Language Learning and Development

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hlld20

Acquisition of a rare variant: ne-realization in the negative utterances of French children and their caregivers

Yiran Chen & Kathryn Schuler

To cite this article: Yiran Chen & Kathryn Schuler (2025) Acquisition of a rare variant: ne-realization in the negative utterances of French children and their caregivers, Language Learning and Development, 21:2, 173-192, DOI: 10.1080/15475441.2024.2380686

To link to this article: https://doi.org/10.1080/15475441.2024.2380686

Published online: 28 Aug 2024.

LANGUAGE LEARNING AND DEVELOPMENT
2025, VOL. 21, NO. 2, 173–192
https://doi.org/10.1080/15475441.2024.2380686

Routledge
Taylor & Francis Group

Acquisition of a rare variant: ne-realization in the negative utterances of French children and their caregivers

Yiran Chen and Kathryn Schuler

Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

A growing body of evidence suggests children acquire linguistic variation and follow grammatical constraints on its use from a young age. While prior research predominantly focuses on variables for which both variants are well-attested, stable variation in natural language often includes variants with an unbalanced distribution. In many cases, one variant is extremely rare, which poses a learning challenge for children and may require revision of accounts of children's acquisition of variation. The negative morpheme ne in French is one such rare variant as it is only rarely realized (<10%) in adult 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-realization in the conversational speech of 14 monolingual French-learning children (ages 0.10–8.01) and their caregivers. We found that, for both French children and their caregivers, ne-realization was rare and followed the same linguistic constraint found in interadult speech. Importantly, children showed probabilistic production following the constraint extremely early on – at age 2.05 on average. Further, we found a strong influence of input on variation acquisition: caregivers with higher ne-realization rates had children who used more *ne* and from a younger age. Taken together, our results suggested that young children can acquire statistically rare but systematically patterned variants, adding further evidence that they are skilled and sophisticated language learners.

Introduction

In the past half century, the field of sociolinguistics has made significant progress characterizing how speakers acquire, represent and use stable variation in the grammar. Across many languages, developmental sociolinguists have shown that children have no trouble acquiring variable forms, producing variants with the same patterns of use, and governed by the same grammatical and social constraints, that are observed in adult speakers (Labov, 1989; Miller, 2013; Nardy et al., 2014; Roberts, 1994, 1997, 2002; Shin, 2016; Smith et al., 2007, 2013 among others). However, while many different variables have been examined, the majority of studies have focused on variables that have relatively balanced distribution, or variables for which both variants are well-attested (e.g., 60% v 40%). For example, one of the most widely studied variables in English is T/D deletion (Guy, 1980; Labov et al., 1968; Wolfram, 1969). While word-final T/D is more likely to be deleted in some contexts compared to

CONTACT Yiran Chen cheny39@sas.upenn.edu University of Wisconsin–Madison, Waisman Center, 1500 Highland Ave, Room 535, Madison, WI 53705, USA

others (e.g., in monomorphemic words, 84.56% when followed by a consonant, 42.98% when followed by a vowel, Wolfram, 1969, p. 62), the baseline rate of deletion ranges from around 45% to 65% in colloquial speech in different varieties of North American English (e.g., Hazen, 2011; MacKenzie & Tamminga, 2021; Wolfram, 1969).

While stable variation often involves variation between two frequent forms, there are several cases of (apparently stable) variation in which one form is rare. For instance, in modern spoken French, the negative morpheme *ne* is rarely realized, with realization rates estimated around 10% (e.g., Culbertson, 2010) or lower in some dialects (below 1% in Canadian French) (e.g., Sankoff & Vincent, 1980). Similarly, in Buckie Scottish English, the standard pronunciation of "ou" – pronouncing the word "house" as *house* instead of *hoose* – occurs under 1% of the time in inter-adult speech (Smith et al., 2007). Importantly, despite their infrequent realization, these forms are not becoming any more infrequent or obsolete (e.g., Poplack & St-Amand, 2007). Rather, they seem to be part of stable variation in the language that learners acquire just as robustly as more frequent forms. A question that remains unknown is whether the distribution of variants affects how the variable is acquired. This question is particularly important in understanding the acquisition of sociolinguistic variation, given children's tendency to *regularize* their language input, especially when it contains probabilistic alternations of multiple forms (e.g., Hudson Kam & Newport, 2005). Importantly, one predominant pattern of such regularization is that children often regularize to the statistically dominant form. How, then, does an extremely rare variant get learned?

Acquisition of variation

One of the clearest demonstrations of children's regularization behavior in natural language acquisition is the detailed documentation of the input and production of ASL-acquiring children born to parents who were non-native signers (Ross, 2001; Ross & Newport, 1996; Singleton & Newport, 2004). These children used ASL with much greater consistency than their parents by boosting the frequency of the most frequent forms of their parents' language (Singleton & Newport, 2004). Children's regularization behavior is also widely observed in well-controlled experimental settings (e.g., Hudson Kam & Newport, 2005). When presented with an artificial language in which a determiner was present 60% of the time and absent 40% of the time, adults matched this distribution in their production, but most children regularized, either always using the determiner or always producing the noun without the determiner (Hudson Kam & Newport, 2005). In this and other similar artificial language learning experiments, using the dominant form deterministically is the most common regularization pattern (e.g., Austin et al., 2022; Hudson Kam & Newport, 2005). According to the regularization literature, learning stable variation where one variant is rare could be more difficult because children will have a stronger tendency to regularize to the dominant form during acquisition, circumventing the acquisition of the rare form.

While developmental sociolinguists have provided substantial evidence that children are able to produce variable forms early on, very few explicit proposals have been developed to explain *how* children acquire them. Many have underscored that quantity of input is an important factor, given that the more language input children get, the sooner they acquire the variable pattern (Hendricks et al., 2018; Shin, 2016). Others have commented that the more complex a pattern is, the longer it takes to acquire the variable (Johnson & White, 2020; Shin, 2016). To the best of our knowledge, only one formal proposal has been made. Shin and Miller (2022) proposed that, when learning a variable, children first produce only one variant (due to regularization), then they produce both variants (but only in mutually exclusive contexts), and finally they produce both variants in overlapping contexts. According to this view, for rare variants, we should expect to see an initial period where only the dominant form is attested (due to particularly strong regularization tendency) and the rare form should emerge later and (at least initially) occur in only certain contexts.

Table 1. Examples of French negative sentences.

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

Taken together, prior research on children's acquisition of variation seems to suggest that acquisition of a rare form could pose a challenge for children. We might expect a rare form to be acquired later (due to its infrequency in the input) or to be regularized out of children's productions, either temporarily for an initial period as predicted by Shin and Miller (2022), or for an extended period as predicted by the literature on children's regularization behavior (e.g. Hudson Kam & Newport, 2005). On the other hand, rare forms may be attested more often in child-directed speech than in adult-to-adult colloquial speech, since caregivers have been found to boost the use of an infrequent variants – though often the standard form – when they are speaking to their children (Foulkes et al., 2005; Smith et al., 2007).

In the present work, we zoomed in on one variable, ne-realization in colloquial French, to investigate how a rare variant is learned. While *ne* is only realized around 10% of the time, it is still in stable variation in French adults' speech, following systematic linguistic-internal and external constraints. We examined *ne*'s occurrence in both children-directed speech and children's own production to provide a quantitative analysis of the acquisition pattern of *ne*, a rare form of a stable variation.

Ne-realization 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).

The variable realization of *ne* in colloquial French is thought to have originated gradually, sometime in the 17th Century (Hirschbühler & Labelle, 2004; Martineau & Mougeon, 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-realization rates ranging widely from 36.7% to less than 1% in adult speakers[2] (see Table 2).

Table 2. Ne-realization rate in colloquial French by publication and region.

Publication	Region	ne-realization
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 and Vincent (1980)	Canada	<1%

[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-realization 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.

While being rarely attested, ne-realization in colloquial French appears to be conditioned by several factors both internal and external to the language. Among language-internal constraints, ne-realization 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; Marrero & Aguirre, 2003). Speakers are more likely to omit *ne* when the post-verbal negative morpheme is *pas* compared to *jamais, plus, rien*, or *personne* (Armstrong & Smith, 2002). Among language-external constraints, ne-realization 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-realization in French-acquiring children and their caregivers

While a great deal is known about the patterns of ne-realization in interadult colloquial French, less attention has been devoted to understanding how children acquire and use this particular sociolinguistic variable in French. Does children's input reflect the same ne-realization patterns we observe in speech between French-speaking adults? Given that rates of ne-realization are low in adult-to-adult colloquial French (as discussed in section 1.1), one might expect ne-realization rates to be similarly low in child-directed speech. However, it is possible that the rate of ne-realization 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 their use of standard forms in their child-directed speech (e.g. Foulkes et al., 2005; Roberts, 2002; Smith et al., 2007, 2013), particularly when children are young, likely as a way to facilitate learning.

So far, only a few studies have investigated ne-realization 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 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-realization 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 increase 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-realization in child directed speech, we also want to understand whether children's input obeys the same constraints on ne-realization that we observed in adult-to-adult speech. Recall that, for French-speaking adults, ne-realization 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 a sample of 5 mothers followed the subject constraint in their child-directed

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

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 (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.

Finally, and crucially, we also want to understand 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 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 it was not possible to quantify whether ne-realization 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-realization in children and their language input, many aspects of *ne* acquisition remain unclear. In the present study, we aim to investigate how *ne* is acquired, as a window into how rare variants are learned more generally. 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 24,222 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 to facilitate learning (either in general or dependent on their child's age or gender), and 3) whether caregivers follow the robustly attested subject constraint on ne-realization in their child-directed speech, such that this constraint is indeed available to child learners. Our second aim is to provide a quantitative analysis of the acquisition pattern of *ne* by children specifically asking 4) whether and, if so, 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-realization 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 et al., 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

Table 3. Corpora analyzed.

Corpus	Age Range	Region	Recording Year	N Children	Mean Recordings per Child
Leveille	2;01–3.03	Paris	1971–1972	1	33
Champaud	1;09–3.05	Paris	1988–1989	1	32
York (Anne)	1;10–3.05	Paris	1997–1998	1	35
Lyon	0;11–4.00	Lyon	2002–2005	4	83.75
Paris	0;10–8.01	Paris	2005–2008	6	30.16
Yamaguchi	1;03–4.03	Paris	2006–2009	1	31
Total				**14**	

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.

Procedure

To extract the negative utterances from the transcript of 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-realization 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 24,222 negative utterances for corpus region, corpus decade, speaker id, speaker type (caregiver or child), child id, child age, child gender, subject type (NP, clitic pronoun, non-clitic pronoun, or null subject), and whether *ne* was realized or not. Among the 24,222 negative utterances, 6887 were uttered by children. To confirm that children's utterances were a reflection of their own internal grammar, we investigated whether any of the children's negative utterances were direct imitations of a caregiver's immediately preceding utterance. Out of the 6887 negative sentences uttered by children, there were only 14 cases of such direct

imitation (0.2%). Since this is a negligibly small number, we included these utterances in our reported analysis. But note that running our model with these cases removed does not change any of the reported results.

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., 2019).

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 gender, 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-realization rates are decreasing in France and may differ by region (Armstrong & Smith, 2002; Ashby, 1981), 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 gender + corpus region + corpus decade + (0 + child age + subject type || child name)*
(6) Child model: *ne ~ child age * subject type + child gender + corpus region + corpus decade + (1|child name)*

Finally, to test whether a caregiver's rate of ne-realization 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-realization 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-realization rate (log transformed) to predict the child's average ne-realization rate (log transformed). No additional predictors were included in either of these models.[5]

Results

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.

Region and decade of recording

Consistent with reports of regional differences in *ne* usage in interadult speech (Armstrong & Smith, 2002; Ashby, 2001; 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: $\beta = 0.672$, SE = 0.336,

[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]An earlier version of this work appears in the conference proceedings for BUCLD 47 (Chen & Schuler, 2023).

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

	Caregiver model (n=17335)				Child model (n=6887)			
Fixed effects	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 gender - 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

p = .045; Child: β = 0.850, SE = 0.344, p = .013). In contrast, despite reports of the rapid decline of *ne* usage in interadult speech in recent decades (Armstrong & Smith, 2002; Ashby, 1981, 2001), 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 = .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

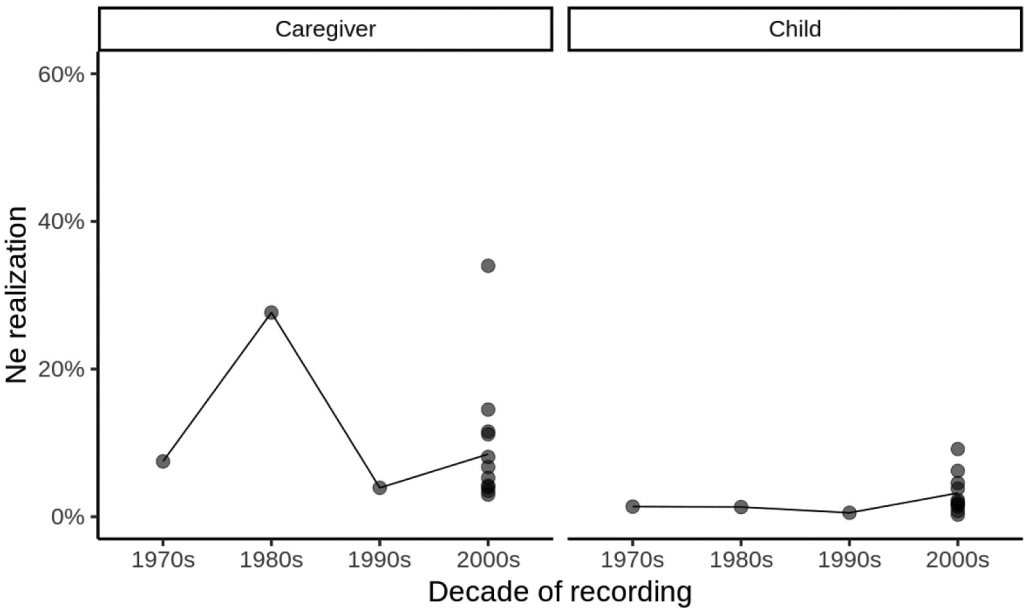

Figure 1. Mean ne-realization by decade of recording for caregivers (left) and children (right). Each dot is the mean rate of ne-realization for a given caregiver (or child) in our sample.

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.

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), p. 3 mothers, 8%; Culbertson (2010), p. 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-realization for all children and their caregivers in our sample, arranged by corpus and caregiver realization rate. For caregivers, we found the average rate of ne-realization 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, $p < .001$). 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.

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,

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

			Age of first		Child ne		Caregivers ne	
Child	Gender	Age range	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**	**188/6887**	**2.73%**	**1371/18335**	**8.49%**

indicating that log odds of ne-realization were significantly lower than chance (50%) in children's productions (β=-4.403, SE = 0.446, $p < .001$). 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-realization was furthest from his caregivers, differing by 26.35%.

Acquisition of variable ne by age and gender

Beyond the average ne-realization rate in children and their caregivers, we also asked whether children's acquisition of *ne* differed by child age or gender. While previous reports of children's ne-realization has been limited – neither Sankoff (2019b) nor Palasis (2015) had sufficient data to analyze the developmental trajectory of *ne* in children – researchers have found age- and gender-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 with *pas* 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). And without exception, every child first categorically produced negative sentences without *ne*. As shown in Table 5, every child's first recorded negative sentence (Age of first neg) is earlier than their first

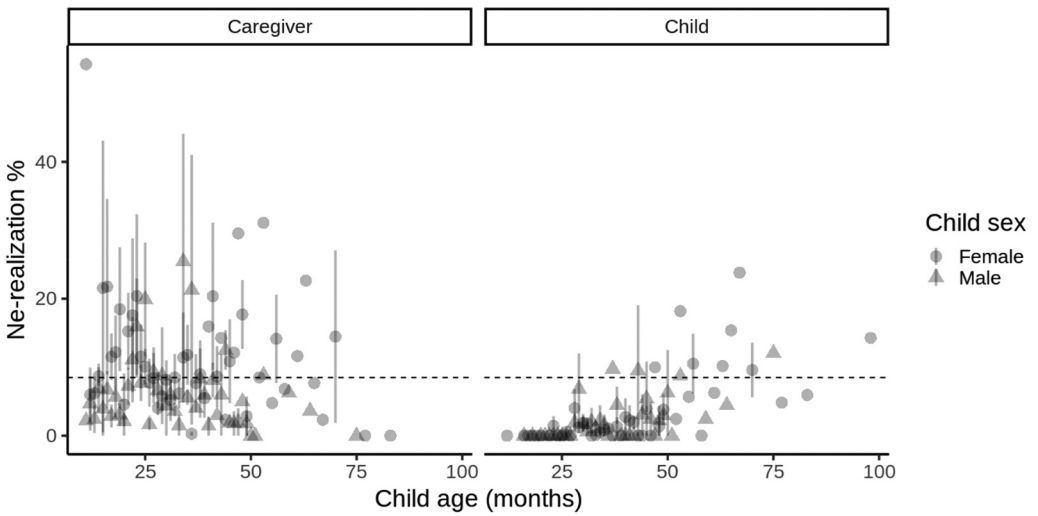

Figure 2. Average ne-realization rate for caregivers and children by child age in months and child gender. Error bars reflect standard error and the dashed line reflects mean ne-realization rate for caregivers (i.e. The average ne-realization in children's input).[6]

[6]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-realization rate and therefore not excluded.

recorded use of *ne* (Age of first ne). 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).

Our results further suggest that once children start to produce *ne*, ne-realization is age-dependent. In our child model, child age is a significant predictor of ne-realization ($\beta = 0.621$, SE = 0.102, $p < .001$). As can be seen in Figure 2, *ne* is unattested in the youngest children, but approaches the adult level of ne-realization as the children grow older. Gender, on the other hand, was not a significant predictor of ne-realization in our child model. Male children are no less likely to realize *ne* than their female peers (β=-0.241, SE = 0.260, $p = .354$).

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 studies have demonstrated that caregivers tend to increase 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, $p = .521$), 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: gender-dependence. Recall from section 3.3 that we did not find children's ne-realization to differ by gender. However, for many sociolinguistic variables, researchers have found

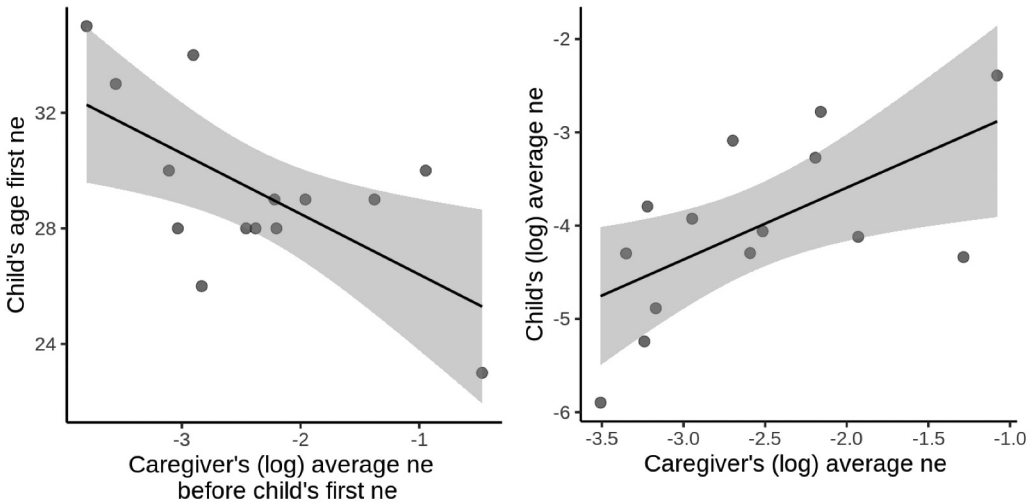

Figure 3. Caregiver's average ne-realization before their child's first attestation (log transformed) predicts the age at which their child produced *ne* (left). Caregiver's average ne-realization rate (log transformed) predicts their child's average ne-realization rate (log transformed).

that the input itself differs by child gender, 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 gender was not a significant predictor of ne-realization in caregivers (β=-0.436, SE = 0.346, p = .204; Table 4), suggesting that caregivers do not use more *ne* with girls than boys.

Finally, while we did not find any age- or gender-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-realization *before* that age (log transformed). As shown in Figure 3 (left), we found that caregivers with the highest ne-realization rate had children who produced *ne* the earliest (β=-2.100, SE = 0.728, p = .014). We also built a second model to predict children's average ne-realization (log transformed) by their caregivers' average ne-realization (log transformed). As shown in Figure 3 (right), we found that caregivers with the highest ne-realization had children who produced more *ne* overall (β = 0.771, SE = 0.285, p = .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 in our sample did not tailor their ne-realization rate to their child's age or gender.

Acquiring the linguistic constraints on variable ne

A final question to consider is whether children and their caregivers follow the linguistic constraints on ne-realization. Recall that we hypothesized that caregivers (and therefore the child's input) would likely obey the linguistic constraints on ne-realization, 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 < .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 < .001), which suggests that the slope of

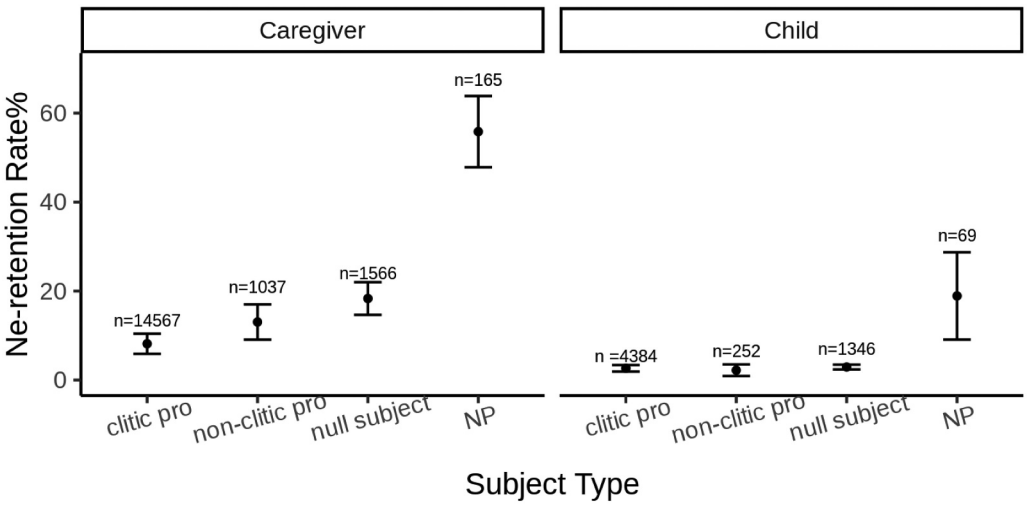

Figure 4. Average ne-realization rate in "verb+pas" sentences by subject type for caregivers and children. (for children, only data after the first attestation of *ne* is included.).

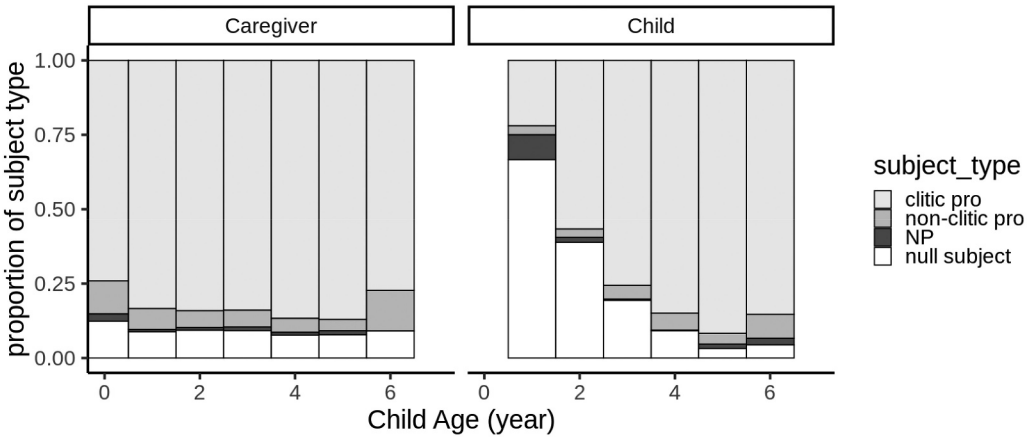

Figure 5. Proportion of each subject type in all negative utterances across children's age.

the trend is also increasing. This reflects what we see in Figure 4 (left): caregiver's ne-realization follows a linear increase from clitic pro to null subject, with a dramatic increase in slope going from null subject to full NP.

Our results suggest that the linguistic constraints on ne-realization 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 ($\beta = 1.397$, SE = 0.279, $p < .001$), with a significant quadratic term ($\beta = 0.640$, SE = 0.292, $p = .028$). As shown in Figure 4 (right), children's ne-realization by Subject type follows a similar trajectory as adults, increasing linearly from pronoun subjects to null subject 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). This null effect could mean that children follow the linguistic constraints on ne-realization 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). However, it is also possible that how children's production of *ne* is modulated by subject type *does* change across age, but our current approach is unable to detect it.

Discussion

The goal of this paper was to investigate the use of a rarely attested variable, *ne*, in the negative utterances of children and their caregivers as a window into the acquisition patterns of these rare variants that participate in systematic linguistic variation. 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); Berit 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-realization rates in child-directed speech (8% and 7.6%, respectively), many previous studies have observed that caregivers increase their use of standard variants in speech to their children (e.g. Foulkes et al., 2005; Smith et al., 2007). Indeed, Sankoff (2019b) 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-realization 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 increase their use of the standard form in speech to their children?

Researchers have hypothesized that parents may facilitate learning of rare variants by boosting their use of these forms in their speech to young children. As reported above, we did not find evidence that caregivers increase their use of *ne* in child-directed speech. One possible explanation is that, unlike other variables where parents' boosting of the standard form is striking (e.g., Smith et al., 2007), *ne* omission may be less salient or stigmatized, and therefore parents do not engage in conscious or unconscious boosting. However, it could also be the case that parents' use of *ne* does differ between interadult and child-directed speech, but our analysis could not detect this difference. Importantly, 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 boosted their ne-realization rates based on their child's age or gender. For other sociolinguistic variables, research suggests that parents increase 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), again to facilitate learning. While Sankoff (2019a) 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 gender reliably predicted caregivers' use of *ne* in our sample. One explanation for this difference could be regional: Olivier and his father spoke Canadian French, a dialect for which ne-realization 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-realization in their child-directed speech?

A prerequisite to children's acquisition of grammatical conditioning factors is that these constraints must be available in speech to young children. We find that, yes, like Culbertson (2010), 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 realization pattern follows the same ordering as what has been reported for interadult speech (Agren, 1973; Armstrong & Smith, 2002; Coveney, 1996, Hanson & Malderez, 2004; Martineau & Mougeon, 2003), confirming that the subject constraints on ne-realization 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 the omission of *ne* (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 the omission of *ne* — it is possible that we have underestimated the true rate of *ne* realization in both children and their caregivers. Still, our analysis offers an important first step in characterizing the nature of *ne* in child-directed speech: though ne-realization is rare in child-directed speech, the subject constraint is available in speech to young children.

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

Our results for child-directed speech seem to suggest that *ne* is just as rare in children's input as in adult-to-adult speech, and caregivers in our sample do not appear to boost their ne-realization rates when children are young (or by gender) to help children acquire this rare form: the *ne* variant is truly rare in child-directed speech. Recall that previous research seems to suggest that rare forms might take longer for children to acquire (because they are infrequent and present fewer learning opportunities) or pass through a period in which they are unattested in children's own productions (due to their regularization tendencies). In line with these predictions, we found children's *ne*-realization is 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*, an optional form, early on, which is consistent with what's previously found with other variables (e.g., Chevrot et al., 2011; Nardy et al., 2014). Importantly, as shown in Figure 2, *ne* is also very 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-realization?

Our results suggest that children's production of *ne* varies depending on the syntactic category of the proceeding subject. Children in our sample obeyed the subject constraint reported for interadult speech (a pattern we also observed in caregivers' negative utterances). This pattern likely reflects children's own linguistic knowledge because it holds true after excluding negative utterances that are direct imitations of caregiver's preceding utterance, although we can't definitively rule out the possibility that priming from utterances in the conversation context that were not immediately preceding might play a role. However, the current analysis does not allow us to firmly conclude whether the learning pattern observed here is in fact morphosyntactic in nature, assumed by prior work on "subject constraint," or driven by specific lexical restrictions. In future work, we plan to examine the lexical items preceding and following children's *ne* to see whether they have truly generalized a constraint on use of *ne* across lexical items. If we observe sufficient variation in the preceding subjects or the following verbs that co-occur with *ne*, we will have more evidence that children have mastered a morphosyntactic constraint on the variation.

Further, we found no interaction between subject type and child age. This could mean that children follow this constraint in similar ways throughout the age range where they produce *ne* examined here (from 2.5 to 8 years). While this would not be surprising in some ways – several developmental sociolinguists have argued that children master grammatical constraints quite early (Smith et al., 2007, 2013) – it would be unexpected in others. First, until now, *ne* was so rarely reported in children's own negative utterances that such an analysis was not possible. Second, the composition of children's early negative utterances is quite different from that of adults, so one might expect ne-realization patterns to reflect this difference. One possibility is 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. However, caution needs to be exercised when interpreting null results. It is also possible that there are in fact age-dependent changes in how children's *ne* production varies with the preceding subject, but our current analysis was unable to detect them. One reason our model may have missed age-related changes in the subject-type constraint is that there are fewer utterances from children older than age 4 in the current dataset (see Table 5). In future work, we plan to examine in further detail the developmental trajectory of each individual child's *ne* productions in their linguistic contexts, to complement the current group-level analysis.

What implications does children's ne-realization have on the acquisition of morphosyntactic variation?

Our quantitative analysis of the nature of ne-realization in French children yields results consistent with many other documented cases of children's acquisition of morphosyntactic variables (e.g., Carvalho et al., 2015; Grinstead, 2004; Newkirk-Turner & Green, 2016; Smith et al., 2013) as well as the 4-step pathway summarized in Shin and Miller's (2022). Shin and Miller (2022) stated that, when learning a variable, children first produce only one of the two forms (Step 1), then produce two forms only in mutually exclusive contexts (Step 2), then produce both forms in some overlapping contexts (Step 3) and finally gradually expand the overlapping contexts (Step 4). In our case of ne-realization, it is true that, before around age 2.05, children never realized *ne*, which could be interpreted as using only one variant: the null form of the variable. As Shin and Miller (2022) have pointed out, this one-variant stage can be accounted for by children' s tendency to regularize inconsistency in input (e.g., Hudson Kam & Newport, 2005). Further, compared to their parents, children showed a more extreme dichotomy of using ne in sentences with NP subjects and omitting ne when the subject is not full NP. This could be an example of children transitioning from Step 2 to Step 3 in Shin and Miller's (2022) terminology: first using the second variant in only one (or at least more limited) contexts before moving on to the more adult-like expression of the variant.

Our results have potential to further refine Shin and Miller's (2022) model, which is a future direction we are keen to pursue. The first avenue is that, with further analysis of our data at an individual level, we can directly examine whether there is truly a stage where the use of *ne* and null are in mutually exclusive contexts – more specifically, whether the initial use of ne is exclusively in the NP subject context. If this turns out to be true, it would provide strong evidence for Shin & Miller's Stage 2. The second avenue is clarifying whether non-production can indeed can be characterized as "regularizing to null." Because one of the variants at hand is "null" – as are many cases where children's acquisitions of morphosyntactic variables are examined (e.g., Carvalho et al., 2015; Grinstead, 2004; Newkirk-Turner & Green, 2016; Smith et al., 2013) – there are a few possible explanations for children's initial categorical omission other than regularization. Specifically, do children first develop a deterministic grammar fully omitting 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, but the overt form is just not reflected in production? This point is especially relevant for morphosyntactic variation in French because many scholars have raised the possibility of French grammar being diglossic (Rowlett, 2013), meaning that children's early grammar may be a deterministic one without *ne*. Alternatively, children may have the probabilistic grammar in place earlier, but did not produce *ne* earlier due to the non-trivial demand of producing utterances that are more than two words long (Brown, 1973) at 2 years of age. The importance of differentiating the two possibilities arising from the current investigation converges with Hudson Kam's (2024) for differentiating "two kinds of non-producers" (p.g., 74) in the context of regularizing variation where one variant is null – ones who understand the grammatical function that the null variant serves and the ones who don't. One potential way of disentangling the two possibilities in our specific case of *ne* is to test whether children comprehend "ne … pas" negation (concord negation) before they produce *ne*. That way, we will be able to determine – for the children who categorically omit *ne* — whether their lack of production reflects their lack of knowledge, which would suggest that they have not yet entered Shin and Miller's (2022) Step 1 (c.f., Shin & Miller, 2024), or whether they indeed understand *ne* as a negative morpheme but only refrain from producing it due to regularization or other pressure. For now, we leave these questions aside to explore in future research.

Conclusion

In sum, a growing body of evidence suggests children acquire linguistic variation and follow linguistic constraints on its use from a young age. While prior research predominantly focuses on variables for which both variants are well-attested, stable variation in natural language often includes variants with an unbalanced distribution. In many cases one variant is extremely rare, which poses a learning challenge for children and may require revision of accounts of children's acquisition of variation. Here, we find that children are able to acquire one such rare variant, despite being realized around 8% in their input. Further, their production of the variant is conditioned by the same linguistic constraint as adult speakers, potentially from a young age, adding further evidence that young children are skilled and sophisticated language learners.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Funding for this work was provided by the University of Pennsylvania to K. Schuler.

ORCID

Yiran Chen http://orcid.org/0000-0002-0695-3940
Kathryn Schuler http://orcid.org/0000-0003-2962-731X

Data availability statement

The data that support the findings of this study are openly available in an OSF repository at https://osf.io/mjtbz/.

Open scholarship

This article has earned the Center for Open Science badge for Open Data. The data are openly accessible at https://do i.org/10.1080/15475441.2024.2380686

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), 153–173. https://doi.org/10.1016/S0388-0001(01)00015-8
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. (1981). The loss of the negative particle ne in French: A syntactic change in progress. *Language*, *57*(3), 674–687. https://doi.org/10.2307/414345
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. https://doi.org/10.1017/S0959269501000114
Austin, A. C., Schuler, K. D., Furlong, S., & Newport, E. L. (2022). Learning a language from inconsistent input: Regularization in child and adult learners. *Language Learning and Development*, *18*(3), 249–277. https://doi.org/10.1080/15475441.2021.1954927
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. https://doi.org/10.1017/S0305000900009491
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01
Berit Hansen, A., & Malderez, I. (2004). Le ne de négation en région parisienne: une étude en temps réel. *Langage et Société*, *107*, 5–30. https://doi.org/10.3917/ls.107.0005
Brown, R. (1973). *A first language: The early stages*. Harvard University Press.
Carvalho, A. M., Orozco, R., & Shin, N. (Eds.). (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Georgetown University Press.
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.
Chen, Y., & Schuler, K. (2023). Variable ne in the negative utterances of French children and their caregivers. In P. Gappmayr & J. Kellogg (Eds.), *Proceedings of the 47th annual Boston University conference on language development* (pp. 104–117). Cascadilla Press.
Chevrot, J. P., Nardy, A., & Barbu, S. (2011). Developmental dynamics of SES-related differences in children's production of obligatory and variable phonological alternations. *Language Sciences*, *33*(1), 180–191. https://doi.org/10.1016/j.langsci.2010.08.007
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*. Elm Bank.
Culbertson, J. (2010). Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language*, *86*(1), 85–132. https://doi.org/10.1353/lan.0.0183
Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, *35*, 99–127. https://doi.org/10.1017/S0305000907008276

Dimroth, C. (2010). The acquisition of negation. In L. R. Horn (Ed.), *The expression of negation* (pp. 39–72). De Gruyter Mouton.
Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word*, *14*(1), 47–56. https://doi.org/10.1080/00437956.1958.11659655
Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, *81*(1), 177–206. https://doi.org/10.1353/lan.2005.0018
Grinstead, J. (2004). Subjects and interface delay in child Spanish and Catalan. *Language*, *80*(1), 40–72. https://doi.org/10.1353/lan.2004.0024
Guy, G. R. (1980). Variation in the group and the individual: The case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1–36). Academic Press.
Hazen, K. (2011). Flying high above the social radar: Coronal stop deletion in modern Appalachia. *Language Variation and Change*, *23*(1), 105–137. https://doi.org/10.1017/S0954394510000220
Hendricks, A. E., Miller, K., & Jackson, C. N. (2018). Regularizing unpredictable variation: Evidence from a natural language setting. *Language Learning and Development*, *14*(1), 42–60. https://doi.org/10.1080/15475441.2017.1340842
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.
Hudson Kam, C. L. (2024). Thinking about shin & Miller (2022)'s step 1. *Language Learning and Development*, *20*(1), 73–75. https://doi.org/10.1080/15475441.2023.2239793
Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195. https://doi.org/10.1080/15475441.2005.9684215
Johnson, E. K., & White, K. S. (2020). Developmental sociolinguistics: Children's acquisition of language variation. *Wiley Interdisciplinary Reviews Cognitive Science*, *11*(1), e1515. https://doi.org/10.1002/wcs.1515
Kluyver, T., (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. https://doi.org/10.1017/S0954394500000120
Labov, W., Cohen, P., Robins, C., & Lewis, J. (1968). *A study of the non-standard english of negro and puerto rican speakers in New York city. Volume i: phonological and grammatical analysis*. Columbia University.
MacKenzie, L., & Tamminga, M. (2021). New and old puzzles in the morphological conditioning of coronal stop deletion. *Language Variation and Change*, *33*(2), 217–244. https://doi.org/10.1017/S0954394521000119
MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.
Marrero, V., & Aguirre, C. (2003). Plural acquisition and development in Spanish. In S. Montrul & F. Ordóñez (Eds.), *Linguistic theory and language development in Hispanic languages: Papers from the 5th Hispanic linguistics symposium (Vol.2)* (pp. 275–296). Cascadilla Press.
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://doi.org/10.1353/lan.2003.0090
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. https://doi.org/10.1017/S095439451300015X
Miller, K., & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. *Language Acquisition*, *19*(3), 223–261. https://doi.org/10.1080/10489223.2012.685026
Morgenstern, A., & Parisse, C. (2012). The Paris corpus. *Journal of French Language Studies*, *22*(1), 7–12. https://doi.org/10.1017/S095926951100055X
Nardy, A., Chevrot, J., & Barbu, S. (2014). Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *Language Variation and Change*, *26*(3), 273–301. https://doi.org/10.1017/S0954394514000131
Newkirk-Turner, B. L., & Green, L. (2016). Third person singular-s and event marking in child African American English. *Linguistic Variation*, *16*(1), 103–130. https://doi.org/10.1075/lv.16.1.05new
Palasis, K. (2015). Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua*, *161*, 125–143. https://doi.org/10.1016/j.lingua.2014.11.012
Plunkett, B. (2002). Null subjects in child French interrogatives: A view from the York Corpus. In C. D. Pusch & W. Raible (Eds.), *Romance Corpus Linguistics: Corpora and Spoken Language* (pp. 441–452). Gunter Narr Verlag.
Pooley, T. (1996). *Chtimi: The urban vernaculars of northern France*. Multilingual Matters.
Poplack, S., & St-Amand, A. (2007). A real-time window on 19th-century vernacular French: The Récits du français québécois d'autrefois. *Language in Society*, *36*(5), 707–734. https://doi.org/10.1017/S0047404507070662
Purcell, A. K. (1984). Code shifting Hawaiian style: Children's accommodation along a decreolizing continuum. *International Journal of the Sociology of Language*, *1984*(46), 71–86. https://doi.org/10.1515/ijsl.1984.46.71
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
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*(2), 249–266. https://doi.org/10.1017/S0954394500001897
Roberts, J. L. (2002). Child language variation. In J. Chambers, N. Schilling-Estes, & P. Trudgill (Eds.), *Handbook of language variation & change* (pp. 333–348). Blackwell.
Romaine, S. (1978). Postvocalic /r/ in scottish english: sound change in progress. In P. Trudgill (Ed.), *Sociolinguistic patterns in british english* (pp. 144–157). Arnold.
Ross, D. S. (2001). *Disentangling the nature-nurture interaction: The effects of exposure to non-native input on the language acquisition process* [Doctoral dissertation]. University of Rochester.
Ross, D. S., & Newport, E. L. (1996). The development of language from non-native linguistic input. In A. Stringfellow & D. Cahana-Amitay (Eds.), *Proceedings of the 20th annual Boston University conference on language development* (pp. 634–645). Cascadilla Press.
Rowlett, P. (2013). Do French speakers really have two grammars? 1. *Journal of French Language Studies*, *23*(1), 37–57. https://doi.org/10.1017/S095926951200035X
Sankoff, G. (2019a). Age grading, style, and language change: A lifespan perspective. In R. Blake & I. Buchstaller (Eds.), *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). Georgetown University Press.
Sankoff, G., & Vincent, D. (1980). The productive use of ne in spoken montréal French. In G. Sankoff (Ed.), *The social life of language* (pp. 295–310). University of Pennsylvania Press.
Shin, N. (2016). Acquiring constraints on morphosyntactic variation: Children's Spanish subject pronoun expression. *Journal of Child Language*, *43*, 914–947. https://doi.org/10.1017/S0305000915000380
Shin, N., & Miller, K. (2022). Children's acquisition of morphosyntactic variation. *Language Learning and Development*, *18*(2), 125–150. https://doi.org/10.1080/15475441.2021.1941031
Shin, N., & Miller, K. L. (2024). Children's acquisition of morphosyntactic variation: A reply to commentaries. *Language Learning and Development*, *20*(1), 83–96. https://doi.org/10.1080/15475441.2023.2239798
Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American sign language from inconsistent input. *Cognitive Psychology*, *49*(4), 370–407. https://doi.org/10.1016/j.cogpsych.2004.05.001
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. https://doi.org/10.1017/S0954394507070044
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. https://doi.org/10.1515/ling-2013-0012
Suppes, P., Smith, R., & Leveillé, M. (1973). The French syntax of a child's noun phrases. *Archives de psychologie*, *42*, 207–269.
Wickham, H., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686
Wolfram, W. A. (1969). *A sociolinguistic description of Detroit Negro speech*. Center for Applied Linguistics.
Yamaguchi, N. (2012). *Parcours d'acquisition des sons du langage chez deux enfants francophones* [Doctoral dissertation]. Université Sorbonne Nouvelle Paris 3.