Variable ne in the negative utterances of French children and their caregivers

Yiran Chen & Kathryn Schuler

Department of Linguistics

University of Pennsylvania

cheny39@sas.upenn.edu

Competing interests: The authors declare none.

Short title: Variable ne in French children and caregivers

Variable ne in the negative utterances of French children and their caregivers

Abstract

Standard French expresses clausal negation through preverbal *ne* and post-verbal *pas* (e.g., Je *ne* sais *pas*.'I don't know'). However, *ne* is variable and only rarely realized (< 10%) in colloquial speech. While previous research suggests clear constraints on ne-realization in interadult speech, less is known about *ne* in French children's early production and input. In the current study, we analyze ne-retention in the conversational speech of 14 monolingual French-learning children (ages 0;11-8;01) and their caregivers. We found that, for both French children and their caregivers, ne-retention was rare and followed the same linguistic constraint found in interadult speech. Importantly, children followed the constraint from their earliest ne-attestations — age 2;05 on average — supporting extremely early sociolinguistic competence. Further, our results suggested a strong influence of input on variation acquisition: caregivers with higher ne-retention rates had children who used more *ne* and from a younger age.

Keywords: acquisition, variation, morphosyntax, child directed speech, French

Acknowledgements: We thank members of the Language Variation and Cognition Lab as well as the audience at the 15th Arizona Linguistics Circle conference (ALC15) for helpful feedback and suggestions. We are especially grateful to Meredith Tamminga for invaluable input. Funding for this work was provided by the University of Pennsylvania to K. Schuler.

Introduction

Ne-omission in colloquial French

In standard French, verbs are negated by adding *ne* before the verb and a negative element after (e.g. *pas, jamais, plus, rien,* or *personne*) However, in colloquial speech, speakers frequently drop preverbal *ne* (see Table 1).

Table 1. Examples of French negative sentences

		Standard French	**Colloquial French**
a.	'I don't know'	*Je ne sais pas* I neg know not	*Je sais pas* I know not
b.	'I never smoke'	*Je ne fume jamais* I neg smoke never	*Je fume jamais* I smoke never
c.	'I don't want any more cherries'	*Je ne veux plus de cerises* I neg want no-more of cherries	*Je veux plus de cerises*[1] I want no-more of cherries
d.	'There is nothing/no one'	*Il n' y a rien/personne* pro neg there has nothing/no one	*Il y a rien/personne* pro there has nothing/no one

The variable omission of *ne* in colloquial French is thought to have originated gradually, sometime in the 17th Century (Hirschbühler & Labelle, 2004; Martineau & Mongeon, 2003; Palasis, 2015), becoming increasingly common only in recent decades (Agren, 1973; Armstrong & Smith, 2002; Ashby 1981, 2001). As a result, the standard form *ne* occurs only rarely in modern colloquial French, with reported ne-retention rates ranging widely from 36.7% to less than 1% in adult speakers[2] (see Table 2).

[1] This French sentence would be ambiguous – it could also mean 'I want more cherries.'

[2] As suggested by Culbertson (2010), the wide range of ne-retention reported from these corpora is likely explained by regional differences and the composition of the corpus under investigation. The Coveney (1996) corpus, for example, involved conversations between two strangers — a context in which speakers are more likely to use the standard form — while Armstrong (2002) also included conversations between friends.

Table 2. ne-retention rate in colloquial French by publication and region

Publication	Region	ne-retention
Ashby 1981	Tours	36.7%
Ashby 2001	Tours	15.5%
Coveney 1996	Sommes	18.8%
Pooley 1996	Roubaix	7%
Armstrong 2002	Lorraine	1.8%
Sankoff & Vincent 1980	Canada	<1%

Despite being rarely attested, ne-retention in colloquial French appears to be conditioned by several factors both internal and external to the language. Among language-internal constraints, ne-retention has been shown to depend on both the preceding subject and the post-verbal negative element. Speakers are most likely to realize *ne* when it is preceded by full NPs, followed by null subjects, followed by non-clitic pronouns (*ceci, qui, cela, lui, ça*), followed by clitic pronouns (*je,tu, il(s), elle(s), nous, vous, on, ce*) (Agren, 1973; Armstrong & Smith, 2002; Coveney, 1996; Martineau & Mongeon, 2003). Speakers are more likely to omit *ne* when the post-verbal negative morpheme is *pas* (Armstrong & Smith 2002). Among language-external constraints, ne-retention has also been shown to be conditioned by social factors. For example, *ne* is realized more often by linguistically conservative speakers and may be used to signal higher socio-economic status (Armstrong & Smith, 2002; Coveney, 1996). It is also one of the French variables that most actively participates in style-shifting, with *ne* realized much more frequently in formal contexts (Coveney, 1996).

Ne-omission in French-acquiring children and their caregivers

While a great deal is known about the patterns of ne-retention in interadult colloquial French, less attention has been devoted to understanding how children acquire and use this particular sociolinguistic variable in French. One question of interest is whether children's input reflects the same ne-retention patterns we observe in speech between French-speaking adults. Given that

rates of ne-retention are low in adult-to-adult colloquial French (as discussed in section 1.1), one might expect ne-retention rates to be similarly low in child-directed speech. However, another possibility is that the rate of ne-retention in children's input is actually higher than in adult-to-adult colloquial French. Across many other sociolinguistic variables, caregivers have been shown to increase, or boost, their use of standard forms in their child-directed speech (e.g. Foulkes, Docherty & Watt, 2005; Roberts, 2002; Smith, Durham & Fortune, 2007; Smith, Durham & Richards, 2013). For example, in Tyneside, England, Foulkes and colleagues (2005) found that, while word-medial and word-final prevocalic /t/s are rarely realized as the standard form [t] in interadult speech (~10%), caregivers realize the standard form much more often in their speech to young children (59%). Similarly, in Buckie Scotland, Smith and colleagues (2007) found that, instead of using the local monophthong form, pronouncing the word "house" as "hoose", nearly categorically as they do in interadult speech, parents boosted their use of the standard form, "house", to 43% when interacting with their children. Researchers have argued that this boosting is the caregiver's way of facilitating learning (either implicitly or explicitly), particularly when the standard form is rarely attested in colloquial speech.

Another question of interest is whether ne-retention in child-directed speech is dependent on the child in some predictable way. For many sociolinguistic variables, researchers have found caregivers' use of the standard form in child-directed speech to depend on the age and/or gender of the child. For example, both Foulkes et al. (2005) and Smith et al. (2007), found that parents used the standard form less as their children grew older, presumably a function of gradually treating their children as equal — or more adult-like —conversational partners. Foulkes et al. (2005) further reported that such parental modulation is more prevalent in speech to girls than in speech to boys, with caregivers using the standard form more often with girls, perhaps out of implicit or explicit motivation to bias female children to the more positively-evaluated variant.

So far, only a few studies have investigated ne-retention in child-directed speech and findings have been mixed. Choi (1986) and Culbertson (2010), for example, found that French-speaking caregivers retained *ne* only rarely in speech to their children, at similar rates that that have been reported for interadult corpora (Choi 1986[3] reported 8% (p.g. 70) and Culbertson 2010 reported 7.6% (p.g. 95)). On the other hand, Sankoff (2019b) analyzed two French-Canadian child-caregiver dyads — Adele (1;9–2;6) and her mother and Olivier (1;11-4;1) and his father — and found ne-retention rate to be significantly boosted over the interadult level. These parents used *ne* in nearly 20% of all negative utterances to their children — 19.8% (N=49) and 18.4% (N=86) respectively — while French-Canadian adults are reported to retain *ne* only 1% of the time (Sankoff & Vincent, 1980). Further, Sankoff (2019b) found that Olivier's father used much less *ne* as Olivier approached four — consistent with the hypothesis that parents initially boost their use of the standard form, using it less often (and more aligned with rates in interadults speech) as their children grow older (Foulkes et al., 2005; Smith et al., 2007).

Besides the *rate* of ne-retention in child directed speech, another interesting question is whether children's input obeys the same sociolinguistic constraints on ne-retention that we observed in adult-to-adult speech. Recall that, for French-speaking adults, ne-retention is conditioned on both language-internal factors (the preceding subject and the post-verbal negative element) as well as language-external factors (the formality of the social context and the speaker's socioeconomic status). Research on whether caregivers obey these same constraints in speech to their children has been limited, but there is at least some evidence to suggest they do. For example, Culbertson (2010) found that mothers followed the subject constraint in their child-directed speech, realizing *ne* most often with DP subjects, next most often with null subjects, and least often with subject clitics. In a more recent analysis, Sankoff

[3] Choi 1986 analyzed the caregiver speech of two French mothers and one French-Canadian mother, but they were not analyzed separately.

(2019b) noted that Olivier's father mostly used *ne* in formal social contexts like discipline and teaching, retaining *ne* only rarely in casual contexts like play. However, both studies relied on relatively small datasets (Culbertson, 2010: 5 mothers; Sankoff, 2019b: one father), leaving room for future work to both corroborate and expand these findings. Indeed, a larger sample of French-speaking caregivers would allow us to paint a more nuanced picture of children as they acquire the sociolinguistic constraints on *ne*, including how these constraints develop in the French-acquiring child and do (or do not) change with the child's age or gender.

A final interesting question is whether and how French-acquiring children use *ne* in their own productions. To date, only a handful of attestations of *ne* have ever been reported in children's speech, making this aspect of children's acquisition particularly difficult — even impossible — to study quantitatively. While Choi (1986) and Culbertson (2010) did not specifically analyze *ne* in children's early productions, Sankoff's (2019b) reanalysis of Choi's (1986) data confirmed that Adele did not produce *ne* at all during the study (1;9–2;6), despite her adult-like mastery of other negative elements in the grammar. Olivier, on the other hand, did produce *ne* a few times, first at the age of 2;09 and twice (out of 20 negative utterances) in his last recording at age 4 (Sankoff, 2019b). Beyond Sankoff (2019b), to our knowledge, only one additional study has investigated children's production of *ne*, and on a somewhat larger scale. Palasis (2015) reported that ne-realization was extremely rare for children, and did not significantly increase as children grew older (1.2% in children aged 2;4-4;0 and 1.8% in children aged 3;6-4;11). In fact, across the two corpora analyzed, *ne* was attested in only 20 total utterances. While, Palasis (2015) noted that children's use of *ne* seemed to always follow a clitic or non-clitic subject pronoun, *ne* was so rarely attested that they could not quantify whether ne-retention changed over age, or whether children followed the linguistic or social constraints on ne-attestation reported in adult-to-adult speech.

The present study

To summarize, while a small number of studies have investigated ne-retention in children and their language input, many aspects of *ne* acquisition remain unclear. In the present study, we seek to address these open questions by investigating *ne* on a larger scale. We report a corpus analysis of 14 French families across 6 different CHILDES corpora, with recordings of children as young as 1 year and up to 8 years. Our analysis includes 24222 negative utterances, 6887 of which were uttered by children themselves (including children as young as 1 year). Our first aim is to understand the nature of *ne* in children's language input by asking 1) whether *ne* is rare in child directed-speech (as it is in interadult colloquial French), 2) whether caregivers show significant boosting of the standard *ne* form in speech to their children (either in general or dependent their child's age or gender), and 3) whether caregivers follow the robustly attested subject constraint on ne-retention in their child-directed speech. Our second aim is to provide a quantitative analysis of the acquisition pattern of the *ne* variable by children — the first of its kind to our knowledge — specifically asking 4) at what age children begin to show variable use of ne, and 5) whether and by what age children match the subject constraint on ne-retention in their own productions.

Method

Participants

Participants were 14 monolingual French children (8 boys and 6 girls) and their caregivers, selected from 6 different French corpora in the CHILDES database (Bassano & Mendes-Maillochon, 1994; Champaud, 1994; Demuth & Tremblay, 2008; MacWhinney, 2000; Morgenstern & Parisse, 2012; Plunkett, 2002; Suppes, Smith & Leveillé, 1973; Yamaguchi, 2012). To ensure our analysis would capture a representative sample of French-learning children's everyday language experience, we selected only corpora recorded in French homes while families engaged in natural conversation. Further, we selected only families whose

children had at least one recording beyond age 2;05. In a preliminary analysis of the Paris corpus (Morgenstern & Parisse, 2012), we found 2;05 to be the average age at which children first produced *ne*. The 6 corpora are described in Table 3, arranged by year of recording.

Table 3. Corpora Analyzed

Corpus	Age Range	Region	Recording Year	N Children
Leveille	2;01-3;03	Paris	1971-1972	1
Champaud	1;09-3;05	Paris	1988-1989	1
York (Anne)	1;10-3;05	Paris	1997-1998	1
Lyon	0;11-4;00	Lyon	2002-2005	4
Paris	0;10-8;01	Paris	2005-2008	6
Yamaguchi	1;03-4;03	Paris	2006-2009	1
Total				**14**

Procedure

To extract the negative utterances from each corpus, we used a regular expression pattern match targeting post-verbal *pas. Pas* is the most common negator in French and is the first expression of clausal negation children acquire after the anaphoric negator “non” (no) (Dimroth, 2010). As mentioned in the introduction, the ne-omission variable also extends to other negative elements like *personne, jamais, plus, que*, etc. However, we excluded these other negative elements from our analysis because they are significantly more difficult to extract and code. For example, many of these utterances are ambiguous when *ne* is omitted (1) and are therefore impossible to code. Extraction would also be more difficult because many of these negative elements can occur not only post-verbally, but also in subject positions (2). Thus, in the present work, we focus our analysis on negative utterances containing post-verbal *pas*, leaving these more complex negative elements for future work.

(1) Il a plus d' oreille, lui.

‘He has no ear any more’ or ‘He has more ears’ [Anae (2;09) ’s mother]

(2) Personne n’a le droit de rentrer.

‘Nobody has the right to go back in.’ [Madeleine (6;01)]

After extraction, we manually checked each utterance for errors and removed any non-alternating contexts in which it is impossible to realize *ne*. Specifically, we removed utterances containing words that were incorrectly tagged as verbs, where pas negates an adjective or adverb (see (3) where *peut-être* (maybe) was wrongly tagged as a verb). We also removed utterances in which *pas* was used in a fixed expression where it is impossible to realize *ne*. For example, in the expression “ou pas (or not)” (4).

(3) Peut-être pas tout de suite.
‘Maybe not right away’ [Theotime (2;05) ’s mother]

(4) Je dois encore te sortir des enfants ou pas?
‘Should I still take the children outside for you or not?’ [Julie (4;08) ’s mother]

After this manual cleaning, we coded the remaining 24222 negative utterances for corpus region, corpus decade, speaker id, speaker type (caregiver or child), child id, child age, child sex, subject type (NP, clitic pronoun, non-clitic pronoun, or null subject), and whether *ne* was realized or not.

Analysis

We conducted our analysis in *R* (R Core Team, 2021) via Google Colab, a cloud-based Jupyter notebook (Kluyver et al., 2016). Data wrangling and figure creation were accomplished with the *tidyverse* (Wickham et al., 2009).

To determine the rate of ne-realization in caregivers and children, and the constraints governing their patterns of use, we built separate logistic mixed-effect regression models for child and caregiver utterances using the *lme4* package (Bates et al., 2015). In each model, we

predicted whether *ne* was realized by child age (in months, scaled and centered), child sex, and subject type (an ordinal category with four levels: clitic, pronoun, non-clitic pronoun, null-subject, and NP) with random by-child intercepts and random slopes for child age and subject type[4] . To test the acquisition of the subject constraint on ne-realization over age, we also included the interaction between child age and subject type. And because research has suggested that ne-retention rates are decreasing in France and may differ by region (Ashby, 1981; Armstrong and Smith 2002), we included both corpus region (Paris vs Lyon) and corpus decade (an ordinal category with four levels: 1970s, 1980s, 1990s and 2000s) as fixed effects in both models. Complete models in *lme4* syntax are shown below in (5) and (6).

(5) Caregiver model: *ne ~ child age * subject type + child sex + corpus region + corpus decade + (0 + child age + subject type || child name)*

(6) Child model: *ne ~ child age * subject type + child sex + corpus region + corpus decade + (1|child name)*

Finally, to test whether a caregiver's rate of ne-retention impacted their child's patterns of use, we built two additional simple linear regression models. In one model, we used the caregiver's average ne-retention before their child's first attestation (log transformed) to predict the age at which their child first produced *ne*. In another model, we used the caregiver's average ne-retention rate (log transformed) to predict the child's average ne-retention rate (log transformed). No additional predictors were included in either of these models.

Results

[4] Note that the child model failed to converge when including random slopes for child age and subject, so we simplified the random effects structure in the child model to include only random by-child intercepts

We begin with the full model coefficients from our caregiver and child mixed-effects models (see Table 4). In the sections that follow, we describe these results in more detail in the context of our research questions.

Table 4. Model coefficients. Significant predictors are highlighted in gray.

Fixed effects	Caregiver model (n=17335)				Child model (n=6887)			
	β	SE	z	p	β	SE	z	p
(Intercept)	-1.873	0.453	-4.138	<0.001	-4.403	0.446	-9.879	<0.001
Child age	-0.070	0.109	-0.642	0.521	0.621	0.102	6.084	<0.001
Subject type - Linear	2.282	0.255	8.957	<0.001	1.397	0.279	5.003	<0.001
Subject type - Quadratic	1.006	0.222	4.529	<0.001	0.640	0.292	2.194	0.028
Subject type - Cubic	0.218	0.184	1.179	0.238	0.494	0.303	1.630	0.103
Child sex - Male	-0.436	0.346	-1.262	0.204	-0.241	0.260	-0.926	0.354
Region - Paris	0.672	0.336	1.998	0.045	0.850	0.344	2.472	0.013
Decade - Linear	-0.254	0.535	-0.476	0.634	0.167	0.372	0.449	0.653
Decade - Quadratic	-0.362	0.614	-0.589	0.556	0.755	0.549	1.374	0.169
Decade - Cubic	1.749	0.724	2.416	0.016	0.954	0.704	1.356	0.175
Child age x Subject type.L	-0.175	0.136	-1.285	0.199	0.202	0.181	1.112	0.266
Child age x Subject type.Q	-0.152	0.121	-1.253	0.210	-0.013	0.189	-0.070	0.944
Child age x Subject type.C	0.002	0.104	0.014	0.989	0.276	0.199	1.389	0.165

Region and decade of recording

Consistent with reports of regional differences in *ne* usage in interadult speech (Ashby, 2001; Armstrong & Smith, 2002; Coveney, 1996; Pooley, 1996), region of recording was a significant predictor of ne-realization in both our caregiver and child models. As shown in Table 4, caregivers and children in Paris were significantly more likely to realize *ne* than those in Lyon (Caregivers: β=0.672, SE=0.336, p=0.045; Child: β =0.850, SE=0.344, p=0.013). In contrast, despite reports of the rapid decline of *ne* usage in interadult speech in recent decades (Ashby, 1981, 2001; Armstrong & Smith 2002), we did not find a decreasing trend in ne-realization by decade in our sample. Our child model revealed no significant trends in ne-realization across decades, suggesting that the rate of ne-realization in children's productions has been stable

since the 1970s. Our caregiver model on the other hand revealed a significant cubic trend (β=1.749, SE= 0.724, p=0.014), suggesting a non-linear change in ne-realization rates in child-directed speech across decades. Indeed, this can be seen clearly in Figure 1: ne-realization among our caregivers increases from the 1970s to the 1980s, but decreases again between the 1980s and the 2000s.

We want to emphasize that one should interpret our observed differences in region and decade with caution, given the large individual differences among caregivers in our sample, even in the same corpus. To cite a dramatic example, though both Julie and Theophile's parents are middle-class Parisians in their thirties in the early 2000s, Julie's parents realized *ne* 33.99% of the time, whereas Theophile's parents realized *ne* only 3.5% of the time. Indeed, these individual differences coupled with the over-representation of data sampled from Paris in the early 2000s may have skewed any apparent-time or regional differences we might otherwise observe in the population.

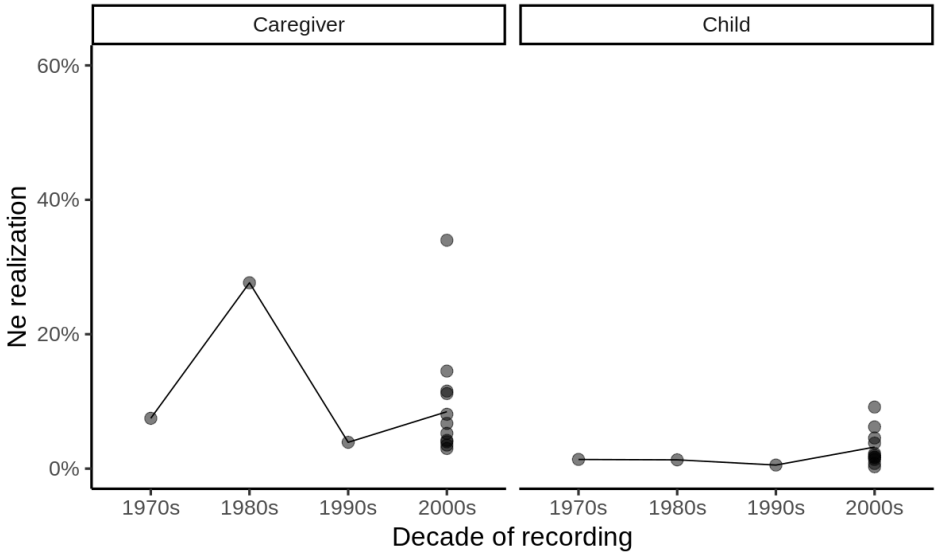

Figure 1. Mean ne-realization by decade of recording for caregivers (left) and children (right). Each dot is the mean rate of ne-realization for a given caregiver (or child) in our sample.

Rate of ne-realization by children and their caregivers

We turn next to the rate of ne-realization by children and their caregivers. We hypothesized that *ne* would be rare in child-directed speech, given that *ne* is reported to be rare among adult speakers of colloquial French (see Table 2) and in small samples of child-directed speech by French-speaking mothers (Choi (1986): 3 mothers, 8%; Culbertson (2010): 5 mothers, 7.6%). Further, we hypothesized that *ne* would be similarly (or even more) rare in children's own productions, given that there have been only 23 attestations of *ne* ever reported in children's negative utterances (3 in Sankoff's (2019b) analysis of Olivier and 20 in Palasis (2015)'s analysis of two corpora).

Table 5 shows the overall rate of ne-retention for all children and their caregivers in our sample, arranged by corpus and caregiver retention rate. For caregivers, we found the average rate of ne-retention to be relatively low (mean=8.49%), though the range among individuals was quite wide. Julie's caregivers realized *ne* most often, in 243 of 715 negative utterances (33.99%) while Anais's caregivers realized *ne* the least, in just 46 of 2636 negative utterances (3.00%). Results from our logistic mixed-effect model confirm our observations: *ne* is rarely attested in speech to young children. Our caregiver model has a significant negative intercept, indicating the log odds of ne-realization are significantly lower than chance (50%) among caregivers in our sample (β=-1.873, SE=0.453, p<0.001). Our results are consistent with the low rates of ne-realization reported by Choi (1986) and Culbertson (2010), lending further support to the notion that *ne* occurs only rarely in child-directed speech.

Table 5. Use of ne in negative utterances (neg) by each child and their caregivers, arranged by corpus. Asterisk (*) indicates the first negative occurs in the child's first available recording.

Child	Sex	Age range	Age of first		Child ne		Caregivers ne	
			neg	ne	ne/neg	%	ne/neg	%
Leville								
Phillippe	M	2;01-3;03	2;01	2;02	11/806	1.36	104/1389	7.49
Champaud								
Gregoire	M	1;09-3;05	1;09*	2;05	2/153	1.31	52/188	27.66
York								
Anne	F	1;10-3;05	1;10*	1;11	3/567	0.53	17/434	3.92
Lyon								
Marie	F	1;00-4;00	1;00*	2;05	9/554	1.63	294/2026	14.51
Nathan	M	1;00-3;00	1;10	2;06	3/152	1.97	83/1582	5.25
Theotime	M	0;11-3;00	1;04	2;04	3/397	0.76	78/1857	4.20
Anais	F	1;00-3;00	1;11	2;09	1/364	0.55	49/2636	3.00
Paris								
Julie	F	0;10-8;01	1;11	2;06	24/262	9.16	243/715	33.99
Antoine	M	1;00-6;03	1;06	2;05	38/612	6.21	179/1552	11.53
Anae	F	1;04-5;10	1;04*	2;04	23/606	3.80	135/1209	11.17
Leonard	M	1;08-3;02	1;08*	2;04	4/232	1.72	52/643	8.08
Madeleine	F	1;00-6;11	1;07	2;04	47/1031	4.56	74/1099	6.73
Theophile	M	1;00-4;11	2;02	2;11	9/662	1.36	64/1827	3.50
Yamaguchi								
Adrien	M	1;03-4;03	2;08	2;10	11/489	2.25	47/1178	3.99
Mean of all children			**1;08**	**2;05**		**2.73%**		**8.49%**

For children, we found the average rate of ne-realization to be even lower than their caregivers (mean = 2.73%). Among the children in our sample, Julie realized *ne* most often, in 24 of her 262 (9.16%) negative utterances, while Anne realized *ne* the least, in just 3 of her 567 (0.53%) negative utterances. As in our caregiver model, our child model revealed a significant negative intercept, indicating that that log odds of ne-realization were significantly lower than chance (50%) in children's productions (β=-4.403, SE=0.446, p<0.001). We can also observe from Table 5 that children retained *ne* less often than their caregivers on average (mean difference = 7.72%), and no child retained *ne* more often than their caregivers. Adrien matched

his caregivers most closely, retaining *ne* only 1.74% less often than they did, while Gregoire's ne-retention was furthest from his caregivers, differing by 26.35%.

Acquisition of variable ne by age and sex

Beyond the average ne-retention rate in children and their caregivers, we also asked whether children's acquisition of *ne* differed by child age or sex. While previous reports of children's ne-retention has been limited — neither Sankoff (2019) nor Palasis (2015) had sufficient data to analyze the developmental trajectory of *ne* in children — researchers have found age- and sex-dependent patterns in children's acquisition of other sociolinguistic variables. For example, many researchers agree that, while children can produce variable forms from a young age, they may not show adult-like knowledge of the constraints governing this variation until they are older (e.g. Labov, 1989; Miller, 2013; Roberts, 1997; Shin, 2016; Smith et al., 2007, 2013). Further, for some variables, female children are more likely to use standard (or socially marked) variants than their male peers (e.g. Fischer, 1958; Purcell, 1984; Roberts, 1997; Romaine, 1978;).

We begin by describing the age at which children in our sample first produce *ne* in negative utterances. As shown in Table 5, on average, children produce their first negative utterance at 1 year, 8 months. The earliest attested negative utterance comes from Marie at age 1 year (e.g., *Il a pas sommeil*. 'He's not sleepy.'), while the latest comes from Adrien, who produced no negative sentences until 2 years, 8 months (e.g., *Ouh, c'est pas là.* 'Oh, it's not there.'). On average, nine months after producing their first negative sentence, children produce *ne* for the first time (mean age = 2;05). Annie is the first to realize *ne* in a negative sentence at 1 year, 11 months (e.g., *Ils n' entendent pas.* 'They can't hear.'), while Theophile is the last, at 2 years, 11 months (e.g. *(Je) n' ai pas fait encore.* 'I haven't done it yet'). Taken together, these observations suggest that children begin producing negative sentences around 1 year, 8

months, but do not produce the standard form, *ne*, until sometime between 23 and 36 months of age (approximately 2-3 years).

While children do not initially produce *ne* in negative sentences, our results suggest that ne-retention is indeed age-dependent in children. In our child model, child age is a significant predictor of ne-retention (β=0.621, SE=0.102, p<0.001). As can be seen in Figure 2, *ne* is unattested in the youngest children, but approaches the adult level of ne-retention as the children grow older. Sex, on the other hand, was not a significant predictor of ne-retention in our child model. Male children are no less likely to realize *ne* than their female peers (β=-0.241, SE=0.260, p=0.354).

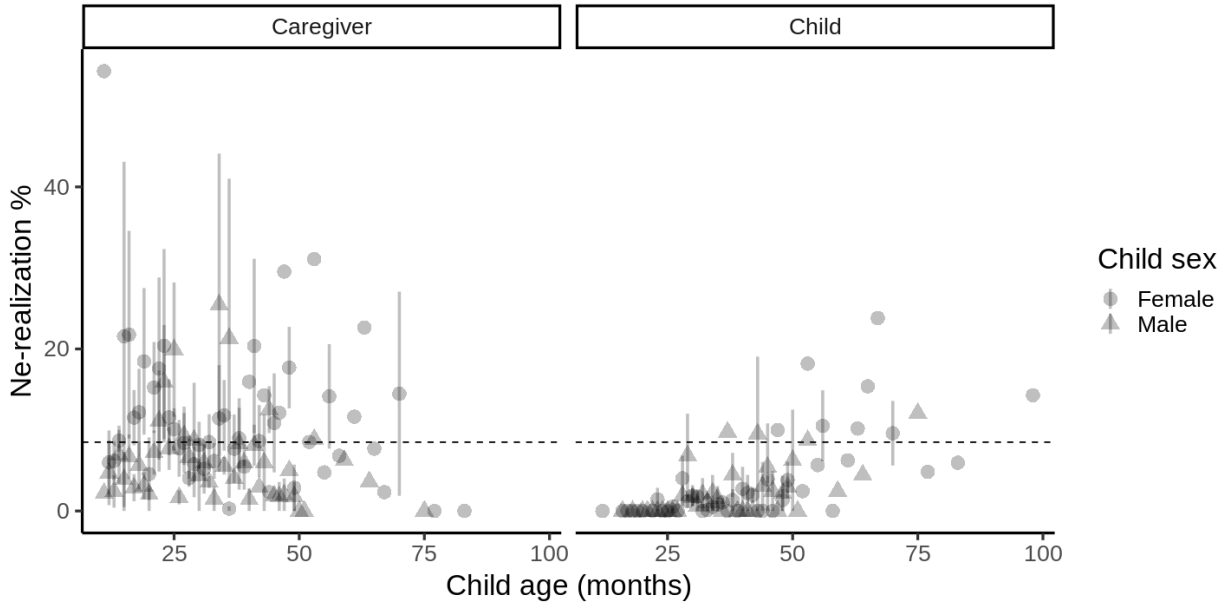

Figure 2. Average ne-retention rate for caregivers and children by child age in months and child sex. Error bars reflect standard error and the dashed line reflects mean ne-retention rate for caregivers (i.e. the average ne-retention in children's input). [5]

The role of the input on the acquisition of variable ne

To investigate the role of children's input on their acquisition of *ne*, we first ask whether caregivers bias their children toward ne-realization early in the acquisition process. Previous

[5] We are aware that the first datapoint at 11 months for caregivers is very high. However, it is within 3 standard deviations of the mean ne-retention rate and therefore not excluded.

studies have demonstrated that caregivers tend to boost their use of the standard form in child-directed speech when their children are young, then gradually reduce usage as their children age (Foulkes et al., 2005; Smith et al., 2007). Researchers have hypothesized that such boosting may have a facilitative effect on the acquisition of (otherwise rarely attested) standard forms. However, while preliminary evidence suggests a similar age-dependent pattern for ne-realization in child-directed French — in her reanalysis of Olivier (Choi, 1986), Sankoff (2019b) reported that Olivier's father realized *ne* less often as Olivier approached age four — we found no such pattern in our sample of child-directed speech. In our caregiver model, child age was not a significant predictor of ne-realization (β=-0.070, SE=0.109, p=0.521), indicating that caregivers in our sample did not boost their ne-realization rate when their children were young (or otherwise adjust their *ne* usage based on their child's age, see Figure 2).

Next, we investigated another property of the input commonly reported in the developmental sociolinguistic literature: sex-dependence. Recall from section 3.3 that we did not find children's ne-retention to differ by sex. However, for many sociolinguistic variables, researchers have found that the input itself differs by child sex, with caregivers using more social marked variants with girls than boys (e.g. Foulkes et al., 2005). For *ne*, however, we found no such pattern. In our caregiver model, child sex was not a significant predictor of ne-realization in caregivers (β=-0.436, SE=0.346, p=0.204; Table 4), suggesting that caregivers do not use more *ne* with girls than boys.

Finally, while we did not find any age- or sex-dependent patterns in our sample of child-directed speech, we did find that children are sensitive to the input from their caregivers in other ways. For example, we built a simple linear model to predict the age at which children first produced *ne* by their caregivers' average ne-retention before that age (log transformed). As shown in Figure 3 (left), we found that caregivers with the highest ne-retention rate had children who produced *ne* the earliest (β=-2.100, SE=0.728, p=0.014). We also built a second model to predict children's average ne-retention (log transformed) by their caregivers' average

ne-retention (log transformed). As shown in Figure 3 (right), we found that caregivers with the highest ne-retention had children who produced more *ne* overall (β=0.771, SE=0.285, p=0.019). Taken together, these results indicate that children are indeed sensitive to the *ne* in their input — caregivers who use more *ne* have children who produce *ne* earlier and at higher rates — but caregivers do not tailor their ne-retention rate to their child's age or sex.

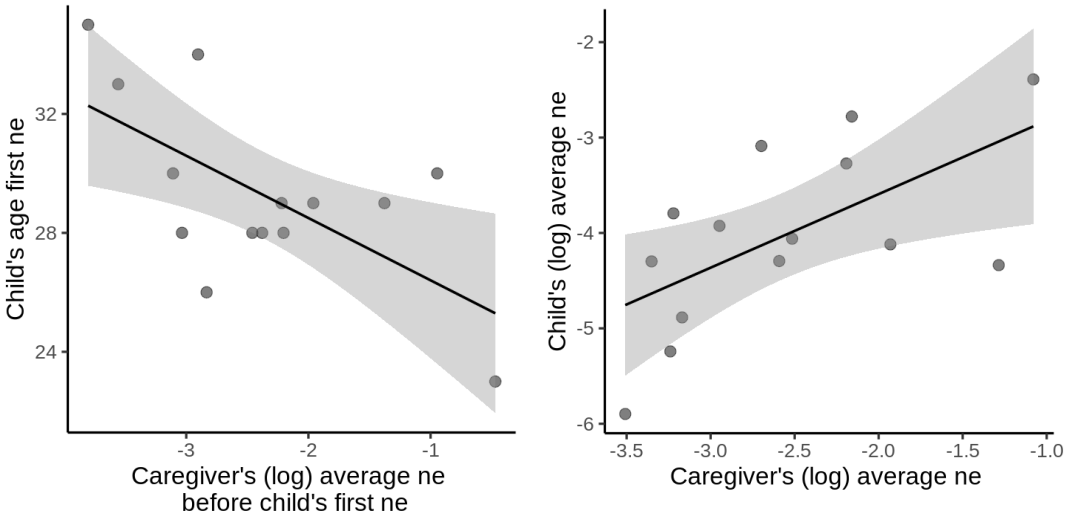

Figure 3. Caregiver's average ne-retention before their child's first attestation (log transformed) predicts the age at which their child produced *ne* (left). Caregiver's average ne-retention rate (log transformed) predicts their child's average ne-retention rate (log transformed).

Acquiring the linguistic constraints on variable ne

A final question to consider is whether children and their caregivers follow the linguistic constraints on ne-retention. Recall that we hypothesized that caregivers (and therefore the child's input) would likely obey the linguistic constraints on ne-retention, given such constraints have been observed in adult-to-adult speech (Agren, 1973; Armstrong & Smith, 2002; Coveney, 1996; Martineau & Mongeon, 2003). Recall that speakers are most likely to realize *ne* with full NPs, followed by null subjects, non-clitic pronouns, and clitic-pronouns (in that order). To test this, we included Subject type as an ordered factor in our models with four levels (clitic pronoun, non-clitic pronoun, null subject, full NP). As predicted, our adult model revealed a significant linear trend (β=2.282, SE=0.255, p<0.001), indicating that the log-odds of ne-realization

increase by subject type for caregivers (when ordered from clitic pronoun to full NP). Further, our adult model revealed a significant quadratic trend as well (β=1.006, SE=0.222, p<0.001), which suggests that the slope of the trend is also increasing. This reflects what we see in Figure 4 (left): caregiver's ne-retention follows a linear increase from clitic pro to null subject, with a dramatic increase in slope going from null subject to full NP.

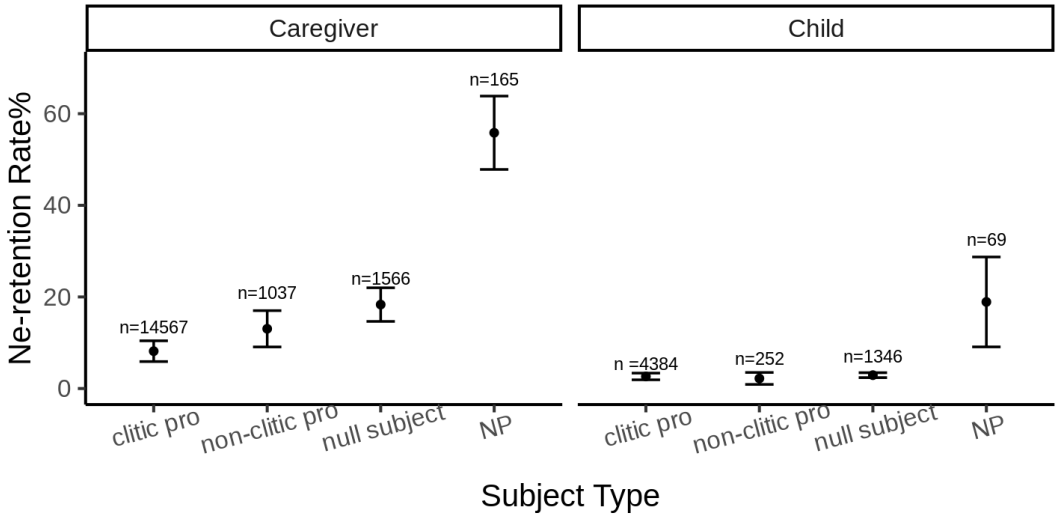

Figure 4 Average ne-retention rate in "verb+pas" sentences by subject type for caregivers and children. (For children, only data after the first attestation of *ne* is included.)

Our results suggest that the linguistic constraints on ne-retention are available in French children's input, but have children themselves acquired these constraints? Results from our child model suggest that they have. We found the same results for children as we did for caregivers: a significant linear trend in Subject type (β=1.397, SE=0.279, p<0.001), with a significant quadratic term (β=0.640, SE=0.292, p=0.028). As shown in Figure 4 (right), children's ne-retention by Subject type follows the same trajectory as adults, increasing linearly from clitic pro to NP, with a much steeper slope between null subject and full NPs.

Finally, to test whether children's mastery of linguistic constraints changes with age, we included a Child age by Subject type interaction in our model. We found no significant interaction (see Table 4), suggesting that children follow the linguistic constraints on ne-retention

as early as they produce it. While this may seem surprising, given the composition of children's negative sentences changes dramatically with age (see Figure 5), it aligns with developmental work on other sociolinguistic variables suggesting that children master grammatical constraints on variation from a very young age (2-3 years) (e.g., Smith et al., 2007, 2013).

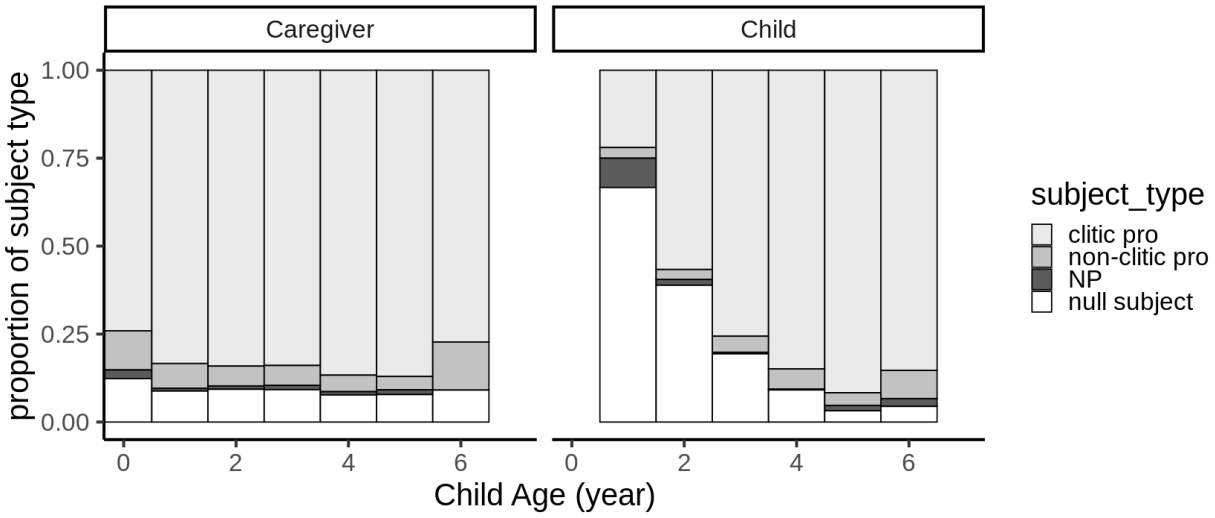

Figure 5 Proportion of each subject type in all negative utterances across children's age

Discussion

The goal of this paper was to investigate the use of variable *ne* in the negative utterances of children and their caregivers. In this section, we return to our primary research questions and discuss our results in the broader context of children's acquisition of linguistic variation.

Is ne rare in child-directed speech?

Our results show that, in general, *ne* is as rare in child-directed speech as it is in interadult colloquial French. Caregivers in our sample realized *ne* in 8.49% of their child-directed negative utterances, which is similar to many recent reports of interadult speech (Ashby, 2001: 15.5% (Tours); Coveney, 1996: 18.8% (Somme); Hanson & Malderez, 2004: 8.2% (Paris);

Pooley, 1996: 7% (Roubaix)). Recall that this similarity was not guaranteed: while Choi (1986) and Culbertson (2010) reported similarly low ne-retention rates in child-directed speech (8% and 7.6% respectively), many previous studies have observed that caregivers boost their use of standard variants in speech to their children (e.g. Foulks et al., 2005; Smith et al., 2007). Indeed, Sankoff (2019) found evidence of such boosting for *ne* among Canadian French speaking families: Adele and Olivier's parents used *ne* in nearly 20% of their negative utterances — dramatically more often than Canadian French interadult speech (1%). In our current dataset, however, although there are also individual caregivers with higher ne-retention rate (e.g. Julie's mother: 33.99%), most caregivers in our dataset realize *ne* rarely, at a rate comparable to what's reported for inter-adult speech.

Do caregivers boost their use of the standard form in speech to their children?

As reported above, we did not find evidence that caregivers boost their use of *ne* in child-directed speech. Importantly, however, our findings compare caregivers' speech with previous reports of ne-realization in interadult speech — a completely different sample of speakers, who are likely to differ in many ways (geographical region, age, socioeconomic status, method of data collection, etc). While this is the best available comparison to date, future studies would ideally compare child-directed and adult-directed speech in the same speakers. One way to accomplish this would be to record conversations between caregivers and their children in the evening; then, continue recording these caregivers with their partners or friends after their children have gone to bed.

Beyond general boosting of the standard form, we also asked whether caregivers' ne-relization was predicted by child age or sex. For other sociolinguistic variables, research suggests that parents boost their use of the standard form when their children are young (Foulkes et al., 2005, Smith et al 2007) and use more socially-favored variants with girls than boys (Foulkes et al., 2005). While Sankoff (2019) found evidence of an age-dependent pattern

for *ne* specifically — Olivier's father used *ne* less often as Olivier grew up — we found that neither child age nor sex reliably predicted caregivers' use of *ne* in our bigger sample. One explanation for this difference could be regional: Olivier and his father spoke Canadian French, a dialect for which ne-retention is reported to be exceptionally low (1%). Perhaps caregiver boosting is employed most often (or is most necessary for acquisition) under circumstances when a variant is extremely rare.

Do caregivers follow the subject constraint on ne-retention in their child-directed speech?

Yes, like Culbertson (2010) we find that parents' child-directed speech is conditioned by the same subject constraint that conditions interadult speech. As shown in Figure 4, *ne* is realized most often in caregivers' speech when preceded by NP subjects, followed by null-subjects, followed by non-clitic pronouns, and is least favored when preceded by clitic pronouns. Note that this retention pattern follows the same ordering as what has been reported for interadult speech (Agren,1973; Armstrong & Smith, 2002; Coveney, 1996; Hanson & Malderez, 2004; Martineau & Mongeon, 2003), confirming that the subject constraints on ne-retention are available in French children's language input.

Recall that we constrained our analysis to negative utterances containing post-verbal *pas* because doing so made our analysis more tractable. As we described in our study methods, other negative elements are both harder to extract (they can occur in subject positions as well as after the verb) and harder to code (the speaker's intended meaning can be ambiguous when *ne* is omitted). However, we are aware that another linguistic constraint on *ne* is the post-verbal negative element itself, where *pas* negatives are most likely to trigger *ne* omission (Armstrong & Smith, 2002). Given our analysis excludes other negative elements, we are unable to determine whether this constraint is available in child-directed speech or whether children follow it in their own negative utterances. Further, because our analysis focuses on pas specifically — the negative utterances that are most likely to trigger *ne* omission — it is possible that we have

underestimated the true rate of *ne* retention in both children and their caregivers. Still, our analysis offers an important first step in characterizing the nature of *ne* in child-directed speech and children's own productions.

At what age do children begin to show variable use of ne?

Our results suggest that children show variable use of *ne* from the moment they produce it. Turning first to when children begin to produce *ne*, we found *ne* production to be delayed compared to children's negative utterances in general. As reported in Table 2, *ne* begins to occur in children's productions as they approach age two (mean age = 2;05), an average of 9 months after their first *pas* negative. Our results are not surprising, given previous findings that children acquire "optional" or variably realized morphemes later than obligatory morphemes. Marrero and Aguirre (2003), for example, found that children acquiring Spanish dialects with variable /s/ lenition first produced the overt plural marker when they were age 3;0, over a year later than children acquiring the non-leniting dialects. Similarly, Miller and Schmitt (2012) found that children acquiring a leniting variety of Spanish take longer to associate a plural interpretation with the presence of a plural marker than children acquiring a non-leniting variety.

Turning next to variable use of *ne*, we found that children produce *ne* variably from their first attestation. As shown in Figure 2, *ne* is rare in children's productions. In fact, our results suggest that, on average, children have not yet matched their parents' rate of realization (2.73% vs 8.49%), though they do use *ne* more often as they grow older. This is reminiscent of the age-dependent ne-realization reported by Sankoff (2019a) for Canadian-French speaking adults. The same speaker of Canadian French, who was interviewed 24 years apart, increased his ne-realization from 0.5% at age 22 to 4.5% at age 45. Sankoff (2019a) noted that, while speakers apparently internalize the probabilistic nature of *ne* as children, such protracted age dependence might indicate an evolving understanding of the social meaning of the variant, and/or a speaker refraining from using a variant until they have reached the appropriate age and

status. While we did not attempt to code the social context of each negative utterance in our sample, it is reasonable to assume that children have fewer stylistically appropriate occasions to employ the standard variant. For example, while parents have many occasions to teach or discipline their children — a social context that invites the more formal *ne* — children likely have many fewer such opportunities. Indeed, a closer examination of children's mastery of the social constraints on ne-realization is called for. In future work, we plan to specifically analyze the topic and context (e.g. school vs play vs discipline etc.) of children's negative utterances to determine whether children control the social constraints on *ne* and from what age.

Do children match the subject constraints on ne-retention?

Yes, our results suggest that children not only produce *ne* variably, but also condition their use of *ne* on the proceeding subject. Children in our sample obeyed this subject constraint in exactly the same order reported for interadult speech (a pattern we also observed in caregivers' negative utterances). Further, we found no interaction with child age, suggesting that children control this linguistic constraint from their very earliest *ne* productions. While this finding is not surprising in some ways — several developmental sociolinguists have argued that children master grammatical constraints quite early (Smith et al., 2007, 2013) — it is unexpected in others. First, until now, *ne* was so rarely reported in children's own negative utterances that such an analysis was not possible. Our results offer the first evidence that children follow the grammatical constraints on ne-retention from a young age. Second, the composition of children's early negative utterances is quite different from that of adults, so one might expect ne-retention patterns to reflect this difference. Our analysis confirms that, though children do not develop an adult-like pattern of negative sentence production until they are older (see Figure 5), their use of variable *ne* follows the subject constraint from their earliest productions.

While we have provided a quantitative analysis of the nature of "ne" in French children and their caregivers, we have not addressed whether the development of *ne* production is an

accurate reflection of French-acquiring children's mental grammar. Our analysis does not provide an answer, and instead leaves open several such questions. Specifically, do children first develop a deterministic grammar, either fully omitting or fully incorporating the standard form, before they begin to express variable "ne" in their own productions? Or do they develop a probabilistic grammar from the very start? In either case, how does their acquisition of variable "ne" interact with their developing understanding of negative concord? For now, we leave these questions aside to explore in future research.

References

Agren, J. (1973). *Étude sur quelques liaisons facultatives dans le français de la conversation radiophonique.* Uppsala University Press.

Armstrong, N. (2002). Variable deletion of French ne: a cross-stylistic perspective. *Language Sciences,* 24(1), 153-73.

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.

Ashby, W. (1981). The loss of the negative particle ne in French: a syntactic change in progress. *Language*, 57, 674-687.

Ashby, W. (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.

Bassano, D. & Mendes-Maillochon, I. (1994). Early grammatical and prosodic marking of utterance modality in French: a longitudinal case study. *Journal of child Language*, 21(3), 649-675.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Champaud, C. (1994). *The development of verb forms in French children at around two years of age: some comparisons with Romance and non-Romance languages.* Paper presented at the First Lisbon Meeting on Child Language, Lisbon, Portugal.

Choi, S. (1986). *A cross-linguistic development study of negation in English, French and Korean* (Doctoral dissertation). State University of New York at Buffalo.

Coveney, A. (1996). *Variability in spoken French. A sociolinguistic study of interrogation and negation.* Exeter: Elm Bank.

Culbertson, J. (2010). Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language*, 86(1), 85-132.

Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (Eds.), *The expression of negation* (pp. 39-72). De Gruyter Mouton.

Demuth, K. & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, 35, 99-127.

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word*, 14(1), 47-56.

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 81, 177–206.

Hirschbühler, P. & Labelle, M. (2004). Proclisis and enclisis of object pronouns at the turn of the 17th century: the speech of the future Louis XIIIth. In R. Gess & D. Arteaga (Eds.), *Historical Romance linguistics: retrospective and perspectives* (pp. 187-208). John Benjamins, Amsterdam.

Kluyver, T., et al. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Scmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87-90). IOS Press.

Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1(1), 85-97.

MacWhinney, B. (2000). *The CHILDES project: The database (Vol. 2).* Psychology Press.

Marrero, V. & Aguirre, C. (2003). Plural acquisition and development in Spanish. In *Linguistic Theory and Language Development in Hispanic Languages: Papers from the 5th Hispanic Linguistics Symposium (Vol.2)* (pp. 275-296). Somerville: Cascadilla Press.

Martineau, F. & Mougeon, R. (2003). A sociolinguistic study of the origins of ne deletion in European and Quebec French. *Language*, 118-152.

Miller, K. & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition*, 19(3), 223-261.

Miller, K. (2013). Acquisition of variable rules:/s/-lenition in the speech of Chilean Spanish-speaking children and their caregivers. *Language Variation and Change,* 25(3), 311-340.

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

Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua*, 161, 125-143.

Plunkett, B. (2002). Null Subjects in child French interrogatives: A view from the York Corpus. *Romance corpus linguistics: Corpora and spoken language*, 441-452.

Pooley, T. (1996). *Chtimi: the Urban Vernaculars of Northern France.* Clevedon: Multilingual Matters.

Purcell, A. K. (1984). Code shifting Hawaiian style: Children's accommodation as a decreolizing continuum. *International Journal of the Sociology of Language*, 46, 71–86.

R Core Team. (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria.

Roberts, J. L. (1994). *Acquisition of variable rules: (-t, d) deletion and (ing) production in preschool children* (Doctoral dissertation). University of Pennsylvania.

Roberts, J. L. (1997). Hitting a moving target: Acquisition of sound change in progress by Philadelphia children. *Language Variation and Change*, 9, 249–66.

Roberts, J. L. (2002). Child language variation. In J. Chambers, N. Schilling-Estes & P. Trudgill (Eds.), *Handbook of language variation & change*. Oxford: Blackwell.

Romaine, S. (1978). Postvocalic /r/. In *Scottish English: Sound change in progress* (pp. 144-157).

Sankoff, G. & Vincent, D. (1980). The productive use of ne in spoken Montréal French. In *The social life of language* (pp. 295-310). University of Pennsylvania Press.

Sankoff, G. (2019a). Age grading, style, and language change: a lifespan perspective. In *The Routledge companion to the work of John R. Rickford* (pp. 308-315). Routledge.

Sankoff, G. (2019b). Transmission revisited. In D. Lightfoot & J. Havenhill (Eds.), *Variable properties in language: Their nature and acquisition* (pp. 155-166).

Shin, N. (2016). Acquiring constraints on morphosyntactic variation: children's Spanish subject pronoun expression. *Journal of Child Language*, 43, 914–947

Smith, J., Durham, M., & Fortune, L. (2007). "Mam, ma troosers is fa'in doon!" Community, caregiver and child in the acquisition of variation in Scottish dialect. *Language Variation and Change*, 19(1), 63–99.

Smith, J., Durham, M., & Richards, H. (2013). The social and linguistic in the acquisition of sociolinguistic norms: caregivers, children, and variation. *Linguistics and philosophy,* 51(2), 285–324.

Suppes, P., Smith, R. & Leveillé, M. (1973). The French syntax of a child's noun phrases. *Archives de Psychologie*, 42, 207–269.

Wickham, H., et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

Yamaguchi, N. (2012). *Parcours d'acquisition des sons du langage chez deux enfants francophones* (Doctoral dissertation). Université Sorbonne Nouvelle Paris 3.

Footnotes

1. This French sentence would be ambiguous – it could also mean ‘I want more cherries.’
2. As suggested by Culbertson (2010), the wide range of ne-retention reported from these corpora is likely explained by regional differences and the composition of the corpus under investigation. The Coveney (1996) corpus, for example, involved conversations between two strangers — a context in which speakers are more likely to use the standard form — while Armstrong (2002) also included conversations between friends.
3. Choi 1986 analyzed the caregiver speech of two French mothers and one French-Canadian mother, but they were not analyzed separately.
4. Note that the child model failed to converge when including random slopes for child age and subject, so we simplified the random effects structure in the child model to include only random by-child intercepts
5. We are aware that the first datapoint at 11 months for caregivers is very high. However, it is within 3 standard deviations of the mean ne-retention rate and therefore not excluded.