

18-month-olds fail to use recent experience to infer the syntactic category of novel words

Naomi Havron*,

Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS,
Ecole Normale Supérieure – PSL University, Paris, France

naomi.havron@mail.huji.ac.il

Mireille Babineau*,

Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS,
Ecole Normale Supérieure – PSL University, Paris, France

mireillebbineau@gmail.com

and Anne Christophe

Laboratoire de Sciences Cognitives et Psycholinguistique, DEC-ENS/EHESS/CNRS,
Ecole Normale Supérieure – PSL University, Paris, France

anne.christophe@ens.fr

Havron, N., Babineau, M., & Christophe, A. (2020). 18-month-olds fail to use recent experience to infer the syntactic category of novel words. *Developmental Science*, e13030. Doi:10.1111/desc.13030

*Joint first authors

Corresponding authors: naomi.havron@mail.huji.ac.il, mireillebbineau@gmail.com

The Stage 1 approved protocol detailing the method, analyses and criteria for exclusion of participants is accessible on Open Science Framework (OSF) with the following link: https://osf.io/ujesx/?view_only=f3b068cdee0043b9b4cd6db6ee8fba34. The materials, collected data, and data analyses are also freely available to readers through the same link.

Research highlights

- We asked whether 18-months-old infants rely on recent linguistic experience to infer the syntactic category of a novel word.
- We habituated infants to a novel word appearing in an ambiguous context, and biased them to expect it to be either a noun or a verb, using familiar words appearing in this context.
- We found no evidence that infants in the noun group were expecting a noun nor that infants in the verb group were expecting a verb.
- There was no difference between groups, thus, no evidence for expectation adaptation in 18-month-old infants.

Abstract

Infants are able to use the contexts in which familiar words appear to guide their inferences about the syntactic category of novel words (e.g., “This is a” + “dax” -> *dax* = object). The current study examined whether 18-month-old infants can rapidly adapt these expectations by tracking the distribution of syntactic structures in their input. In French, *la petite* can be followed by both nouns (*la petite balle*, “the little ball”) and verbs (*la petite mange*, “the little one is eating”). Infants were habituated to a novel word, as well as to familiar nouns or verbs (depending on the experimental group), all appearing after *la petite*. The familiar words served to create an expectation that *la petite* would be followed by either nouns or verbs. If infants can utilize their knowledge of a few frequent words to adjust their expectations, then they could use this information to infer the syntactic category of a novel word – and be surprised when the novel word is used in a context that is incongruent with their expectations. However, infants in both groups did not show a difference between noun and verb test trials. Thus, no evidence for adaptation-based learning was found. We propose that infants have to entertain strong expectations about syntactic contexts before they can adapt these expectations based on recent input.

Keywords: Word categorization, Language acquisition, Infants speech processing, Syntactic adaptation, Prediction, Syntactic bootstrapping

There is a general consensus in the field of language acquisition that infants use syntactic context to bootstrap their learning of the meaning of words. This is known as *the syntactic bootstrapping hypothesis* (e.g., Arunachalam, 2016; Bernal, Lidz, Millotte, & Christophe, 2007; Fisher, Gertner, Scott, & Yuan, 2010; Gleitman, & Lederer, 1999; Gleitman, 1990; Jin & Fisher, 2014; Arunachalam & Waxman, 2010). For example, toddlers can use the distributional information that articles tend to be followed by nouns (which tend to be words for objects), and pronouns tend to be followed by verbs (which tend to denote actions), to infer whether a novel word is likely to refer to an object or an action (e.g., He & Lidz, 2017; Bernal et al. 2007; Oshima-Takane, Ariyama, Kobayashi, Katerelos, & Poulin-Dubois, 2011). It is an open question whether these are stable expectations, or whether infants are able to flexibly change them to fit evidence from their input. Here, we examine whether infants are able to rapidly adjust their expectations for a syntactic category following distributional evidence from recent input. Such rapid adaptation of expectation would be in line with *prediction*-based learning.

Prediction¹ has been put forward as a mechanism that can capture key aspects of language acquisition. Under this proposal, children could acquire their language by predicting what words or types of words speakers will use in context, and use error signals to adjust their predictions (e.g., Chang, Dell, & Bock, 2006; Elman, 1993). If prediction is to explain key elements in language acquisition, then this would require that infants from an early age possess prediction mechanisms that can operate at the level of syntactic representations², and that they are able to adapt their representations rapidly enough to enable learning from context. The first premise is clearly shown in syntactic bootstrapping studies: Shortly after their first birthday, infants can already use context to predict a syntactic category (e.g., determiners will be followed by nouns; Cauvet et al., 2014; Kedar, Casasola, Lust, & Parmet, 2017; Melançon & Shi, 2010). The second premise is an open question, however, since at the moment we do not know how fast syntactic expectations are

¹ Note that the term *prediction* is not used here to talk about anticipation. The term ‘prediction’, as used by Andy Clark (2013) and others, relates to the idea that we try to make sense of incoming sensory input using expectations derived from our current view or representation of the world. We adapt these representations if current input is unexpected. This induced change in our representations following a discrepancy between input and expectations may well happen after several sentences have been processed, and is thus not defined by the timeline of the predictive process.

² These do not have to be equivalent to adult syntactic categories, but to any classification of words to types, such as "actions" and "objects".

formed, and how flexible they are. It is possible that a large quantity of data has to be accumulated before predictions are formed or changed. This seems to be a hidden assumption of some accounts of learning which emphasize the importance of overall frequency of word-sequences in corpora (e.g., Hoff & Naigles, 2002; Lieven, 2010, though note that these proposals do not necessarily contradict accounts which assign more weight to recent input). If infants' expectations are based on aggregated statistics, performed on large quantities of stored information, then recent distributional information should have a limited effect on infants' syntactic predictions. If, however, infants constantly track distribution of syntactic structures and adapt their representations rapidly enough to enable learning from context, then a short exposure to new information should be sufficient to change their predictions. In the current study we test whether 18-month-old infants can adapt to recent distributional information about syntactic categories. If they do, this will show that flexibility and rapid adaptation to recent distributions of structure is already present at the beginning stages of language learning, and that it is powerful and flexible enough to guide prediction-based syntactic categorization.

Computational studies have shown that models relying on distributional information such as the position of determiners and pronouns in sentences, are useful for syntactic categorization (Brusini, Amsili, Chemla, & Christophe, 2014; Chemla, Mintz, Bernal, & Christophe, 2009; Mintz, 2003; Mintz, Newport, & Bever, 2002; Saint Clair, Monaghan, & Christiansen, 2010). These studies show that a considerable amount of information concerning syntactic categories can be obtained from distributional information in the input. However, all these models faced some challenges, such as how to categorize homophones that can appear both in noun or verb contexts (i.e., I watch vs. the watch) or how to handle the fact that some contexts can predict both nouns and verbs. For example, *le* ("the" or "it") is an ambiguous function word in French, which can appear either before a verb (*je le vois*, "I see **it**") or before a noun (*c'est le lapin*, "it's **the** bunny"). One thing that may be lacking in these models is that they aggregate over stored information rather than changing their predictions when new information does not fit predictions. For example, in the Brusini et al. (2014) model, if a given context was encountered with nouns 50 times before, the model will have to encounter 51 instances where the same context contains a verb before changing its prediction from *noun* to *verb*. Under a prediction-based account of learning, however, every new instance of a verb following this context should carry weight in the generation of expectations. This is consistent with the behavior of adults, who can be made to change their expectations in the

course of one experimental session (Fine, Jaeger, Farmer, & Qian, 2013; though see Harrington Stack, James, & Watson, 2018).

As regards to children, it was recently shown that at a young age they are in fact able to rapidly utilize recent distributional information to infer the meaning of novel words. Beretti, Havron, and Christophe (under revision) found that 4-to-5-year-old children adapt to the properties of recent linguistic input to interpret an ambiguous sentence. The study examined children's interpretation of ambiguous sentences, which could be interpreted either as ungrammatical or as introducing a novel homophone: for instance, 'the eats' is ungrammatical if 'eats' is recognized as a verb (it should be 'she eats'), but grammatical if the participant interprets 'eats' as a novel noun. The results showed that children's expectations depended on whether the speaker previously produced ungrammatical sentences, or previously introduced new homophones. That is, children were changing their predictions for a syntactic category depending on recent experience. Similarly, Havron, de Carvalho, Fiévet and Christophe (2019) used a syntactic context that can precede both nouns and verbs to manipulate what syntactic category children should expect. In French, *la petite* can be followed either by a noun (*la petite balle*, "the little ball"), or by a verb (*la petite mange*, "the little one is eating"). By manipulating which familiar words followed this context in an induction phase, the researchers made this syntactic context become predictive of different categories for different groups of participants. They found that children for whom *la petite* preceded familiar verbs were more likely than children for whom it preceded familiar nouns, to infer that a novel word appearing in this context refers to an action. This shows that 3-to-4 year-old children adapt rapidly to the distribution of nouns and verbs following *la petite*, and use it to infer novel words meaning (whether *la petite nuve* means "the little nuve" or "the little one is nuving"). Both studies, however, focused on young children and not on infants, and it is still unclear whether the same ability to rapidly integrate novel syntactic information that was found in young children is already present in infants, or whether infants need more consistent evidence to adapt such syntactic expectations. In order to show that prediction-based learning is a key component in learning, one must test young infants, in the earliest stages of language acquisition.

In the current study we ask whether 18-month-old infants can use recent distributional information to categorize novel words appearing in an ambiguous context – that is, one for which the accumulated evidence is inconclusive as to whether the following word should be categorized as a verb or a noun. As in Havron and colleagues (2019), we will use the context *la petite* which

can either be followed by verbs or by nouns. However, unlike Havron and colleagues, we will not use an eye-tracking paradigm with two simultaneous videos, which would be too demanding for 18-month-old infants, but rather, a habituation paradigm, more suitable for this age.

A similar paradigm has been used by Shi & Melançon (2010) to show that young infants can categorize novel words based on the syntactic context in which they appeared. They familiarized French-learning 14-month-olds with novel words appearing in either noun or verb contexts (i.e., preceded by determiners for one group of infants, and by pronouns for the other group). Both groups were presented with the same two types of test trials: a noun context, with a determiner that was withheld from the habituation phase; and a verb context, with a pronoun that was withheld from the habituation phase. The researchers found that 14-month-old infants in the noun condition looked longer when the same words appeared in a verb context - incongruent with their expectations if they have categorized the novel words as nouns - than when they appeared in a different noun context. This suggests that 14-month-old infants already know enough about noun contexts to have syntactic expectations. In contrast, infants in the verb condition showed no difference in looking times to the two test trial types, suggesting that their knowledge of verb contexts was not sufficient to support generalization to a new verb context.

In a recent study, however, when familiar verbs were added to the habituation phase, 14-month-olds were able to categorize novel verbs using frequent verb contexts (Babineau, Shi & Christophe, in press). The novel words appeared in the same verb contexts as the familiar words (e.g., *il mange*, *il dort*, *il dase*, “he is eating, he is sleeping, he is dazing”). In this condition, infants looked longer when the novel words were presented in a noun context (the surprising event if the novel words were categorized as verbs) than when the novel words were presented in a verb context (the expected event), thus showing evidence of also categorizing verbs, and not only nouns, using verb-predictive contexts. The authors proposed that the presence of familiar verbs might have facilitated the categorization of the novel words through the activation of previously encountered and stored verb contexts. These findings suggest that infants benefit from the presence of familiar words to facilitate the use of syntactic contexts for categorization – however, they leave open the question of whether infants can rapidly adapt their predictions of a word’s syntactic category based on changes in the input.

In the current study, the habituation phase will be used to create different expectations for the different groups of participants about which syntactic category will follow the syntactic context

la petite. Infants will be repeatedly exposed to *la petite* followed by either familiar nouns or familiar verbs, depending on the experimental group. If infants can learn the distributional properties of syntactic contexts, and rapidly adjust their expectations, then they will infer that the novel word appearing in the same context is of the same syntactic category as these familiar words – thus adjusting their expectations to recent information about which category this context predicts. To test whether infants have assigned the novel word to a noun or a verb syntactic category, we will expose them, in the test phase, to a noun context (a determiner) and a verb context (a pronoun), as in Shi and Melançon (2010) and Babineau et al (in press). If infants have categorized the novel word correctly – despite the fact that it never appeared in an unambiguously noun or verb context – they should be surprised, and listen longer to the trials in which the word appears in a context incongruent with the syntactic category of the familiar words in the habituation phase. Take, for example, an infant who would be habituated to *la petite* followed by familiar verbs, played alongside a novel word appearing in this context (*la petite crale*). In the test phase she should listen longer to *une crale* (“a crale”), because it is incongruent with the expectation that “crale” will appear in verb contexts - than to *tu crales* (“you are craling”), which is congruent with the expectation she has formed during the habituation phase. If infants show this pattern of results, this would suggest that infants constantly track syntactic contexts, and utilize them to update expectations about novel words’ syntactic categories. This could thus be a mechanism through which infants exploit words they already know in order to build expectations about the kinds of properties that novel words, occurring in similar contexts, are likely to possess.

Method

Participants

Fifty-eight French-learning 18-month-olds from the Paris area participated in the study (29 in each condition; 28 girls, age: $M = 17.9$ months, $SD = .22$, range = 17.54-18.43). Infants were monolinguals, and had no known hearing problems. An additional 29 infants came to the lab but were not included in the final sample for one of the following reasons: not being able to finish the experiment because of crying/fussiness/walking away ($n = 23$), failing to meet the preset habituation criterion ($n = 1$), parental interference ($n = 4$), or a technical problem ($n = 1$). Moreover, three infants were excluded because they turned out to be raised in a bilingual environment (i.e.

they had less than 70% of exposure to French). Parents provided informed consent. The local ethics committee approved this research.

We had planned (and preregistered our intention) to not pre-specify a sample size, but instead to conduct a Bayesian analysis and run subjects until the Bayes Factor (BF₁₀) for the interaction between the two groups is 3 (indicating that the data are 3 times more likely to have occurred under H₁) or less than 0.3, indicating evidence for H₀ (see Schönbrodt, Zehetleitner, & Perugini, 2015 and analyses section below). As planned, we tested 40 infants (20 in each condition) before running our first analysis, given that studies with less than 20 infants per cell give less consistent results (Oakes, 2017), and added a total of eight infants for every additional analysis. Since the first and second analysis did not meet our pre-specified inference criterion, we ran a third analysis (see Results section). Since we did not cancel appointments that were already taken with parents, two more infants were run before this final analysis was performed (which led to 29 infants in each group instead of 28).

Design

A habituation paradigm was implemented with a central fixation procedure. The experiment contained two phases, a habituation phase and a test phase. Infants were randomly assigned to one of two conditions (noun vs. verb). Infants in the noun condition were familiarized with trials presenting familiar nouns in the ambiguous context *la petite* (e.g., *la petite balle*, “the little ball”) and one novel word in the same context (*la petite crale*), along with trials presenting familiar nouns and verbs in unambiguous contexts (e.g., *c’est ma balle*, “it’s my ball”, and *je mange*, “I eat”). In contrast, infants in the verb condition were familiarized with trials presenting familiar verbs and the same novel word in the ambiguous context (e.g., *la petite mange*, “the little one is eating”, and *la petite crale*), along with trials presenting familiar verbs and nouns in unambiguous contexts. Hence, familiar nouns and verbs were presented in the habituation phase of both conditions, but the *la petite* context was followed exclusively by familiar nouns in the noun condition, and by familiar verbs in the verb condition. This ensured that, if infants are surprised to hear the novel word in the incongruent context, this is because they categorized it based on the recent distributions of the ambiguous context *la petite*, and not simply based on the syntactic category of the familiar words presented during habituation. Once infants had reached the habituation criterion, the habituation phase ended and the test phase started automatically.

In the test phase, both groups were presented with the same stimuli containing the novel word. There were two types of test trials: Verb contexts involving the pronoun *tu* (“you”), and noun contexts involving the article *une* (“a”). For each group, one type of test trials was congruent with the syntactic category assigned to the novel word during the habituation and the other one was incongruent. For instance, for infants who were habituated to *la petite* as a verb context, the test trials presenting the novel word preceded by the pronoun *tu* (i.e. *tu crales*) was congruent with the learned category, whereas the test trials presenting the novel word preceded by the article *une* (i.e. *une crale*) was incongruent. Note that both trial types contained contexts that were not used in the habituation, the only difference between the two trial types is that one is a noun context, and the other a verb context.

Stimuli

Speech stimuli included one novel word (*crale*), which could be interpreted as either a noun or a verb, preceded by the ambiguous context *la petite*, which can either mean “the little one” (e.g. *la petite mange* “the little one is eating”) or “the little” (e.g., *la petite balle* “the little ball”). They also included sentences where *la petite* is followed by familiar words. A total of eight familiar words, four for each group, appeared in the *la petite* context: four verbs (Verb Condition) and four nouns (Noun Condition; see Table 1). Though the ambiguous context *la petite* is permissible in French both before a noun and before a verb, it is more likely to be followed by a noun. A search in the Paris child-directed-speech corpus (Morgenstern & Sekali, 2009) found 720 instances where it was followed by a noun and only six where it was followed by a verb. Importantly, de Carvalho and colleagues (2019) found that 18-month-olds could interpret a novel word that appears following the *la petite* context as either a noun or a verb, suggesting that both readings are available to them. In that case, the noun or verb reading was triggered by a different prosodic structure for sentences involving a novel noun, such as *bamoule* in [*Regarde la petite bamoule*]! – [Look at the little bamoule]!, than those involving a novel verb as in [*Regarde*]! [*La petite*] [*bamoule*]! – [Look]! [The little one] [is bamouling]! (square brackets mark prosodic boundaries). Our study, on the other hand, used ambiguous prosody as in Havron and colleagues (2019; see Fig. 1).

To make the habituation phase less repetitive and more varied, and to ensure that both groups would hear both verbs and nouns in the habituation phase, some sentences did not contain the context *la petite*, but instead contained unambiguous frequent noun or verb contexts (e.g., *Elle*

mange bien. “She eats well.” *C’est ma balle*. “It’s my ball.”). For the test phase, we created sentences with the novel word in a different verb context (following the pronoun *tu* “you”) and a different noun context (following the determiner *une* “a”). For a full list of all sentences used, see Table 1.

A female native French speaker (the last author) recorded multiple repetitions of the 27 sentences in child-directed speech style (sampling frequency 44.1 kHz). All sentences were used in the habituation phase, except for sentences using the pronoun *tu* and the article *une*, which were reserved for the test phase. The final stimuli consisted of one exemplar for each *la petite* sentences, three exemplars for the rest of the 16 habituation sentences, and six exemplars for each of the two test sentences. The exemplar of the ambiguous sentence with the novel word (i.e. *la petite crale*) was carefully chosen to ensure that the prosodic properties would not bias infants towards a noun or a verb interpretation. This was done by intentionally picking a prosody that is appropriate for both nouns and verbs, in which the three words belong to the same prosodic phrase (see Figure 1); otherwise it would have been possible to disambiguate by using different prosodic structures (as shown in de Carvalho, Dautriche, Lin & Christophe, 2017; de Carvalho, Dautriche, & Christophe, 2016). A previous experiment found that adults rate both nouns and verbs as possible when this prosody is used (novel noun or novel verb following *la petite*), though nouns seem to be preferred (Havron, et al., 2018), which fits with the fact that nouns are more frequent than verbs in this construction, as mentioned above. Note that even if there are some prosodic cues that bias towards one interpretation (verb or noun), this does not affect the validity of our results, since the same exemplar was used in both conditions (verb and noun). Therefore, prosody could not create any artefactual difference between groups. If the two groups behave differently in the test phase, this will show that infants were able to adjust their inferences about a novel word's category depending on what they heard in the habituation phase.

We created three different habituation trials containing 32 sentences in different orders (see Table 1 for a full list of sentences). Twenty of these sentences contained familiar words (both verbs and nouns) and 12 sentences contained the novel word. Infants from the verb condition could hear up to (if they would listen until the end of the trial) 12 tokens with the familiar verbs, eight in *la petite* context, four with unambiguous verb contexts, mixed with eight tokens involving familiar nouns (in unambiguous contexts). Infants in the noun condition could hear up to 12 tokens with the familiar nouns, eight in *la petite* context, four unambiguous contexts, mixed with eight tokens

involving familiar verbs in unambiguous contexts involving function words. Within a trial, the order of the sentences was pseudo-randomized with three restrictions: A particular sentence could not appear twice in a row, a sentence with the novel word should appear at least once every three sentences, and the same word could occur only twice in a row. The order in which those three trials were presented during the habituation phase was counterbalanced across infants. The total length of a habituation trial was fixed to 68.5 sec. Test trials contained 16 sentences and were 27.5 sec long. Mean acoustic measures for all stimuli are shown in Table 2.

An animation of a bird moving its mouth in synchrony with the audio stimuli was presented during each trial. We used a silent video of a butterfly perched on a leaf as an attention-getter, to attract the infant's attention between trials. All stimuli are available on the OSF (https://osf.io/ujesx/?view_only=f3b068cdee0043b9b4cd6db6ee8fba34).

Procedure

Infants were tested individually using a habituation paradigm implemented in a central fixation procedure. In a sound-attenuated double-walled booth, they sat on their caregiver's lap about 70 cm from a wall-mounted 27-inch monitor. Sound was presented to the infant through two loudspeakers, positioned on each side of the monitor. Caregivers were instructed not to talk or point towards the screen. They listened to music through noise-canceling headphones to prevent them from hearing the stimuli. A camera above the monitor was connected to a second monitor placed outside the booth. This set-up enabled the experimenter to observe infants' reaction from an adjacent room without being aware of the stimuli presentation.

We presented the stimuli using Habit (Cohen, Atkinson, & Chaput, 2004). The experimenter pressed a computer key to initiate a trial when the infant looked toward the monitor, and she pressed another key whenever a look toward the monitor occurred during a trial. Trials ended either after the maximum trial length was reached or after the infant had looked away from the screen for more than two consecutive seconds. A new trial started when the infant looked back towards the screen.

The habituation criterion was reached when an infant's average looking time during any block of three consecutive trials dropped to less than 65% of the average looking time for the most-attended block (i.e., the 3-trial block that has the longest total looking time). Therefore, the total number of habituation trials each infant received varied (between a minimum of four and a

maximum of 12 trials). The test phase either began with a trial presenting verb syntactic contexts or a trial presenting noun syntactic contexts. The order of presentation was counterbalanced across infants (N V N V or V N V N). The two types of trials alternated for a total of four trials (two of each type).

Results

Planned analyses

Mean listening times to the noun and verb test trials were calculated for each subject. Difference scores were used as the dependent variable for the analyses, i.e. the difference between the two types of test trials (mean listening times during noun test trials minus mean listening times during the verb test trials; a positive difference means longer looking times for noun trials, which is the expected result for the verb condition). We employed sequential hypothesis testing with Bayes factors with a pre-specified inference criterion of $BF_{10} > 3$, indicating evidence for an asymmetry between the size or direction of the difference between trial types in the two conditions (noun versus verb induction), and of $BF_{10} < 0.3$, indicating evidence for H_0 . The Bayes factors was obtained from a Bayesian t-test using the JASP software (JASP team, 2018). We used a Cauchy prior of 1, as recommended by Schönbrodt et al. (2015), but will also examine other priors to assess the robustness of our results with different priors. The Bayesian analysis with the final sample revealed moderate evidence in favor of the null hypothesis ($t(55) = -0.042$, $p = 0.967$, Cohen's $d = -0.011$. $BF_{10} = 0.199$), indicating no difference in infants' difference scores between the two conditions (see Figure 2). Thus, children in the verb condition failed to adapt to recent syntactic distribution, and instead showed the same listening preference ($M = -1.253$, $SD = 7.567$) as children in the noun condition ($M = -1.309$, $SD = 7.031$, i.e. longer listening times during the verb test trials). For the robustness region of this analysis, the use of Cauchy priors with a scale smaller than 0.61 gives results above 0.3, Cauchy priors larger than 1 provide even stronger evidence for H_0 .

As planned, we tested the effect of trial type within each condition (noun and verb induction) by conducting Bayesian one-sample t-tests. Contrary to our expectations, results were moderately in favor of the null hypothesis: Children's average listening times between the two types of test trials (difference scores) were not different from zero for the noun condition (a difference of -1.36 , $SD = 7.16$, 95% CI for difference $[+1.44 ; -4.11]$, $t(27) = -0.987$, $p = 0.332$,

Cohen's $d = -0.187$, Cohen d 's CIs: $[-0.522 ; +0.507]$, $BF_{10} = 0.233$), nor the verb condition (a difference of -1.25 , $SD = 7.57$, 95% CI for difference $[+1.62 ; -4.13]$, $t(28) = -0.892$, $p = 0.38$, Cohen's $d = -0.166$, $BF_{10} = 0.210$). This suggests that after hearing many repetitions of *la petite crale* during the habituation phase, infants failed to build strong expectations about the syntactic context in which the novel word *crale* would be likely to be next encountered (i.e. following a pronoun or a determiner). These findings will be discussed in depth in the Discussion section.

To examine how much each wave of testing changed the strength of our evidence we performed a sequential analysis using JASP. As can be seen in Figure 3, adding new infants shows a steady decreasing pattern, rather than a fluctuating one. Some fluctuations only occur until we tested about 14 participants in each group, from which point evidence towards H1 consistently, but slowly, dropped³.

Lastly, we compared our results to those of Babineau, Shi & Christophe (in press). The analysis comparing the two verb conditions revealed anecdotal, although significant, evidence in favor of H1 ($BF_{10} = 1.686$, $t(55) = -2.208$, $p = 0.031$, Cohen's $d = -0.585$), giving some indication that children in our verb condition behaved differently than children who built predictions from typical/unambiguous verb contexts (following frequent pronouns) in the study of Babineau, Shi, and Christophe. Fourteen out of 29 infants listened longer to incongruent noun test trials than to correct verb test trials in the current study, while 18 out of 28 children showed this expected pattern in the previous study. The effect size for the verb condition was larger in the former study by Babineau et al. (Cohen's $d = 0.438$) than in the present study's verb condition (Cohen's $d = -0.166$). However, the analysis comparing the noun conditions between these two studies revealed anecdotal evidence in favor of the null hypothesis ($t(54) = -0.922$, $p = 0.361$, Cohen's $d = 0.246$, $BF_{10} = 0.294$). Hence, although infants from the noun condition in the current study did not show a looking time difference between the two blocks, while infants from the noun condition in the previous study did – this analysis shows that infants in both experiments (*la petite* vs determiners) behaved similarly and looked longer during the incongruent verb test trials than to the congruent noun test trials (18 out of 28 children in our study vs. 20 out of 28 children in Babineau, Shi &

³When performing this analysis, we noticed that in fact our threshold, a BF_{10} of 0.3 was passed even before we finished testing 20 participants in each condition, and thus, our two additional waves of testing were not strictly necessary. This error was a result of using the default JASP prior, $Cauchy = 0.707$, and not our preregistered one of $Cauchy = 1$, in our intermediate analyses. With the default JASP prior our evidence crosses the $BF_{10} = 0.3$ line only after about 50 participants.

Christophe's study). Note, however, that this effect was larger in the former study's noun condition (difference from zero = -3.2, 95% CI [-6.27, -0.12], Cohen's $d = -0.404$, 95% CI [-0.786; -0.014]) than the one in the present study (a difference of -1.36, 95% CI [+1.44 ; -4.11], Cohen's $d = 0.186$, 95% CI [-0.552; +0.183])⁴.

Exploratory analyses

ANOVA. Above we compared the noun condition infants from the current experience with the noun condition infants from a previous experiment, which used unambiguous sentences; as well as their verb condition to our verb condition. To further explore the relation between these two studies, we ran an ANOVA with difference score as the dependent variable, and condition, experiment and an interaction as our independent variables. We used the default priors in JASP, r scale fixed effects on 0.5 and r scale random effects of 1. The effect of the interaction was BF10 of 2.05, anecdotal evidence for H1 ($F(3, 100) = 4.9$, $p = .028$). Thus, we cannot say with confidence that the infants' performance was importantly different between these two studies without additional evidence.

Simulations. Next we ran several simulations to examine whether increasing the sample size of our study and of Babineau et al. (*in press*) would allow us to draw firmer conclusions. We conducted a set of simulation analyses in which we repeatedly created ever-increasing sample sizes by sampling additional infants from both the Babineau et al and the current study. We created datasets by sequentially sampling 8 more participants from our combined dataset (on average 4 from each experiment) and adding them to our data, then running the same ANOVA as we did with the real data on these datasets. This procedure was repeated 250 times for each sample size.

Larger BF values are found with larger sample size, but many samples (29.9% overall) still fall below the BF10 = 3 threshold. When sample size becomes larger, this proportion decreases, indicating more evidence for H1 (i.e., a larger difference between the noun and the verb conditions in Babineau et al. study than for our study). However, even with the largest sample size (266), 10.4% of samples fall below our threshold.

⁴ The size of the main effect (difference between condition was also bigger in Babineau and colleagues' (*in press*) study: For our study, we had a difference of -0.056 (95% CI for difference [-3.898 ; +3.787], Cohen d 's CIs: [-0.522 ; +0.507], Cohen's $d = -0.011$). In Babineau et al. (*in press*), they had a difference of -6.038, CI for diff [-9.902 ; -2.174], Cohen d 's CIs: [-1.368; -0.284], observed Cohen's $d = -0.830$).

Comparing the noun conditions across experiments. The next simulation we ran included resampling only from the noun conditions in both experiments. We ran this simulation because our analysis of the actual data seemed conflicting – on the one hand the noun group in Babineau et al. showed expectations while ours did not, and on the other hand there was evidence for H0 – no difference between these two groups. Each sample size was sampled 100 times. Our simulations show that increasing the sample size gradually increases evidence for H1 (Babineau and colleagues' noun condition shows greater evidence for expectations than our noun condition), but many samples (35.73%) still fall below the $BF_{10} = 0.3$ line. Even with the largest sample size (289), 21% of samples fall below this line, and 33% fall above the $BF_{10} = 3$ line, which is a very weak support for H1.

Comparing the verb conditions across experiments. To complete the picture, we sampled in the same way from the verb conditions in both experiments. Each sample size was sampled 100 times. Our simulations show that increasing the sample size gradually increases evidence for H1 (Babineau and colleagues' verb condition shows greater evidence for expectations than our verb condition). Most samples (89.1% overall) fall above the $BF_{10} = 3$ line and none fall below the $BF = 0.3$ line. With the largest sample size (289), all samples fall above the $BF_{10} = 3$ line. Why are the results for the verb conditions simulations so much clearer than those for the noun conditions? Note that in our experiments, both groups showed slightly longer looking times for verb trials, that is, both the noun and verb groups behaved as if they expected a noun (though the difference between the two trial types was too small to indicate evidence for true expectations, as detailed above). In Babineau and colleagues (*in press*) on the other hand, the noun group showed the same trend as ours (only stronger), while the verb group showed a difference in the *opposite* direction, indicating a true verb expectation.

Exploring additional factors. Supplementary analyses were conducted in order to investigate possible sources of variability that could have affected infants' performance in the task. Since our study used a habituation paradigm, the length of exposure to the target syntactic distribution during the habituation phase (noun induction vs. verb induction) varied, and could have affected our results. Thus, we examined whether it correlated with the difference scores in either condition by running a linear regression with the difference score as our dependent measure, and habituation length, condition and an interaction as our independent measures. None of our dependent measures

had a significant effect on the difference scores. Moreover, the average habituation length was similar between the two conditions (Noun group: $M = 175.1$ sec, range = 69.9-500.5 sec; Verb group: $M = 154.57$ sec, range = 34.4-279.9 sec), yielding a high average exposure to the novel word (Noun group: 30.9 times; Verb group: 27.3 times).

Following our laboratory practices, parents were asked to complete a French adaptation of the short version of the MacArthur-Bates Communicative Development Inventory (Kern & Gayraud, 2010) within 1 week of participation, as a measure of infants' receptive and productive vocabularies. Scores were obtained for 48 of 58 infants. Reported comprehension vocabulary ranged from 29 to 95 words ($M = 68.5$, $SD = 17.4$), and reported production ranged from 0 to 86 words ($M = 16.6$, $SD = 17.5$). Since it is possible that only infants with a large vocabulary could utilize the ambiguous syntactic frame *la petite* for prediction of syntactic category of the novel word, we examined whether there was a main effect of condition, vocabulary size and an interaction. We ran the same analysis with infants' age instead of vocabulary. Neither the infants' age nor its vocabulary size correlated with performance (i.e. difference scores), and these variables did not interact with conditions (see "Exploratory analyses" on the Open Science Framework).

Discussion

In the current study, we asked whether 18-month-old infants can use recent distributional information to categorize a novel word appearing in an ambiguous context (i.e. the context *la petite*, which can either mean "the little" or "the little one"). To address this question, we used a habituation phase meant to create different expectations about the syntactic category following *la petite*. Depending on the experimental group, infants were exposed to multiple sentences where *la petite* was followed by either familiar nouns or familiar verbs (e.g., Noun induction group: *la petite balle*, "the little ball"; Verb induction group: *la petite mange*, "the little one is eating"). We reasoned that if infants can learn the category predicted by this ambiguous context in the context of the experiment, then they could infer that a novel word appearing in the same context also belongs to the same category (i.e. noun vs verb). If the two groups of infants indeed showed different expectations, then these would have been the result of infants' ability to flexibly change their representations and expectations to fit evidence from their recent input. However, contrary to our initial predictions, we found no evidence that infants were adapting their expectations in the context of the experiment.

We remind the reader that in habituation studies such as ours, a pattern of longer looking times to the incongruent trials is interpreted as indicating surprise, and suggests that infants noticed that these trials were incongruent with their expectations. In the current study, we expected infants exposed to sentences where *la petite* was followed by nouns to look longer at test trials where the novel word was preceded by a pronoun. That is, trials which indicate that the novel word is a verb. If infants were adapting their expectations, then this pattern should not hold for the other group of infants, for whom *la petite* was followed by verbs. This group should either show no preference or be surprised by noun trials (where the novel word was preceded by a determiner). We came to see the noun group as our baseline based on the distributional properties of French, where the syntactic context *la petite* is overwhelmingly followed by nouns in day-to-day speech, especially in infant-directed speech (Havron et al., 2019), where it is used as a form of diminutive. We thus expected the default expectation of infants to be that *la petite* will be followed by a noun. Our prediction was disconfirmed: Infants in the noun group did not show a strong noun expectation, and there was no evidence that they listened longer to verb test trials than to noun test trials.

These results are surprising in light of previous studies showing that 18-month-old infants are capable of interpreting a novel word following the context *la petite* as either a noun or a verb (de Carvalho, He, Lidz, & Christophe, 2019). However, there are two important differences between de Carvalho and colleagues' design and ours. The first is the use of a visual context, a video, to support infants' inferences about the meaning of a novel word, i.e. referring either to a novel action or to a novel agent/object (a penguin). In de Carvalho and colleagues, infants watched videos of a penguin cartwheeling and spinning. They heard two sentences with *la petite + novel-word*, and were expected to interpret one novel word as a verb and one novel word as a noun (depending on the phrasal prosody used). At 18 months, this visual illustration of the action and novel animal might have been crucial for enabling infants to make a strong prediction when the grammatical context is ambiguous. The second critical difference is the presence of prosodic cues. In de Carvalho and colleagues, noun and verb interpretations were manipulated using a very clear and conspicuous noun or verb prosody, which disambiguated between these two interpretations. When listeners heard [*Regarde*], [*la petite*][*doripe*] (“[look], [the little (one)] [is doriping]”), they interpreted *doripe* as a verb. When they heard [*Regarde la petite doripe*] (“[look (at) the little doripe]”), they had no choice but to interpret *doripe* as a noun. Even in the absence of visual referents (with an auditory-only task), French-learning 17-month-olds have been shown to use

disambiguating prosodic cues to categorize words (Massicotte-Laforge & Shi, 2015). In the current study we deliberately selected and recorded an ambiguous prosody, which was shown in a previous study to enable both interpretations in older children (though at least for adults, this prosody favored the noun interpretation, see Havron et. al., 2019). It is quite possible that an unambiguous prosody is a prerequisite for an accurate or clear expectation at 18 months of age. Prosody has been shown to aid language processing and learning in infants as early as six-months of age (e.g., Shukla, White, & Aslin, 2011; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). Prosody is also highly exaggerated in infant-directed speech, possibly in order to facilitate infants' language processing and acquisition (Graf Estes, & Hurley, 2013). While we have no data on the relative frequency of appearance of *la petite* sentences with different types of prosody in French, it would be reasonable to suppose that sentences containing an ambiguous prosody would be less frequent in infant-directed speech than those that carry unambiguous prosodic information. In their everyday lives, children often encounter such ambiguous contexts together with these disambiguating cues (visual context and prosody), which likely serve as needed anchors. Hence, it is possible that at 18 months of age, these extra cues are still guiding infants' interpretation of ambiguous cases, and they are at a loss when they are absent.

Whatever the reason, it is clear that infants in the noun group were not forming strong expectations about the syntactic contexts in which the novel word was most likely to be next encountered (co-occurring with a determiner or a pronoun). Based on previous studies, younger infants can build syntactic expectations about novel words when these words occur in very frequent and unambiguous contexts. That is, it has been shown that infants as young as 14-months have expectations for novel nouns to follow unambiguous noun contexts (such as the determiner *un*, "a", Babineau, Shi, & Christophe, in press; Shi & Melançon, 2010), and for novel verbs to follow unambiguous verb contexts (Babineau, et al., in press). Since our study used ambiguous contexts and ambiguous prosody, we compared the results of our two conditions to those of one of these previous studies (Babineau et. al., in press). This should have allowed us to examine whether infants have stronger expectations when the syntactic context is unambiguous. However, our analysis found only anecdotal evidence that infants have stronger expectations in unambiguous contexts in the verb condition. Additionally, for the noun condition we found evidence that infants' expectations *did not differ* between ambiguous and unambiguous contexts. We ran simulations to examine whether these results might change with a larger sample size, and found that overall the

difference between these two experiments would get larger with larger sample sizes, but the evidence for an interaction would remain weak. We hope that future studies will be able to shed more light on the factors contributing to successful syntactic expectation during infancy.

A large portion of the findings in the field of language acquisition is based on some behavioral measure of infants' allocation of visual attention (Aslin, 2007). This indirect measure has the advantage of being suitable for infants, but it also has the disadvantage of being potentially noisy. Even if we have a sufficient sample size (which might be hard to acquire when testing infants), how can we make sure the study design can lead to robust results? Some experimental paradigms have been shown to be less reliable than others, i.e. yielding smaller effect sizes and more false negatives. For instance, a recent registered large-scale replication report, ManyBabies1, that focused on measuring preference (in that case, for infant-directed over adult-directed speech, in 12-month-olds) found that the single-screen central fixation procedure yielded smaller effects than the head-turn preference procedure (The ManyBabies Consortium, 2019). In our study, we measured older infants' surprise to an incongruence using a habituation method, which was not studied in ManyBabies1. Hence, we cannot be certain whether our chosen method produced more noise to our data than some other method, or whether we have chosen the best available paradigm for testing this particular type of question with 18-month-olds. As infant researchers, we are faced with limited options when it comes to choosing an experimental paradigm and design suitable for infants and toddlers, we therefore encourage our field to further investigate the validity and limits of these tools and to develop new ones.

Given our results, we can conclude that the current study found no evidence that infants at 18-months of age are capable of flexibly and rapidly adapting their expectation. However, the lack of difference between the two test trial types in *either* of our experimental conditions points to an important reservation. Infants (and children, and adults) cannot be expected to change their expectations if they have none in the first place. A prerequisite for expectation adaptation is that expectation exists. A second prerequisite would be that infants understand that their expectations were wrong. If neither of these prerequisites are fulfilled, as it seems to be the case in our study, we cannot expect infants' predictions to change in the course of an experiment.

With that said, recent years have seen multiple publications describing weaker prediction or prediction adaptation the younger the participants (Barrault, Havron, Dautriche, Babineau, de Carhalho, Christophe, *in preparation*; Berreti, Havron & Christophe, under revision; Gambi,

Gorrie, Pickering, & Rabagliati, 2018; Havron et al., 2019). Furthermore, recent studies have failed to replicate high-profile adult studies showing fine-grained syntactic adaptation (Liu, Burchill, Tanenhaus, & Jaeger, 2017; Harrington Stack & James & Watson, 2018). While these studies do not diminish the importance of other studies that did find evidence for prediction and expectation adaptation in children (e.g., Havron et al., 2019; Gambi, Pickering, & Rabagliati, 2016; Ylinen, Bosseler, Junttila, & Huotilainen, 2017), they do cast into doubt on the necessity or central role of prediction in the first stages of language acquisition (and see Huettig, 2015). Rather than assume that prediction or expectation adaptation is necessary for learning, the picture that recent findings draw is more complex. Like other types of linguistic knowledge and cognitive abilities, children's predictions and expectations may become more complex and fine-grained as they mature. Children can rapidly and flexibly adapt these expectations, but they seem to become better at performing this exercise as they grow up.

Last, we wish to discuss general lessons learned from planning, writing and conducting a registered report in developmental psychology, given that this paper is likely to be one of the first registered reports published in the field. *Registered Reports* (Chambers, 2013; Nosek & Lakens, 2014) were put in place in order to counter a list of commonly practiced questionable research practices such as such as HARKing (Hypothesizing After Results are Known), p-hacking (fishing for a significant result or analysis by trying multiple methods) and cherry-picking (choosing one significant result to report and ignoring nonsignificant results or analysis). Since infants are an especially hard-to-reach and to-test population, the value of publishing a registered report might be even greater than in areas where recruitment and testing are easy. Several rounds of peer review would arguably improve the design and put in place safeguards for quality control, which would reduce the file-drawer problem on the one hand, and the incentive to engage in questionable research practices on the other (see Havron, Bergmann & Tsuji, *in press*, for best practice preregistration recommendations in infant studies).

The introduction of registered reports in the field of psychology seems to have had desirable effects. A recent study compared the rate of significant results in registered reports in the field of psychology with a sample of non-registered reports. It was found that while non-registered reports reported 96.05% significant results for their first hypothesis, the rate of significant results in registered reports was only 43.66%. Since registered reports are likely to be better powered, to show less questionable research practices, and to exhibit more methodological rigor due to rounds

of peer review prior to data collection – these numbers suggest a reduction of publication bias and/or a lower rate of false positives in registered reports (Scheel, Schijen & Lakens, 2020).

Our study joins Scheel and colleagues' findings, as we did not find support for our hypothesis. Rather than use null hypothesis testing, we opted to calculate Bayesian factors that would lend support to either our H0 or H1. For our main prediction, results support the null hypothesis, however, a second finding – that neither group showed expectations when compared to chance – makes these null results even harder to interpret. To guard against the possibility of uninterpretable results, we also preregistered a comparison with a previous study using unambiguous sentences, but otherwise similar in design to ours. The comparison between our study and the previous one was overall statistically inconclusive. We found that infants in our verb group were less likely to expect a verb than infants in the previous experiment, but that there was evidence that this was not the case for the noun groups. Simulations run by resampling from our data showed that increasing the sample size would increase the probability to find evidence that our results were different than those of the previous study, but those would remain noisy, with many samples still not showing this kind of evidence when comparing the noun conditions between experiments. The onset of registered reports might compel us as a field to be more accepting of such statistical grey areas. One study in itself is never sufficient to establish an underlying effect, even if well planned and powered - but revealing to the audience of developmental researchers the uncertainty of statistical assessment can promote cumulative science, for example in the form of meta analytic efforts (see for example Bergmann et al., 2018; Tsuji, Cristia, Frank, & Bergmann, 2019).

Author Contributions

All authors made a significant contribution to the study. M. Babineau and N. Havron developed the initial study concept, and along with A. Christophe developed the study design. M. Babineau programmed the experiment, and prepared the stimuli. Testing, data collection, data coding, and statistical analyses were performed by M. Babineau and N. Havron. N. Havron developed the statistical simulations. M. Babineau and N. Havron wrote the manuscript, and A. Christophe provided critical revisions. All authors approved the final version of the manuscript for submission.

Acknowledgements

This research was supported by postdoctoral fellowships to Naomi Havron from the French Embassy in Israel and the Victor Smorgon Charitable Fund, and to Mireille Babineau from the Fyssen Foundation and from the H2020 European Research Council [Marie Skłodowska-Curie grant No 799380). It was also supported by grants to Anne Christophe from the Fondation de France, and the French ANR (ANR-13-APPR-0012 LangLearn, ANR-17-CE28-0007-01 LangAge, and ANR-17-EURE-0017 FrontCog).

The authors thank Anne-Caroline Fiévet for her help in recruiting the children, and Léa Michel who assisted us in the stimuli preparation. We also thank all the parents who participated in this study.

Declaration of Conflicting Interests

The authors declare that they had no conflicts of interest.

References

- Arunachalam, S. (2017). Preschoolers' Acquisition of Novel Verbs in the Double Object Dative. *Cognitive science*, 41(S4), 831-854. <https://doi.org/10.1111/cogs.12368>
- Arunachalam, S., & Waxman, S. R. (2010). Meaning from Syntax: Evidence from 2-year-olds. *Cognition*, 114(3), 442–446. <https://doi.org/10.1016/j.cognition.2009.10.015>.
- Aslin, R.N. (2007). What's in a look? *Developmental science*, 10 (1), 48-53.
<https://doi.org/10.1111/j.1467-7687.2007.00563.x>
- Babineau, M., Christophe, A., & Shi, R. (in press). 14-month-olds exploit verbs' syntactic contexts to build expectations about novel words. *Infancy*.
- Beretti, M., Havron, N. & Christophe, A.(under review). Conflict resolution between syntactic and semantic cues in language acquisition.
- Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9), 3253-3258. <https://doi.org/10.1073/pnas.1113380109>
- Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax Constrains the Acquisition of Verb Meaning. *Language Learning and Development*, 3(33), 325–341.
<https://doi.org/10.1080/15475440701542609>
- Brusini, P., Amsili, P., Chemla, E. & Christophe, A. (2014). Simulation de l'apprentissage des contextes nominaux/verbaux par n-gramme. In: *Proceedings of 21ème Traitement Automatique des Langues Naturelles*, Marseille (France)
- Brusini, P., Amsili, P., Chemla, E., & Christophe, A. (2011). Learning to categorize nouns and verbs on the basis of a few known examples: a computational model relying on 2-word contexts. *Paper presented at society for research on child development biennial meeting*, Montreal (Canada).
- Cauvet, E., Limissuri, R., Millotte, S., Skoruppa, K., Cabrol, D., & Christophe, A. (2014). Function words constrain on-line recognition of verbs and nouns in French 18-month-olds. *Language Learning and Development*, 10(1), 1-18.
<https://doi.org/10.1080/15475441.2012.757970>
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272. <https://doi.org/10.1037/0033-295X.113.2.234>
- Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using

- ‘frequent frames’: what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental science*, 12(3), 396-406. <https://doi.org/10.1111/j.1467-7687.2009.00825.x>
- Christodoulopoulos, C., Roth, D., & Fisher, C. (2016). An Incremental Model of Syntactic Bootstrapping. In *Proceedings of the Seventh Workshop on Cognitive Aspects of Computational Language Learning*, 38-43.
- Christophe, A., Dautriche, I., de Carvalho, A., & Brusini, P. (2016). Bootstrapping the syntactic bootstrapper. *Proceedings of the 40th Annual Boston University Conference on Language Development*, 75–88.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36, 181–204.
<https://doi.org/10.1017/S0140525X12000477>
- Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). *Austin: University of Texas*.
- de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental science*, 19, 235-250.
- de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, 163, 67-79.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid Expectation Adaptation during Syntactic Comprehension. *PloS One*, 8(10), e77661.
<https://doi.org/10.1371/journal.pone.0077661>
- Fisher, C., Gertner, Y. , Scott, R. M. and Yuan, S. (2010), Syntactic bootstrapping. *WIREs Cogni Sci*, (1) 143-149. doi:10.1002/wcs.17
- Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond Associations: Sensitivity to structure in pre-schoolers’ linguistic predictions. *Cognition*, 157, 340–351.
- Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, 173, 351–370. <https://doi.org/10.1016/j.jecp.2018.04.012>
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73, 135–176. [https://doi.org/10.1016/S0010-0277\(99\)00036-0](https://doi.org/10.1016/S0010-0277(99)00036-0)

- Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, 1(1), 3–55. https://doi.org/10.1207/s15327817la0101_2
- Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, 18(5), 797–824. <https://doi.org/10.1111/infa.12006>
- Hallé, P. A., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and Speech*, 51(1), 23-44. <https://doi.org/10.1177/00238309080510010301>
- Huetig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118–135. <https://doi.org/10.1016/j.brainres.2015.02.014>
- Harrington Stack, C.M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & cognition*, 1-14. <https://doi.org/10.3758/s13421-018-0808-6>
- Havron, N., Bergmann, C., & Tsuji, S. (2020, February 25). Preregistration in infant research – a primer. *PsyArxiv*, <https://doi.org/10.31234/osf.io/es2gx>
- Havron, N., de Carvalho, A., Fiévet, A. C. & Christophe, A. 3-4-year-old children rapidly adapt their predictions and use them to learn novel word meanings. *Child Development*. Advance online publication. doi:10.1111/cdev.13113
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73(2), 418–433. <https://doi.org/10.1111/1467-8624.00415>
- JASP Team (2018). JASP (Version 0.9) [Computer software].
- Jin, K., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. *Proceedings of the 38th Annual Boston University Conference on Language Development*.
- Kedar, Y., Casasola, M., Lust, B., & Parnet, Y. (2017). Little Words, Big Impact: Determiners Begin to Bootstrap Reference by 12 Months. *Language Learning and Development*, 13(3), 317-334. <https://doi.org/10.1080/15475441.2017.1283229>
- Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, 120(11), 2546–2556. <https://doi.org/10.1016/j.lingua.2010.06.005>

- Liu, L., Burchill, Z., Tanenhaus, M. K., & Jaeger, T. F. (2017). Failure to replicate talker-specific syntactic adaptation. In *CogSci*.
- Männel, C., & Friederici, A. (2008). Event-related brain potentials as a window to children's language processing. In I. A. Sekerina, E. M. Fernández, H. Clahsen (Eds.) *Developmental Psycholinguistics: On-line methods in children's language processing* (pp 29-72). John Benjamins Publishing Company, Amsterdam, the Netherlands.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91-117. [https://doi.org/10.1016/S0010-0277\(03\)00140-9](https://doi.org/10.1016/S0010-0277(03)00140-9)
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26(4), 393-424. https://doi.org/10.1207/s15516709cog2604_1
- Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, 22(4), 436-469.
- Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M., & Poulin-Dubois, D. (2011). Early verb learning in 20-month-old Japanese-speaking children. *Journal of Child Language*, 38(3), 455-484. <https://doi.org/10.1017/S0305000910000127>
- Scheel, A. M., Schijen, M., & Lakens, D. (2020, February 5). An excess of positive results: Comparing the standard Psychology literature with Registered Reports. *PsyArxiv*, <https://doi.org/10.31234/osf.io/p6e9c>
- Schönbrodt, F. D., Zehetleitner, M., & Perugini, M. (2015). Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences, *Psychological Methods*, 22(2), 322-343. <https://doi.org/10.1037/met0000061>
- Saint Clair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition*, 116(3), 341-360. <https://doi.org/10.1016/j.cognition.2010.05.012>
- Shi, R., & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, 11(3), 407-413. <https://doi.org/10.1111/j.1467-7687.2008.00685.x>
- Shi, R., & Melançon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy*, 15(5), 517-533. <https://doi.org/10.1111/j.1532-7078.2009.00022.x>

- Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences*, *108*(15), 6038–6043. <https://doi.org/10.1073/pnas.1017617108>
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*(2), 249–267. [https://doi.org/10.1016/S0749-596X\(03\)00024-X](https://doi.org/10.1016/S0749-596X(03)00024-X)
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive Top-Down Integration of Prior Knowledge during Speech Perception. *Journal of Neuroscience*, *32*(25), 8443-8453. <https://doi.org/10.1523/JNEUROSCI.5069-11.2012>

Table 1. Stimuli

Habituation	
Verb condition	Noun condition
Target contexts (x20 in total):	
La petite crale (x12)*	La petite crale (x12)*
La petite dort "The little one is sleeping" (x 2)	La petite balle "The little ball" (x2)
La petite donne "The little one is giving" (x 2)	La petite chaise "The little chair" (x2)
La petite mange "The little one is eating" (x 2)	La petite voiture "The little car" (x2)
La petite regarde "The little one is looking" (x 2)	La petite pomme "The little apple" (x2)
Filler contexts (x12 in total)**:	
Je dors "I sleep"	
Je donne "I give"	
Je mange "I eat"	
Je regarde "I look"	
Elle dort bien "She is sleeping well"	
Elle donne bien "She is giving well"	
Elle mange bien "She is eating well"	
Elle regarde bien "She looks well"	
Des balles "Some balls"	
Des pommes "Some apples"	
Des voitures "Some cars"	
Des chaises "Some chairs"	
C'est ma balle "It's my ball"	
C'est ma pomme "It's my apple"	
C'est ma voiture "It's my car"	
C'est ma chaise "It's my chair"	
Test	
Noun context: une crale	
vs.	
Verb context: tu crales	

*The same sound files are used in both habituation groups

** In each habituation trial, there are 12 filler contexts in which the familiar words are preceded by function words. For the Verb condition, there will be eight noun contexts and four verb contexts, whereas for the Noun condition, there will be eight verb contexts and four noun contexts.

Table 2. Average acoustic values (and standard deviations) across multiple exemplars of *la petite* contexts

Acoustic measures	Verb contexts	Noun contexts	Nonword context
Total utterance duration (ms)	1503 (62)	1449 (112)	1509
Target word duration (ms)	796 (115)	805 (178)	822
Target word mean pitch (Hz)	279.2 (22.7)	284.4 (4.5)	275.2
Target word mean intensity (dB)	66 (2.1)	65.9 (1.6)	64.7

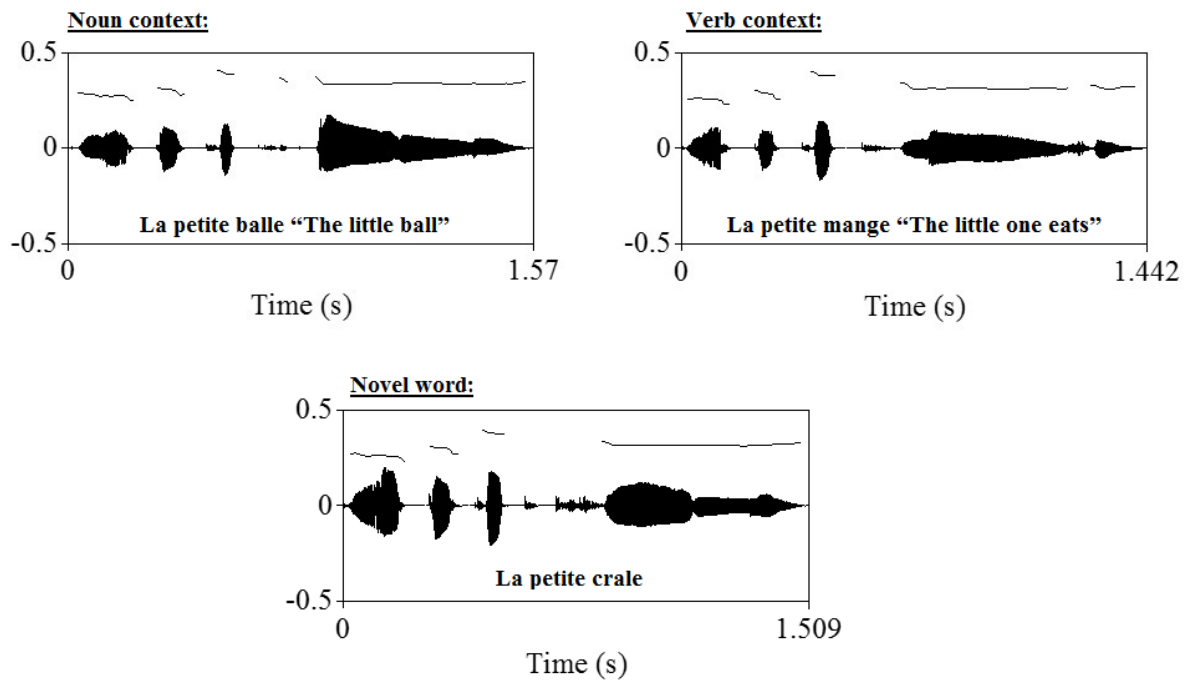


Figure 1. Example of waveforms and pitch contours for familiar verb and noun contexts, along with those for the novel word.

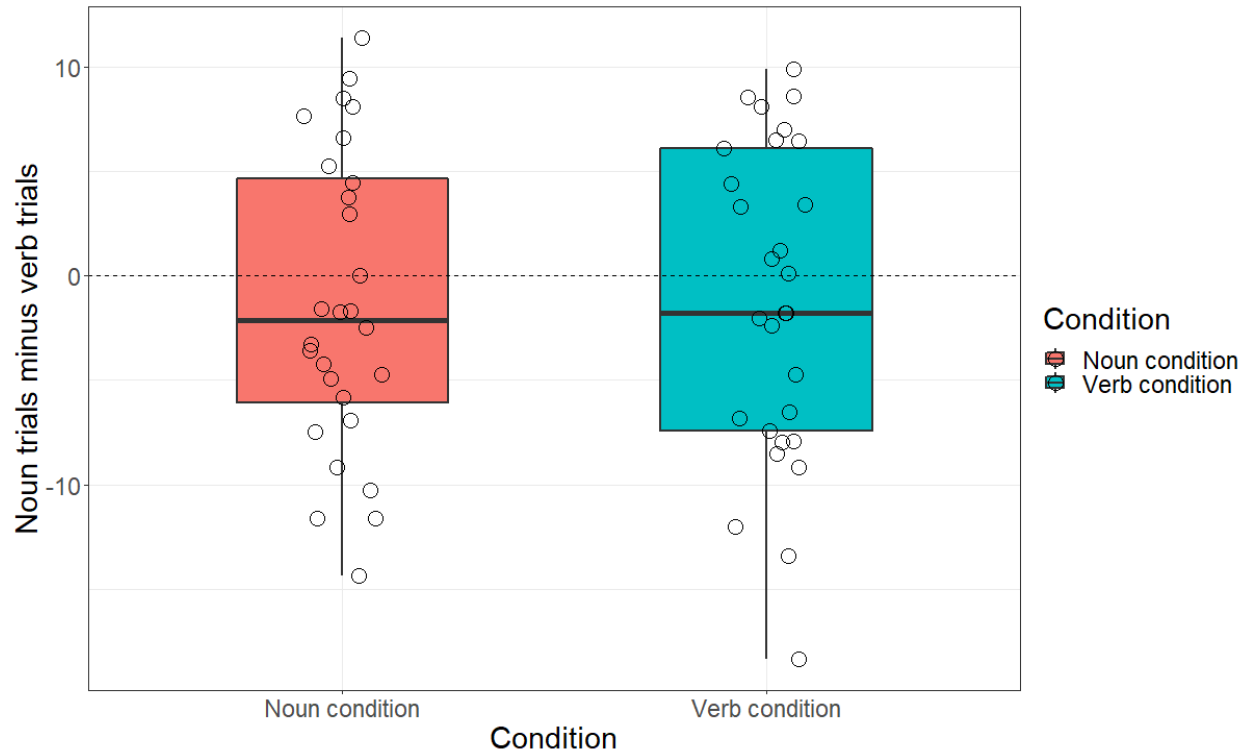


Figure 2. Boxplot of the difference between listening times during the noun and the verb test trials for each condition. The black line represents the median. Each dot represents one participant. A positive value means longer listening times for noun trials.

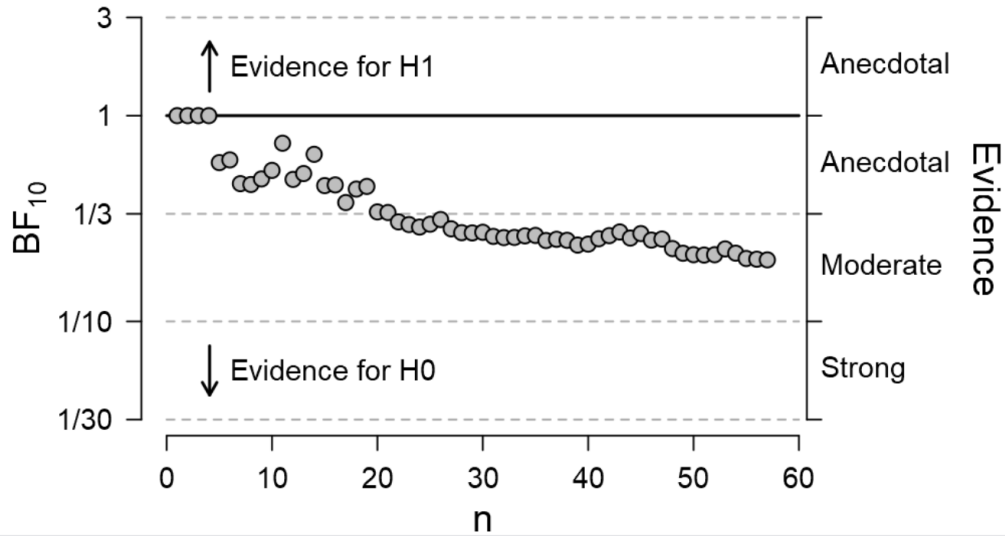


Figure 3. Sequential analysis for the interaction between conditions. Each dot represents one participant, ordered by order of testing. Evidence for H1 steadily decreases as the sample size increases.