia (Eghbaria-Ghanamah et al., 2020).

Finally, to the best of our knowledge, no explicit links are provided to children in the educational system in order to make connections between their early, implicit knowledge of CF and SF, and the new explicit knowledge of SF provided in the primary schoolbooks. Moreover, when these discrepancies between kindergarten and primary-school French are presented to schoolteachers, a demand for collaborative work between the developmental and educational fields clearly emerges.

Limitations of the Study

Naturalistic databases represent invaluable sources for research on child language development. The current dataset nevertheless also presents limitations, especially regarding occurrence counts. First, even though the L1 dataset amounts to nearly 30,000 utterances, the observation of very specific phenomena, such as covariation of variables, often rests on very few occurrences, which obviously raises sampling issues (e.g., Tomasello & Stahl, 2004). The current findings could therefore be complemented with experimental work. Furthermore, the contexts of the CF variants should also be examined and compared with the contexts of the SF variants analyzed in this contribution. Finally, the eL2 dataset is limited in terms of occurrences ($n = 2,154$) and participants ($n = 2$), who in addition have different L1s. The current study of eL2 French is therefore tentative but nevertheless documents two understudied areas of investigation: the acquisition of variability in eL2 and its comparison with child L1.

Conclusions

Much fieldwork in sociolinguistics has highlighted broad differences between Standard French (SF) and Colloquial French (CF). This contribution examined three variable linguistic forms: the types of subjects, the phonological forms of *il* 'he/it' clitics in preconsonantal contexts, and negation. The variants were investigated in adult L1 (adult- and child-directed speech), child L1 and eL2 datasets (16 L1 and two eL2 children between 2;6 and 5;11) to compare the distributions of the variants in each group and examine the contexts for the emergence of the SF variants in kindergarten speech.

This contribution strongly suggested that French-speaking kindergarteners, whether L1 or eL2, acquire CF first. The children predominantly uttered the three CF variants for the three phenomena under scrutiny: doubled rather than singleton NPs, elided rather than full *il* clitics in preconsonantal contexts, and simple rather than combined negation (see overviews in Tables 12 and 13). SF, which always developed after CF, emerged gradually, unevenly, and non-linearly as from 2;9 in the L1 children. The singleton NPs appeared first (between 2;9 and 4;9) and in all L1 and eL2 children. Complex negation (12/16 L1 children) was often the second SF variant (between 2;10 and 5;5), and full preconsonantal *il* clitics (15/16 L1 children) mainly came later (between 3;0 and 5;6) but also emerged in children who did not yet utter preverbal negation (see Figure 1). The two eL2 children also spotted the variability between CF and SF but with a slightly delayed timeline and divergences compared to L1, probably due to their respective L1s (i.e., Russian and European Portuguese). The current investigation thus suggested that L1 and eL2 kindergarten children seem to have developed sufficient socio-pragmatic abilities to switch from one variety to the other according to context and/or addressee. In line with previous work on diglossia and recent statistical investigations on adult French, we claimed that CF and SF emerge as two distinct grammars.

The study also pointed to the early and possibly crucial role of rote-learned nursery rhymes and songs and shared reading in the early development of oral SF and therefore suggested pedagogically oriented exposure to these contexts in kindergarten to ensure broader familiarization of all children before they reach primary school and written SF. We thus hope that our linguistic understanding of the acquisition paths of CF and SF in kindergarteners will also be useful for educational and clinical purposes to give attention to the children who are less at ease with early oral and later written SF. Hopefully, the current study will also contribute to the understanding of language development in other countries with similar linguistic situations.

References

- Armstrong, N. & Smith, A. (2002). The influence of linguistic and social factors on the recent decline of French *ne*. *Journal of French Language Studies*, 12(1), 23–41. <https://doi.org/10.1017/S0959269502000121>
- Ashby, W. J. (1976). The loss of the negative morpheme *ne* in Parisian French. *Lingua*, 39(1-2), 119–137. [https://doi.org/10.1016/0024-3841\(76\)90060-7](https://doi.org/10.1016/0024-3841(76)90060-7)
- Ashby, W. J. (1981). The loss of the negative particle *ne* in French: A syntactic change in progress. *Language*, 57(3), 647–687. <https://doi.org/10.2307/414345>
- Ashby, W. J. (1982). The drift of French syntax. *Lingua*, 57(1), 29–46. [https://doi.org/10.1016/0024-3841\(82\)90069-9](https://doi.org/10.1016/0024-3841(82)90069-9)
- Ashby, W. J. (1984). The elision of /l/ in French clitic pronouns and articles. In E. Pulgram (Ed.), *Romanitas: Studies in Romance Linguistics* (pp. 1–16). University of Michigan Press.
- Ashby, W. J. (1988). The syntax, pragmatics, and sociolinguistics of left- and right-dislocations in French. *Lingua*, 75(2/3), 203–229. [https://doi.org/10.1016/0024-3841\(88\)90032-0](https://doi.org/10.1016/0024-3841(88)90032-0)
- Ashby, W. J. (2001). Un nouveau regard sur la chute du *ne* en français parlé tourangeau: s'agit-il d'un changement en cours? *Journal of French Language Studies*, 11(1), 1–22. <https://doi.org/10.1017/S0959269501000114>
- Auger, J. (1994). *Pronominal clitics in Québec Colloquial French: A morphological analysis*. (PhD Dissertation), University of Pennsylvania.
- Bartoń, K. (2025). *MuMIn: Multi-Model Inference*. R package version 1.48.11. <https://doi.org/10.18637/jss.v067.i01>
- Bentz, C., & Winter, B. (2013). Languages with more second language learners tend to lose nominal case. *Language Dynamics & Change*, 3(1), 1–27. <https://doi.org/10.1163/22105832-13030105>
- Bates, D., Kliegl, R., Vasishth, S., Baayen, H. (2015). *Parsimonious mixed models*. <https://doi.org/10.48550/arXiv.1506.0496704967>
- Blanche-Benveniste, C. (1994). Quelques caractéristiques grammaticales des "sujets" employés dans le français parlé des conversations. In M. Yaguello (Ed.), *Subjecthood and Subjectivity. Proceedings of the Colloquium "The Status of the Subject in Linguistic Theory"* (pp. 77–107). Editions Ophrys.
- Blanche-Benveniste, C. (2003). Les formes grammaticales de réalisation des sujets et leur inégale représentation en français contemporain. In J.-M. Merle (Ed.), *Le sujet* (pp. 73–90). Editions Ophrys.

- Bretz, F., Hothorn, T. & Westfall, P. (2011). *Multiple comparisons using R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420010909>
- Buson, L., Nardy, A., Rousset, I., & Zhang, C. (2022). Pratiques stylistiques en contexte scolaire: usages contrastés du 'ne' de négation chez les adultes et les enfants d'une école maternelle française. In F. Neveu, S. Prévost, A. Steuckardt, G. Bergounioux, & B. Hama (Eds.), *SHS Web of Conferences: Congrès Mondial de Linguistique Française - CMLF 2022* (Vol. 138, pp. 06018). EDP Sciences. <https://doi.org/10.1051/shsconf/202213806018>
- Cabredo Hofherr, P. (2004). Les clitiques sujets du français et le paramètre du sujet nul. *Langue française*, 141, 99–109. <https://doi.org/10.3406/lfr.2004.6592>
- Chervel, A. (2011). Histoire de l'orthographe française : réformes, enseignement, niveau. In O. Bertrand & I. Schaffner (Eds.), *Variétés, variations & formes du français* (pp. 195–210). Palaiseau: Les Editions de l'Ecole Polytechnique.
- Chevrot, J.-P. (forthcoming). Acquiring Sociolinguistic Variation in a First Language: Toward a Socialization-Based Framework. In Y. Asahi, A. D'Arcy & P. Kerswill (Eds.), *The Routledge Handbook of Variationist Sociolinguistics*. Routledge. <https://hal.science/hal-04377247v1>
- Clark, E. V. (2004). Pragmatics and language acquisition. In L. R. Horn & G. L. Ward (Eds.), *The Handbook of Pragmatics* (pp. 562–577). Blackwell. <https://doi.org/10.1111/b.9780631225485.2005.00027.x>
- Costa, J., Fiéis, A. & Lobo, M. (2015). Input variability and late acquisition: Clitic misplacement in European Portuguese. *Lingua*, 161, 10–26. <https://doi.org/10.1016/j.lingua.2014.05.009>
- Coveney, A. (2002). *Variability in Spoken French: A sociolinguistic study of interrogation and negation*. Elm Bank.
- Coveney, A. (2011). A language divided against itself? Diglossia, code-switching and variation in French. In F. Martineau & T. Nadasdi (Eds.), *Le français en contact : Hommages à Raymond Mougeon* (pp. 51–85). Presses de l'Université de Laval.
- Culbertson, J. (2010). Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language*, 86(1), 85–132. <https://www.jstor.org/stable/40666300>
- De Cat, C. (2005). French subject clitics are not agreement markers. *Lingua*, 115(9), 1195–1219. <https://doi.org/10.1016/j.lingua.2004.02.002>
- De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511519789>
- Demuth, K. & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 35(1), 99–127. <https://doi.org/10.1017/S0305000907008276>

- Donaldson, B. (2017). Negation in near-native French: Variation and sociolinguistic competence. *Language Learning*, 67(1), 141–170. <https://doi.org/10.1111/lang.12201>
- Eghbaria-Ghanamah, H., Ghanamah, R., Shalhoub-Awwad, Y., Adi-Japha, E., & Karni, A. (2020). Recitation and listening to nursery rhymes in the familiarization with a literacy language in kindergarteners: Not kids' stuff. *Developmental Psychology*, 56(12), 2195–2211. <http://dx.doi.org/10.1037/dev0001124>
- Ferguson, C. A. (1959). Diglossia. *Word*, 15(2), 325–340. <https://doi.org/10.1080/00437956.1959.11659702>
- Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist*, 66(6), 103–114. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00060
- Gan, Q. (2024). Different registers, different grammars in second language production? The dative alternation in spoken and written Chinese learner English. *Lingua*, 309, 103790. <https://doi.org/10.1016/j.lingua.2024.103790>
- Ghimenton, A., Nardy, A. & Chevrot, J.-P. (2021). Introduction: Bridging contexts to document sociolinguistic variation in acquisition *Sociolinguistic Variation and Language Acquisition across the Lifespan* (pp. 2–8). John Benjamins. <https://doi.org/10.1075/silv.26.int>
- Grohmann, K. K., Kambanaros, M., Leivada, E., & Rowe, C. (2016). A developmental approach to diglossia: bilectalism on a gradient scale of linguality. *Poznan Studies in Contemporary Linguistics*, 52(4), 629–662. <https://doi.org/10.1515/psicl-2016-0025>
- Guy, G. R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, 52, 63–71. <https://doi.org/10.1016/j.pragma.2012.12.019>
- Guy, G. R. & Hinskens, F. (2016). Linguistic coherence: Systems, repertoires and speech communities. *Lingua*, 172(173), 1–9. <http://dx.doi.org/10.1016/j.lingua.2016.01.001>
- Hansen, A. B. & Malderez, I. (2004). Le *ne* de négation en région parisienne : une étude en temps réel. *Langage et société*, 107(1), 5–30. <https://doi.org/10.3917/ls.107.0005>
- Howard, M., Lemée, I. & Regan, V. (2006). The L2 acquisition of a phonological variable: The case of /l/ deletion in French. *Journal of French Language Studies*, 16(1), 1–24. <https://doi.org/10.1017/S0959269506002298>
- Kumarage, S., Donnelly, S. & Kidd, E. (2024). A meta-analysis of syntactic priming experiments in children. *Journal of Memory and Language*, 138, 104532. <https://doi.org/10.1016/j.jml.2024.104532>
- Johnson, E. K. & White, K. S. (2020). Developmental sociolinguistics: Children's acquisition of language variation. *WIREs Cognitive Science*, 11(1), e1515. <https://doi.org/10.1002/wcs.1515>

- Lambrecht, K. (1984). A pragmatic constraint on lexical subjects in spoken French. *Chicago Linguistic Society*, 20, 239–256.
- Labov, W. (1963). The Social Motivation of a Sound change. *Word*, 19(3), 273–309. <https://doi.org/10.1080/00437956.1963.11659799>
- Labov, W. (1964). Stages in the acquisition of standard English. In R. W. Shuy, A. L. Davis & R. F. Hogan (Eds.), *Social Dialects and Language Learning* (pp. 77–104). National Council of Teachers of English.
- Le Normand, M.-T. (2007). Evaluation de la production spontanée du langage oral et de l'activité sémantique du récit chez l'enfant d'âge préscolaire. *Rééducation orthophonique*, 231, 53–71.
- Legendre, G., Culbertson, J., Barrière, I., Nazzi, T., & Goyet, L. (2010). Experimental and empirical evidence for the status and acquisition of subject clitics and agreement marking in adult and child Spoken French. In V. Torrens, L. Escobar, A. Gavarro, & J. Gutiérrez (Eds.), *Movement and clitics. Adult and child grammar* (pp. 333–360). Cambridge Scholars Publishing.
- Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., . . . Singmann, H. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.8.5). <https://cran.r-project.org/web/packages/emmeans/index.html>
- Liang, Y., Donati, C. & Burnett, H. (2024). French subject doubling: A third path. *Isogloss. Open Journal of Romance Linguistics*, 10(7), 1–28. <https://doi.org/10.5565/rev/isogloss.420>
- MacWhinney, B. (2000a). *The CHILDES Project. Tools for Analyzing Talk: Transcription Format and Programs* (Vol. 1, 3rd edition). Lawrence Erlbaum Associates, or at <http://childes.talkbank.org/>
- MacWhinney, B. (2000b). *The CHILDES Project. Tools for Analyzing Talk: The Database* (Vol. 2, 3rd edition). Lawrence Erlbaum Associates, or at <http://childes.talkbank.org/>
- Martineau, F. & Mougeon, R. (2003). A sociolinguistic study of the origins of *ne* deletion in European and Quebec French. *Language*, 79(1), 118–152. <https://www.jstor.org/stable/4489387>
- Massot, B. (2010). Le patron diglossique de variation grammaticale en français. *Langue française*, 168(4), 87–106. <https://doi.org/10.3917/lf.168.0087>
- Massot, B. & Rowlett, P. (2013). Special issue: L'hypothèse d'une diglossie en France. *Journal of French Language Studies*, 23(1). <https://doi.org/10.1017/S0959269512000336>
- McGillycuddy, M., Warton, D.I., Popovic, G., Bolker, B.M. (2025). Parsimoniously fitting large multivariate random effects in glmmTMB. *Journal of Statistical Software*, 112(1), 1–19. <https://doi.org/10.18637/jss.v112.i01>

- Morin, Y.-C. (1979). La morphophonologie des pronoms clitiques en français populaire. *Cahier de linguistique*, 9, 1–36. <https://doi.org/10.7202/800076ar>
- Nagy, N., Blondeau, H. & Auger, J. (2003). Second language acquisition and “real” French: An investigation of subject doubling in the French of Montreal Anglophones. *Language Variation and Change*, 15(1), 73–103. <https://doi.org/10.1017/S0954394503151034>
- Nardy, A., Chevrot, J.-P. & Barbu, S. (2013). The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics*, 51(2), 255–284. <https://doi.org/10.1515/ling-2013-0011>
- Newport, E. L., Gleitman, H. & Gleitman, L. R. (1977). Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In C. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–149). Cambridge: Cambridge University Press.
- Palasis, K. (2012a). *Romance French Child Corpus 1 & 2* (2;5-4;11). Retrieved from: <http://childes.talkbank.org/access/French/Palasis.html>
- Palasis, K. (2012b). Economy principles within child speech: When the nominative Clitic does not surface. In S. Ferré, P. Prévost, L. Tuller, & R. Zebib (Eds.), *Selected Proceedings of the Romance Turn IV Workshop on the Acquisition of Romance Languages* (pp. 261–284). Newcastle: Cambridge Scholars Publishing.
- Palasis, K. (2013). The case for diglossia: Describing the emergence of two grammars in the early acquisition of Metropolitan French. *Journal of French Language Studies*, 23(1), Special Issue 'L'hypothèse d'une diglossie en France', Massot, Benjamin & Paul Rowlett (eds.), 17–35. <https://doi.org/10.1017/S0959269512000348>
- Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua*, 161, 125–143. <http://dx.doi.org/10.1016/j.lingua.2014.11.012>
- Palasis, K. (2024). *Variabilité et acquisition du français hexagonal : comment la syntaxe vient aux enfants*. (Habilitation à Diriger des Recherches, Section CNU 07), Université Côte d'Azur. <https://hal.science/tel-04837205>
- Palasis, K. (2026). *Recherches sur la variabilité et le développement du français langue maternelle : quels liens avec son enseignement de la maternelle au lycée ?* Journée Recherche INSPE, Nice, 4 février 2026. <https://hal.science/hal-05495309>
- Parisse, C. & Le Normand, M.-T. (2000). How children build their morphosyntax: The case of French. *Journal of Child Language*, 27(2), 267–292. <https://doi.org/10.1017/S0305000900004116>
- Pope, M. K. (1934). *From Latin to Modern French*. Manchester University Press.

- Quené, H. & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1–2), 103–121. <https://doi.org/10.1016/j.specom.2004.02.004>
- Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, 94, 85–114. <https://doi.org/10.1016/j.cogpsych.2017.02.004>
- Sénéchal, M., Pagan, S., Lever, R., & Ouellette, G. P. (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. <https://doi.org/10.1080/10409280701838710>
- Snow, C. (1972). Mother's speech to children learning language. *Child Development*, 43, 549–565. <https://doi.org/10.2307/1127555>
- Sorace, A. (2003). Near-nativeness. In C. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 130–151). Blackwell. <https://doi.org/10.1002/9780470756492.ch6>
- Taguchi, N. (Ed.) (2019). *The Routledge handbook of second language acquisition and pragmatics*. Routledge. <https://doi.org/10.4324/9781351164085>
- Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language*, 31, 101–121. <https://doi.org/10.1017/S0305000903005944>
- Veneziano, E. (2000). Interaction, conversation et acquisition du langage dans les trois premières années. In M. Kail & M. Fayol (Eds.), *L'acquisition du langage. Le langage en émergence. De la naissance à 3 ans* (pp. 231–265). Presses Universitaires de France. <https://doi.org/10.3917/puf.fayol.2000.01.0231>
- Voeikova, M. & Krasnoshchekova, S. (2020). The use of pronouns as a developmental factor in early Russian language acquisition. In N. Gagarina & R. Musan (Eds.), *Referential and relational discourse coherence in adults and children* (pp. 171–205). De Gruyter Mouton. <https://doi.org/10.1515/9781501510151-008>
- Washington, J. A. & Seidenberg, M. S. (2022). Language and dialect of African American children. In E. Saiegh-Haddad, L. Laks & C. McBride (Eds.), *Handbook of literacy in diglossia and in dialectal contexts: Psycholinguistic, neurolinguistic, and educational perspectives* (pp. 11–32). Springer. <https://doi.org/10.1007/978-3-030-80072-7>
- Wasik, B. A. & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading and language development in preschool classrooms. *Journal of Educational Psychology*, 93(2), 243–250. <https://doi.org/10.1037/0022-0663.93.2.243>
- Winter, B. (2019). *Statistics for Linguists: An Introduction Using R* (1st ed.). Routledge. <https://doi.org/10.4324/9781315165547>

Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179, 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>

Data, Code and Materials Availability Statement

The full dataset (Colloquial French and Standard French utterances, per year and per child in alphabetical order), the individual tables with counts of occurrences and percentages, the CLAN codes for data extraction and the analysis script files are available on the Open Science Foundation website (OSF, <https://osf.io/29twx>). The exhaustive lists of the occurrences for the three Standard French variants (singleton NPs, full clitics and complex negation) are also provided in the Appendixes.

Ethics Statement

Ethics approval was obtained from the ethics committee of the University of Nice Sophia Antipolis. All children's parents gave informed written consent before taking part in the study. Children provided verbal assent as appropriate.

Authorship and Contributorship Statement

Katerina Palasis: conceptualization, data collection and curation, writing – original draft, review & editing. **Mar Cordero-Rull:** formal analysis, writing – review & editing. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

We are very grateful to Katherine Messenger, Jean-Pierre Chevrot and an anonymous reviewer for their very thorough and constructive comments on the previous versions of the manuscript. We also thank the school, teachers, teaching assistants, parents, and of course children, who all together made this three-year data collection possible. We also thank Brian MacWhinney for his help with sharing the dataset on the CHILDES platform. Finally, thanks also go to our colleagues at the BCL Lab and the Acquisition Team for their support throughout this long-standing project.

Appendixes

Appendix 1: Children's Gender, Age Range and Linguistic Status

Female	CAR	ELA	ELE	LAN	LIN	LIZ	LOU	LUS	MAI	NIN	RIA
Start	3;2.10	3;0.6	3;4.28	2;9.12	2;11.20	3;0.19	3;4.29	2;10.5	2;8.14	2;9.29	2;7.28
End	5;8.19	5;6.15	5;11.6	5;3.21	5;5.29	5;6.28	5;11.7	5;4.14	5;2.23	5;4.7	5;2.6
Status	L1	eL2	L1	L1	L1	L1	L1	L1	L1	L1	eL2

Male	ENZ	KEL	LSN	MAS	MAT	VIC	WIL
Start	3;1.15	3;5.7	2;11.22	2;6.14	2;11.23	3;3.14	2;9.27
End	5;7.24	5;11.16	5;6.0	5;0.23	5;6.1	5;9.23	5;4.5
Status	L1	L1	L1	L1	L1	L1	L1

Appendix 2: Full List of Utterances with Singleton NPs

L1					
n	Year	Session	CHI	Utterance	
1	Y1	XI.43a	ENZ	le beau vert ici [/] est pas ici .	
2	Y1	IV.19d	KEL	oh le train est là.	
3	Y1	II.5	LUS	ta photo n'est pas là? (LUS, 2;10)	
4	Y1	III.13e	LUS	par(ce)+que [/] par(ce)+que [/] par(ce)+que les grenouilles 0 [*] est pieds nus .	
5	Y1	VI.24a	LUS	là tous les oiseaux s' en vont .	
6	Y1	VII.27a	LUS	alors un autre garçon va à sa place !	
7	Y1	VIII.31a	LUS	et après+que Caramel voit par la fenêtre +...	
8	Y1	VIII.31a	LUS	+, si maman est là .	
9	Y1	IX.34b	LUS	et pour écrire les enfants vient [*] avec toi .	
10	Y1	X.38a	LUS	et la souris veut casser ses pics .	
11	Y1	X.38a	LUS	alors <les taupes> [/] <les taupes> [/] <les taupes> [/] les taupes va [*] dans l(e) nid .	
12	Y1	X.38c	LUS	+, les vers+de+terre sont tout gluants .	
13	Y1	XI.43a	LUS	tout+l(e)+monde a déjà fait .	
14	Y1	XI.43c	LUS	tout+le+monde a gagné .	
15	Y1	XII.44a	LUS	tout+le+monde a répondu .	
16	Y1	XII.44c	LUS	alors je le retourne et personne me le retourne comme ça .	
17	Y1	XII.44d	LUS	alors la maman prépare quelque+chose dans le four .	
18	Y1	XIII.46a	LUS	<et mon> [/] et ma petite fourchette marche marche .	
19	Y1	XIII.46a	LUS	et ma petite cuillère marche marche .	
20	Y1	XIII.46b	LUS	vole mon petit chagrin xxx .	
21	Y1	XIII.46e	LUS	euh le poisson nage sur moi .	

22	Y1	III.10b	MAI	si quand [/] quand euh le [/] le papa est [/] est fâché .
23	Y1	XIII.46e	MAI	quelqu+un veut choisir un p(e)tit bonbon ?
24	Y1	VIII.32a	NIN	<pas moi> ["] dit le chat .
25	Y1	VIII.32a	NIN	<pas moi > ["] dit le canard .
26	Y1	VIII.32a	NIN	<pas moi > ["] dit le cochon .
27	Y1	VIII.32a	NIN	<pas moi > ["] dit le chat .
28	Y1	VIII.32a	NIN	<pas moi > ["] dit le chat .
29	Y1	VIII.32a	NIN	<pas moi > ["] dit le canard .
30	Y1	VIII.32a	NIN	<pas moi > ["] dit le cochon .
31	Y1	VIII.32a	NIN	<pas moi > ["] dit le cochon .
32	Y1	XIII.48b	NIN	et comme les chats ont besoin de xx .
33	Y1	IX.36	VIC	le facteur n'est pas passé.
34	Y1	I.4	WIL	mon [/] mon petit chat 0 [*] appelle Moon .
35	Y2	IX37b	CAR	alors aujourd'hui petit chat fait des bêtises
36	Y2	VI22a	ELE	tout+l(e)+monde a mangé
37	Y2	VII29b	ELE	regarde Juliette est en vacances
38	Y2	VII29b	ELE	Juliette est en vacances
39	Y2	VII29b	ELE	voilà Juliette est en vacances voilà
40	Y2	IX37a	ELE	tout+l(e)+monde faisait la fête et tout+l(e)+monde dansait
41	Y2	IX37a	ELE	tout+l(e)+monde faisait la fête et tout+l(e)+monde dansait
42	Y2	IX37a	ELE	après elle soufflait et tout+l(e)+monde goûtait et disait miam miam
43	Y2	X41c	ELE	tout+le+monde me donne +/.
44	Y2	III10a	KEL	euh [/] euh ma maman m' a xx quand j' étais endormi
45	Y2	III10a	KEL	ma maman est allée me donner mon p(e)tit+déjeuner
46	Y2	V19c	KEL	des [/] des [/] des [*] écureuils vit [*] dans les arbres
47	Y2	V19d	KEL	mais lui n' est pas d'accord
48	Y2	V19d	KEL	et là tout+l(e)+monde est content
49	Y2	VII27a	KEL	oui parc(e)+que [/] parc(e)+que ma mère a dit ça
50	Y2	VII27c	KEL	quelqu+un a abîmé [*] .
51	Y2	VII27c	KEL	pourquoi quelqu+un l' a abîmé xxx ?
52	Y2	X40b	LIN	Juliette fait son anniversaire xxx moi .
53	Y2	X40b	LIN	Juliette fait ses courses .
54	Y2	X41c	LIZ	personne ne l'a lui .
55	Y2	V20b	LAN	le dragon noir <eh ben> [/] eh ben était méchant !
56	Y2	VII29b	LOU	oh le monsieur se dit oh mais moi j'ai un drôle de nez !
57	Y2	X38b	LOU	papa m'a donné les médicaments pour+que je me soigne.
58	Y2	IV14d	LSN	des fois ben oui hein le camion poubelle passe .
59	Y2	IV15a	LUS	i faut les retourner pa(rce+que) [/] pa(r)c(e)+que sinon tout+l(e)+monde va voir les cartes .
60	Y2	IV15b	LUS	i fallait les cacher pour que personne les voie .
61	Y2	IV15c	LUS	l' histoire est finie .

62	Y2	III9a	MAI	eh ben i(l) [/] le Père+Noël m' a apporté une Barbie .
63	Y2	VII27d	MAI	oui mais je veux que Kelian finit .
64	Y2	X39d	MAI	Juliette part en voyage .
65	Y2	II5d	MAS	parce que mon chien me fait jamais piquer.
66	Y2	IV15b	MAT	là le monsieur regardait qu+est+c(e)+qu' [*] i(l) y+avait là+bas
67	Y2	V19d	MAT	mais des fois Liz va à la cantine .
68	Y2	V18b	NIN	il enlève sa veste et tout+le+monde enlève sa veste .
69	Y2	V18b	NIN	et là les [/] les gens regardent tout .
70	Y2	VI23d	NIN	après on dort et après les papas et les mamans arrivent .
71	Y2	VIII32d	NIN	Juliette mange parce+qu'elle a un papillon après .
72	Y2	X39d	NIN	non Juliette fait des courses .
73	Y2	X39d	NIN	et après Juliette pique ses chaussures +//.
74	Y2	X39d	NIN	Juliette veut écraser le rouge à lèvres de sa mère .
75	Y2	X39d	NIN	Juliette prend [/] vole le sac de sa maman .
76	Y2	X39d	NIN	Juliette pr(end) [/] prend sa robe pour [/] pour être belle .
77	Y2	X39d	NIN	et Juliette pique le collier de sa mère .
78	Y2	VIII30a	VIC	mes ballons sont gros
79	Y2	III10a	WIL	pendant+qu(e) le soleil se lève ben moi [/] moi je me réveillais .
80	Y2	IX35a	WIL	et aussi j'étais y allé mais pas dans la piscine où maman travaille .
81	Y3	II10b	CAR	oui <et après> [/] et après tout+le+monde dort .
82	Y3	VIII46a	CAR	après [/] après quelqu+un a mis des oeufs .
83	Y3	II10b	ELE	" Juliette fait les courses" et Juliette xxx .
84	Y3	II10b	ELE	y+a [/] <c' est Juliette> [/] Juliette fait son anniversaire .
85	Y3	II10b	ELE	Juliette fête son anniversaire .
86	Y3	II10b	ELE	et Juliette fait partir tout+l(e)+monde xxx .
87	Y3	II10b	ELE	<et Juliette prend sa> [/] et chacun prend la part du [/] du gâteau .
88	Y3	VIII46b	ELE	euh là tout l(e) monde va voir tes cartes .
89	Y3	I5b	ENZ	"comme un escargot porte sur son dos sa maisonnette aussitôt qu' il pleut il est tout heureux i(l) sort sa tête" .
90	Y3	IV23c	ENZ	on a pas l(e) droit de regarder parc(e)+que j' ai pas envie que quelqu+un regarde .
91	Y3	X58a	ENZ	Mesdames et messieurs à tous ! euh le spectacle va commencer dans deux minutes .
92	Y3	I5c	KEL	" Juliette fête son anniversaire" .
93	Y3	I5c	KEL	" Maman prépare un gâteau au chocolat fourré de fraises et de bonbons <et goûte> [/] et le goûte" .
94	Y3	IV24c	KEL	+, " un vieux monsieur s' avance avec sa canne dans la main" .
95	Y3	IV24b	KEL	tout+l(e)+monde a cinq ans et demi .
96	Y3	IV24b	KEL	< tout+l(e)+monde dans notre classe a cin(q)> [/] tous les bleus ont cinq ans et d(e)mi .
97	Y3	VIII46b	KEL	et pourquoi personne me demande à moi ?
98	Y3	VIII46b	KEL	si quelqu+un a la famille de la maison d'aujourd'hui je l'ai hein !

99	Y3	II9b	LAN	quand Noël est passé .
100	Y3	II9b	LAN	c' est quand Noël est passé .
101	Y3	IV20b	LAN	<et [/] et le bouton va au bouton bleu> [<] .
102	Y3	VIII44a	LAN	et moi j(e) connais un endroit où personne me connaît .
103	Y3	X57a	LAN	euh ben <j' ai> [/] j' ai attendu que Maman pousse pour 0 [*] euh i(l) fasse crotte bébé .
104	Y3	X57c	LAN	et puis l' oiseau se moque tout [?] de lui .
105	Y3	X57c	LAN	et puis comme le chien était très dang(e)reux i(l) rentre à la maison .
106	Y3	X57c	LAN	l' oiseau dit "allez allez vas+y" !
107	Y3	X57c	LAN	et puis l' autre montre ses petites dents toutes cassées .
108	Y3	X57c	LAN	ensuite le chien lui a attrapé l' oi(seau) [/] le [/] le chat .
109	Y3	X60a	LIZ	après Tom est allé mettre ses chaussures .
110	Y3	II12b	LOU	non Juliette va faire les courses !
111	Y3	II12b	LOU	oh " Juju fête son anniversaire" !
112	Y3	IV23c	LOU	<" Juliette va au cirque"> [>] !
113	Y3	IV19a	LSN	ça explique quand les hommes sont apparus sur la Terre .
114	Y3	IV19a	LSN	+, voilà tout+l(e)+monde tombe dans l' eau hein .
115	Y3	IV19a	LSN	ah ben tout+l(e)+monde a quatre ans et moi j(e) vais avoir six .
116	Y3	IV19c	LSN	la Tour_Eiffel vo:le !
117	Y3	IV19c	LSN	ben tu sais que là tous les gens sont dans le château hein !
118	Y3	X55a	LSN	tout l(e) monde veut !
119	Y3	X55c	LSN	le pépé est comme ça .
120	Y3	II9c	LUS	tout+l(e)+monde est là sauf lui +...
121	Y3	VIII44c	LUS	" Juliette fait des courses" .
122	Y3	VIII44c	LUS	alors Juliette fait les courses .
123	Y3	X55c	LUS	Juliette part en vacances .
124	Y3	X55c	LUS	Martine Martine part en vacances hi hi .
125	Y3	IV21a	MAI	[...] et après ça fait " les petits poissons dans l' eau nagent nagent bien sur l' eau" .
126	Y3	VIII43b	MAI	un jour <j' ai> [/] <j' ai> [/] <j' ai> [/] mon papa m' appelait derrière là et [/] et [/] et [/] et <j' étais en+train+de gli(sser)> [/] <j' étais> [/] j' étais en+train+d(e) glisser sur [/] sur le tapis et [/] et [/] et [/] et i(l) croyait que j' allais tomber .
127	Y3	I3b	MAS	Sanbarbe se lave !
128	Y3	I3c	MAS	Sanbarbe s' assoit pour jouer aux cartes .
129	Y3	I3c	MAS	Sanbarbe est sur l'eau tout mouillé.
130	Y3	I3c	MAS	Sanbarbe va se laver.
131	Y3	I3c	MAS	Sanbarbe est debout.
132	Y3	I3c	MAS	Sanbarbe est +/.
133	Y3	I3c	MAS	Sanbarbe était pas content.
134	Y3	X56a	MAS	allez le vol va commencer .
135	Y3	X56a	MAS	le petit avion va commencer !

136	Y3	X56c	MAS	Franklin fait du foot .
137	Y3	X56c	MAS	Franklin fait du foot et avec ses amis .
138	Y3	X56c 175	MAS	Juliette va à l' école .
139	Y3	X56c	MAS	quelqu+un m(e) l' a volé .
140	Y3	I4b	MAT	euh “ Sans_Barbe veut rechercher son trésor” .
141	Y3	II11c	MAT	là les éléphants se jettent .
142	Y3	X56c	NIN	là son copain à Juliette tape des pieds et des mains .
143	Y3	II11b	VIC	les rennes sont là+haut et l(e) Père_Noël sont [*] là+haut hein .
144	Y3	II11b	VIC	les rennes sont là+haut et l(e) Père_Noël sont [*] là+haut hein .
145	Y3	I2c	WIL	et le chat est là .
146	Y3	I2c	WIL	de le [?] papa et l(e) chat est là .
147	Y3	I2c	WIL	et le chat est là .
148	Y3	II7b	WIL	Jean avait [/] avait trop peur .
149	Y3	II7c	WIL	et le chat est là !
150	Y3	II7c	WIL	le chat est là .
151	Y3	II7c	WIL	tout l(e) monde se réveille .
152	Y3	II7c	WIL	<tout est en> [/] tous [/] tous les éléphants est [*] en toutes les couleurs mais sauf Elmer .
eL2				
153	Y3	I	ELA	Elmer était différent et euh de couleur .
154	Y3	I4b	ELA	et voilà Elmer est pas déguisé mais tous les éléphants i(ls) sont déguisés .
155	Y3	I4c	ELA	quelqu+un a déchiré .
156	Y3	II11c 36	ELA	alors Barbar [: Sanbarbe] a perdu ses tatouages .
157	Y3	I3b	RIA	mais Cath(e)rine est là !
158	Y3	I3c	RIA	mais Sanbarbe ne veut pas que +...
159	Y3	I3c	RIA	“ la maison est en carton” .
160	Y3	II7c	RIA	tout l(e) monde se réveille xxx pa(r)c(e)+que y+a <deux O> [/] trois O là .

Appendix 3: Full List of Utterances with Full Clitics

L1				
n	Year	Session	CHI	Utterance
1	Y1	VIII31a	LUS	et après oh il dit <qui est là> ["] !
2	Y1	XI41d	MAI	il ne chante plus . (singing)
3	Y1	XI41d	MAI	il ne saute plus . (singing)
4	Y1	XI41d	MAI	il ne danse plus . (singing)
5	Y1	XI41d	MAI	il ne parle plus . (singing)
6	Y1	XI41d	MAI	il ne xx plus . (singing)
7	Y1	XI41d	MAI	il ne xx plus . (singing)

8	Y1	XI41d	MAI	il ne court plus . (singing)
9	Y1	XI41d	MAI	il ne chante plus . (singing)
10	Y1	XI41d	MAI	il ne parle plus . (singing)
11	Y1	VI26a	NIN	et le petit lapin youpi il ne pleure plus !
12	Y1	VIII32a	NIN	il dit <pas moi > ["] le petit cochon .
13	Y1	XIII48b	NIN	et le canard il fait poum dans l' eau .
14	Y2	VIII31c	CAR	il ne sait pas nager
15	Y2	V19d	KEL	lui il n' est pas d'accord
16	Y2	V19d	KEL	et puis lui il n' a pas envie
17	Y2	IX34a	KEL	il commence à devenir xxx
18	Y2	VII26b	LAN	pourquoi <il n' y> [/] il n' y va pas ?
19	Y2	VIII33a	LOU	et puis comme il avait peur eh ben il ne voulait pas glisser
20	Y2	VIII33a	LOU	mais lui il ne voulait toujours pas
21	Y2	I1b	MAI	il se +/.
22	Y2	VI21d	MAS	xxx il est comme ça parc(e)+que [*] il s' est euh fait xxx [...].
23	Y2	IV14b	NIN	et c' est pour ça qu' il ne marche pas .
24	Y2	IX37b	VIC	il [?] voulait renverser 0 [*]
25	Y2	V16c	WIL	il nous [?] crachait du feu du feu xxx .
26	Y3	IV24c	ELE	alors là i(l) [/] il jette à manger pour [/] à table .
27	Y3	I5b	ENZ	“comme un escargot porte sur son dos sa maisonnette aussitôt qu' il pleut il est tout heureux i(l) sort sa tête” .
28	Y3	VIII48a	ENZ	ben ma soeur euh il [*] m' en a trouvé +/.
29	Y3	I5b	KEL	il [/] il pieut [: pleut] [*] .
30	Y3	IV24b	KEL	“ il faut arrêter ça” .
31	Y3	IV24c	KEL	et (i)l [/] il voit quoi ?
32	Y3	IV24c	KEL	lui là il doit arriver ici .
33	Y3	X57c	LAN	alors il essaie de l' attraper mais il n' aime pas l' eau !
34	Y3	X57c	LAN	et puis il montre ses plus grandes dents .
35	Y3	IV22c	LIZ	<là y+a les> [/] là elle veut des pâtes et là il lui renverse tout dessus.
36	Y3	IV19a	LSN	il s' appelle Ja:ck .
37	Y3	X55b	LUS	il [/] il s' appelle Gilles .
38	Y3	I2c	MAI	c' est Elmer qui [/] qui s(e) réveille en premier et [/] et il [?] court ve(rs) [/] vers le galop .
39	Y3	I4c	MAT	il va faire un tatouage .
40	Y3	II11c	MAT	il descend .
41	Y3	II11c	MAT	il cherche son trésor .
42	Y3	I6b	VIC	il s'appelle Lxx .
43	Y3	II11b	VIC	quand il neige euh oh là là !
44	Y3	IV23c	VIC	y+a écrit “ il faut tuer l(e) dragon” . (pretending he is reading)
45	Y3	IV23c	VIC	là il [/] il tire sur la fille .
46	Y3	I2c	WIL	parc(e)+que [*] il [?] veut plus marcher .

47	Y3	II7b	WIL	y+a un p(e)tit qui est allé pleurer pa(r)c(e)+que [*] i(l) voulait avoir Stitch mais il l' a eu alors euh il a plus pleuré alors euh .
48	Y3	II7c	WIL	mais Mat il [/] il voulait euh xxx .
49	Y3	X57d	WIL	il 0 [*] prend . (0 = la= la bougie)
eL2				
50	Y2	VIII31d	ELA	eh ben il vient le bateau <et il> [/] et i(l) se met .
51	Y2	VIII31d	ELA	il va mettre les gens en prison .
52	Y2	VIII31d	ELA	il va mettre les gens en prison .
53	Y2	X40b	ELA	Sanbarbe il dort
54	Y2	X40b	ELA	il fait une photo son copain
55	Y2	X40b	ELA	il le referme
56	Y3	I4b	ELA	et là il [/] il dort .
57	Y3	I4b	ELA	et après tout [/] il tombe .
58	Y3	II11a	ELA	après <il a> [/] il boivait [: buvait] [*] de la potion magique
59	Y3	II11c	ELA	et après il pleure parc(e)+que il a plus .

Appendix 4: Full List of Utterances with Complex Negation

L1				
n	Year	Session	CHI	Utterance
1	Y1	II.5	LUS	ta photo n'est pas là?
2	Y1	IV.19f	LUS	non non ce n' est pas 0 [*à] moi .
3	Y1	VIII.31a	LUS	maman elle n' est pas là .
4	Y1	XI41d	MAI	il ne chante plus . (singing)
5	Y1	XI41d	MAI	il ne saute plus . (singing)
6	Y1	XI41d	MAI	il ne danse plus . (singing)
7	Y1	XI41d	MAI	il ne parle plus . (singing)
8	Y1	XI41d	MAI	il ne xx plus . (singing)
9	Y1	XI41d	MAI	il ne xx plus . (singing)
10	Y1	XI41d	MAI	il ne court plus . (singing)
11	Y1	XI41d	MAI	il ne chante plus . (singing)
12	Y1	XI41d	MAI	il ne parle plus . (singing)
13	Y1	XIII.46b	MAI	on n(e) sait pas on va voir .
14	Y1	VI26a	NIN	et le petit lapin youpi il ne pleure plus !
15	Y1	IX.36	VIC	le facteur n'est pas passé (nursery rhyme)
16	Y2	VIII31c	CAR	il ne sait pas nager
17	Y2	VII28a	CAR	ne [?] le touche pas trop sinon tu vas le casser Dylan .
18	Y2	VII28b	CAR	ne le touche pas trop Dylan !
19	Y2	III11a	ENZ	je n(e) sais pas .
20	Y2	IX36a	ENZ	eh pourquoi elle ne [?] roule plus ?
21	Y2	V19d	KEL	mais lui n' est pas d'accord
22	Y2	V19d	KEL	lui il n'est pas d'accord
23	Y2	V19d	KEL	et puis lui il n'a pas envie
24	Y2	V19d	KEL	hm parc(e)+que je n(e) sais pas .

25	Y2	IX34a	KEL	je n(e) [?] sais pas où je vais aller . (narrative)
26	Y2	VII26b	LAN	pourquoi <il n'y> [/] il n'y va pas ?
27	Y2	VIII33a	LOU	et puis comme il avait peur eh ben il ne voulait pas glisser
28	Y2	VIII33a	LOU	mais lui il ne voulait toujours pas
29	Y2	V19b	LSN	déjà j' ai un gros poisson rouge <qui est> [/] qui n' existe pas .
30	Y2	IX37a	LSN	ne dis pas n+importe+quoi .
31	Y2	I2c	LUS	hm non je n(e) [?] me rappelle plus .
32	Y2	II6a	MAI	je n(e) sais pas .
33	Y2	V19c	MAT	non n' aie pas peur c' est un crapaud .
34	Y2	IV14b	NIN	et c' est pour ça qu' il ne marche pas
35	Y2	VIII32a	NIN	ben je n(e) sais pas .
36	Y3	II10a	CAR	ne parle pas de [*] elle .
37	Y3	II10a	CAR	ne parle pas de [*] elle s+il+te+plaît .
38	Y3	X58b	CAR	ah non je n' ai pas . (playing cards)
39	Y3	II10b	ENZ	non non non n' approche pas non !
40	Y3	IV24b	KEL	“je n' aime que toi” .
41	Y3	IV24c	KEL	“je n' y vois rien” .
42	Y3	IV24c	KEL	alors y+a marqué alors “<ah ah tu> [/] ah ah tu ne mont(e)ras pas” .
43	Y3	X57c	LAN	alors il essaie de l' attraper mais il n' aime pas l' eau !
44	Y3	VIII47b	LIZ	ne les touche pas Lou ! (playing cards)
45	Y3	IV22a	MAT	non non je n' ai pas +/- (playing cards)
46	Y3	X60c	MAT	je n' ai pas d(e) maison d' autrefois ! (playing cards)
eL2				
47	Y3	I3c	RIA	mais Sanbarbe ne veut pas que +...

License

Language Development Research (ISSN 2771-7976) is published by TalkBank and the Carnegie Mellon University Library Publishing Service. Copyright © 2026 The Author(s). This work is distributed under the terms of the Creative Commons Attribution-Non-commercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits any use, reproduction and distribution of the work for noncommercial purposes without further permission provided the original work is attributed as specified under the terms available via the above link to the Creative Commons website.