HAL open science

Learning to predict and predicting to learn: Before and beyond the syntactic bootstrapper

Mireille Babineau, Naomi Havron, Isabelle Dautriche, Alex de Carvalho, Anne Christophe

▶ **To cite this version:**

Mireille Babineau, Naomi Havron, Isabelle Dautriche, Alex de Carvalho, Anne Christophe. Learning to predict and predicting to learn: Before and beyond the syntactic bootstrapper. Language Acquisition, 2022, pp.1-24. 10.1080/10489223.2022.2078211 . halshs-03815072

HAL Id: halshs-03815072

https://shs.hal.science/halshs-03815072v1

Submitted on 14 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Learning to predict and predicting to learn: Before and beyond the syntactic bootstrapper

Mireille Babineau[a*], Naomi Havron[b*], Isabelle Dautriche[c,d],
Alex de Carvalho[e], and Anne Christophe[f,g]

(***To appear in Language Acquisition***)

a. Department of Psychology, University of Toronto, Canada

b. School of Psychological Sciences, University of Haifa, Israel

c. Laboratoire de Psychologie Cognitive, Aix-Marseille University, CNRS, Marseille, France

d. Institute of Language, Communication and the Brain, Aix-Marseille University, CNRS, Aix-en-Provence, France

e. Laboratoire de Psychologie du Développement et de l'Éducation de l'Enfant, CNRS, Université de Paris, Paris, France

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

g. Maternité Port-Royal, AP-HP, Faculté de Médecine Paris Descartes, France

Corresponding authors: correspondence concerning this article should be addressed to Mireille Babineau or Naomi Havron. Emails: mireille.babineau@utoronto.ca, nhavron@psy.haifa.ac.il; Address of Dr. Babineau: Department of Psychology, 100 St. George Street, Sidney Smith Hall, Toronto, Canada, M5S 3G3; Address of Dr. School of Psychological Sciences, University of Haifa, 199 Aba Khoushy, Haifa, Israel.

Abstract

Young children can exploit the syntactic context of a novel word to narrow down its probable meaning. This is *syntactic bootstrapping*. A learner that uses syntactic bootstrapping to foster lexical acquisition must first have identified the semantic information that a syntactic context provides. Based on the *semantic seed hypothesis*, children discover the semantic predictiveness of syntactic contexts by tracking the distribution of familiar words. We propose that these learning mechanisms relate to a larger cognitive model: the *predictive processing framework*. According to this model, we perceive and make sense of the world by constantly predicting what will happen next in a probabilistic fashion. We outline evidence that prediction operates within language acquisition, and show how this framework helps us understand the way lexical knowledge refines syntactic predictions, and how syntactic knowledge refines predictions about novel words' meanings. The predictive processing framework entails that learners can adapt to recent information and update their linguistic model. Here we review some of the recent experimental work showing that the type of prediction preschool children make from a syntactic context can change when they are presented with contrary evidence from recent input. We end by discussing some challenges of applying the predictive processing framework to syntactic bootstrapping and propose new avenues to investigate in future work.

1. Introduction

When acquiring their native language, children have to learn its components: among others, the sounds (i.e. phonology), the words (i.e. lexicon) and the way in which words can be combined to form sentences (i.e. syntax). For a long time, children were thought to acquire these components sequentially. This corresponds to what young children produce: first they babble, then they speak in isolated words, and then they start combining words together. However, research into language development draws a different picture—children tackle language acquisition from all fronts at the same time. That is, young children are immersed in all components of their native language(s) and can use partial knowledge in one domain to help refine their representations in another. These synergies have been observed between different domains. For instance, even a very crude proto-lexicon can help the learner acquire a phonological system (Feldman et al., 2013; Martin, Peperkamp, & Dupoux, 2013). In the same vein, young children use syntax to guide word learning (*syntactic bootstrapping*; e.g., Fisher, Jin, & Scott, 2020; Gleitman, 1990; Gleitman et al., 2005; Landau & Gleitman, 1985).

The aim of this paper is to present a theory of how language learning gets off the ground by relying on such synergies between lexical and syntactic acquisition. In particular, we propose that children's early lexicon and their reliance on a predictive processing mechanism can explain how their early syntactic knowledge develops in tandem with their lexical knowledge. We propose that syntactic bootstrapping fits within a broader framework of human cognition: *predictive processing*. According to this framework, human reasoning, learning, planning, and action rely on our brain operating as a "prediction machine" (e.g., Clark, 2013). We survey evidence that prediction also operates within language acquisition and propose that this framework can explain how lexical knowledge gives rise to, and helps refine, syntactic predictions, as well as how syntactic knowledge gives rise to predictions about novel words' meanings. We start off by surveying syntactic bootstrapping research and go on to ask how syntactic bootstrapping gets started—first, by presenting recent evidence supporting the idea that a few early-learned nouns and verbs may help bootstrap syntactic bootstrapping (the semantic seed hypothesis), and then by

surveying evidence for predictive processing in language acquisition and theorizing how these might work together with the semantic seed and syntactic bootstrapping.

2. Syntactic bootstrapping and function words

In order to make predictions about upcoming words, children have been shown to follow a *syntactic bootstrapping* approach, i.e. they can use information present in the syntactic structures of sentences, such as the number and type of arguments, to identify words and infer their probable meanings (e.g., Fisher, Gertner, Scott, & Yuan, 2010; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Trueswell, Papafragou, & Choi, 2011). In its simplest form, this mechanism can rely on co-occurrence patterns and reliable adjacent elements such as function words and morphemes (e.g., determiners, pronouns, auxiliaries, tense endings). Function words' acoustic, distributional, and phonological characteristics enable infants to distinguish them from content words (e.g., nouns, verbs, adjectives) during their first few months of life (Shi & Werker, 2001; 2003; Shi, Werker, & Morgan, 1999). Function words can be particularly useful during early lexical acquisition because they tend to co-occur with specific categories of content words (Chemla, et al. , 2009; Mintz, 2003; Weisleder & Waxman, 2010). For instance, in languages such as French and English, determiners tend to precede nouns, and pronouns tend to precede verbs.

Infants have been shown to identify function words and learn about their potential for lexical acquisition early in development (see e.g., Shi, 2014). Function words generally tend to appear at the boundaries of prosodic units (see e.g., Christophe, et al., 2003; Dryer, 1992; Gutman, et al., 2015; Shi, Morgan, & Allopenna, 1998)–a position that has been shown to be advantageous for encoding word units (Ferry et al., 2016; Johnson, Seidl, & Tyler, 2014; Shukla, Nespor, & Mehler, 2007). Even before their first birthday, infants store and recognize the phonological form of their native language's frequent function words (e.g., Hallé, Durand, & de Boysson-Bardies, 2008; Shafer, Shucard, Shucard, & Gerken, 1998; Shi, Cutler, Werker, & Cruickshank, 2006; Shi, Werker, & Cutler, 2006; see Shi, 2014, for a review), and use them to segment co-occurring content words (Shi & Lepage, 2008). They also learn how these different categories of words relate to one

another. In a recent study (Babineau & Christophe, 2022), infants as young as 11 months were found to track co-occurrence patterns between function words and common content words. French-learning 11-month-olds preferred to listen to lists of grammatically correct phrases in which familiar verbs (e.g., *mange* 'eat') occurred after the pronoun *tu* 'you', and familiar nouns (e.g., *biberon* 'bottle') occurred after the determiner *des* 'some', than to lists of ungrammatical phrases in which verbs followed a determiner and nouns followed a pronoun. Around 14 months of age, infants can build expectations about the type of function words which are likely to co-occur with novel content words based on the contexts in which they have previously been heard (Babineau et al., 2020; Höhle et al. 2004; Shi & Melançon, 2010): A novel word that co-occurs with determiners is expected to also be used with other determiners, not pronouns. Function words may thus help build a kind of proto-syntactic knowledge, which includes the creation of coarse syntactic categories.

This kind of proto-syntactic knowledge may well be the key to unlocking much of the language acquisition problem. Since speech, both infant- and adult-directed, is for the most part made up of sentences rather than isolated words (Aslin, 1993; Brent & Siskind, 2001; van de Weijer, 1998), infants must learn the meanings of many content words from hearing them in context. The syntactic context of a novel word, such as the co-occurring function word and morpheme, can serve as a "zoom lens" allowing learners to figure out which part of the world is being talked about, and thus narrowing down its possible meaning (e.g., Fisher et al., 1994). When presented with a novel word occurring in a noun context, such as *it is a blick*, 12- to-14-month-olds assume that it refers to an object (e.g. Waxman, 1999; Waxman & Booth, 2001). Toddlers, 18 months and older, also assume that *blick* refers to an action if it is heard in a verb context, such as "he's blicking" (Bernal et al., 2007; de Carvalho et al., 2019; He & Lidz, 2017; Oshima-Takane et al., 2011; Waxman et al., 2009)[1]. While the majority of syntactic bootstrapping studies have focused on English and other indo-European languages, a smaller number have looked at infants learning other languages, such as Mandarin Chinese (Lee & Naigles, 2005), Turkish

[1] Infants also exploit the syntactic structure in which novel verbs appear to constrain their specific meaning – for example mapping verbs appearing in transitive structures (e.g., "the girl is blicking the boy") to causal actions (Arunachalam & Waxman, 2010; Dautriche et al., 2014; de Carvalho et al., 2021; Fisher et al., 2010; Jin & Fisher, 2014; Yuan & Fisher, 2009; Yuan et al., 2012).

(Göksun, Küntay, & Naigles, 2008) and Icelandic (Nowenstein, Sigurjónsdóttir, Yang, Ingason, & Wallenberg, 2020). In the last two languages, researchers not only looked at syntactic but also at morphological bootstrapping. Children were able to use both case marking, which is language specific, and the number of arguments in a sentence, which is said to be universal - to infer novel words' meaning.

Inferences about a novel word's syntactic category and lexical features can happen in real time after only two repetitions (de Carvalho, Babineau, Trueswell, Waxman, & Christophe, 2019). The same is true for lexical access to familiar words: when 18-month-olds are presented with familiar object names in a sentence frame, they are faster to orient to the target picture than when the same names are presented in isolation. This suggests that familiar frames may enable the infant to look for, or predict, the end of the sentence more efficiently (Fernald & Hurtado, 2006). Likewise, Zangl and Fernald (2007) and Kedar and colleagues (2018) found that replacing real determiners with nonce determiners disrupted 12- to-24-month-olds' processing of familiar words.

Children's sensitivity to the set of function words that co-occur with specific content words may also be greatly beneficial for learning homophones (words that share the same phonological form but have different meanings and that, more often than not, appear in different syntactic contexts; Dautriche, Fibla, Fiévet, & Christophe, 2018)[2]. Suppose, for instance, that a child already knows one of the meanings of a homophone; for example, the noun meaning of the homophone *lit* (which can mean either "a bed" or "to read" in French). If the child hears this same word-form for the first time in a verb context, e.g., preceded by a pronoun (such as *elle lit* "she's reading"), she can link an additional meaning to the word-form (referring to an action). Indeed, recent work has shown that French-learning 20-month-olds easily learn homophones when the two meanings belong to different syntactic categories (Dautriche, Fibla, Fiévet, & Christophe, 2018). Overall, these results point to the possibility that a learner's knowledge about the distribution of word categories can foster lexical acquisition.

---

[2]There is evidence in the literature showing that infants have already acquired a certain number of homophones around 20 months of age (see e.g., Dautriche et al., 2015; de Carvalho et al., 2017).

3. Bootstrapping the syntactic bootstrapper

Although a large body of evidence shows that young children are able to use the syntactic context of a novel word to infer something about this word's potential meaning, what remained unclear until recently is how they learn which syntactic contexts correspond to which semantic features – for instance, how do children figure out that words occurring in noun contexts can refer to objects, and how do they learn the characteristics of noun contexts in their language? Based on the *semantic seed hypothesis*, children learn these associations between syntactic contexts and semantics by generalizing from a handful of words whose meaning they already know - a semantic seed. After grouping the small inventory of known words into coarse semantic categories, they then learn which syntactic elements tend to occur with these familiar words (Brusini, Seminck, Amsili, & Christophe, 2021; Christophe, Dautriche, de Carvalho, & Brusini, 2016; Gutman, Dautriche, Crabbé, & Christophe, 2014, see also Christodoulopoulos, Roth, & Fisher, 2016). When do infants start to build these connections between lexical and syntactic knowledge? How do infants use their lexicon and predictive abilities to bootstrap syntactic bootstrapping?

The semantic seed hypothesis is based on empirical findings attesting children's early lexical knowledge and categorization abilities. By the end of their first year of life, infants have already managed to build a small lexicon through rich social and multimodal experiences (e.g., Bergelson & Swingley, 2012, 2013, 2015; Parise & Csibra, 2012; Syrnyk & Meints, 2017; Tincoff & Jusczyk, 1999, 2012). This blossoming lexicon could be exploited as a seed for assigning novel words into syntactic categories, as infants could notice the most frequent syntactic contexts in which different types of seed words tend to appear. Many studies have shown that infants have the ability to track such regularities (e.g., Gomez & Gerken, 1999; Hohle, Schmitz, Santelmann, & Weissenborn, 2006; Marchetto & Bonatti, 2013; Saffran et al., 2008; Santelmann & Jusczyk, 1998; van Heugten & Johnson, 2010). Statistical learning skills allow children to compute what sequences of sound form the morphemes and words of the language, in what syntactic contexts these elements tend to appear in, how consistently sounds co-occur, how frequently words occur in similar contexts, etc. (see e.g., Newport, 2016; Saffran, Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996). Thus, using their distributional learning skills, infants could track

the syntactic contexts of the semantic seed (e.g., 'ball' and 'car' appear in "This is a ball", "This is a car"), learn which kind of words tend to appear in these contexts, and later infer that a novel word appearing in the same context is likely to share some properties with these known words (e.g., "This is a"+"dax" -> dax = object). This follows the established proposal that young children expect words that share conceptual or semantic properties to occur in similar syntactic contexts (Gleitman, 1990; Pinker, 1984). These expectations are likely to be subjected to inter-individual differences linked to the breadth of syntactic knowledge. For instance, Wojcik and Saffran (2015) found that 26- to 28-month-olds with more advanced syntactic knowledge (i.e. those who were combining words in their own speech productions) were more likely to encode relations between novel words occurring in the same sentential position (as subject/agent in the training sentences). Given the research findings described above and the current state of knowledge in infancy research, the semantic seed hypothesis describes plausible synergies in early syntactic and lexical acquisition that are supported by recent computational models and experimental work involving young children and adults.

Several computational models have provided more direct evidence supporting the main assumptions of the *semantic seed* hypothesis: 1) the context of a word contains relevant information enabling its categorization, 2) categorization is efficient when it is based on a handful of words that are already known (a seed). Firstly, many unsupervised computational models using the context of words to categorize them have obtained better-than-chance performance (Chemla et al., 2009; Chrupała & Alishahi, 2010; Mintz, 2003; Parisien et al., 2008; Redington et al., 1998; Wang et al., 2011; Weisleder & Waxman, 2010). Secondly, in Brusini et al. (2021), a very simple algorithm extracting 2-word contexts of syntactically-categorized content words (the seed) was sufficient to categorize test words with high precision. In fact, even a limited semantic seed size (8 nouns and 1 verb) provided during the training phase allowed ~90% of new nouns and ~80% of new verbs to be correctly categorized at test. These findings indicate that even with a small lexicon, young language learners could use familiar words to start forming proto-syntactic categories.

Based on Brusini et al.'s (2021) model, children might discover the predictiveness of function words by tracking the immediate context of familiar words. To demonstrate experimentally that young children can indeed learn proto-syntactic categories this way, Babineau et al. (2021) familiarized 3- to 4-year-olds to a seed lexicon co-occurring with a novel syntactic context—the novel function word "ko". Participants watched a training video in which the novel function word "ko" either replaced pronouns (*verb condition*) or articles (*noun condition*) in sentences containing familiar words in French (see Figure 1).

Training video

Noun condition:	**Verb condition:**
Ko *cochon, il veut encore jouer avec* ***ko*** *jolie tortue! Il fait rouler* ***ko*** *balle comme ça.* "**Ko** pig, he wants to play again with **ko** pretty turtle! He rolls **ko** ball this way".	*Le cochon,* ***ko*** *veut encore jouer avec la jolie tortue!* ***Ko*** *fait rouler la balle comme ça.* "The pig, **ko** wants to play again with the pretty turtle! **Ko** rolls the ball this way".

Figure 1. Example of the script used in the two versions of the training video. The same script was used in both conditions, with "ko" replacing pronouns in the verb condition and articles in the noun condition.

After this training phase, participants were shown two videos side-by-side: one depicting an actor holding a novel object, and the other depicting an actor performing a novel action while hearing novel content words co-occurring with the novel function word (e.g. Oh look, ko dase!). Participants in the verb condition looked and pointed significantly more towards the novel actions than participants in the noun condition (see Figure 2). This

study showed that 3- to-4-year-olds managed to rapidly and efficiently undertake a distributional analysis to learn the semantic predictiveness of a novel function word– providing preliminary experimental support for the semantic seed hypothesis.

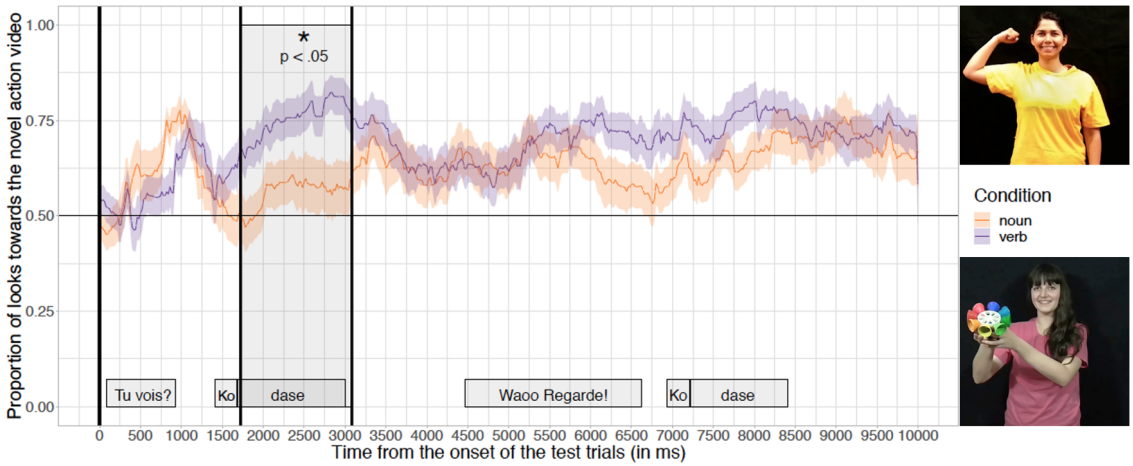

Figure 2. Figure taken from Babineau et al. (2021). Proportion of looks toward the action video, time-locked to the onset of the trial for the verb condition (purple curve) and the noun condition (orange curve). Error bars represent the standard error of the mean. A nonparametric cluster-based permutation test (Maris & Oostenveld, 2007) revealed significant differences between the verb and the noun conditions starting slightly after the onset of the novel function word "ko" (grey time-window). The novel function word "ko" appeared about 1413ms after the onset of the trial (307ms before the significant time-window) and appeared again at 6933ms.

This first attempt of Babineau and colleagues to obtain direct evidence for the learning mechanism that bootstraps the syntactic bootstrapper has two notable limitations: First, this study tested preschoolers while the semantic seed hypothesis entails that this learning strategy would be most useful during the second year of life— when infants have already acquired a restricted, but large-enough, lexicon (e.g., Bergelson & Swingley, 2012) and the necessary computational skills to track distributional patterns (e.g., Gomez & Gerken, 1999; Saffran et al., 2008). Second, the novel functor's position and use was identical with those of familiar function words in French. That is, the novel function word "ko" simply replaced existing function words and could have been treated as such. In order to obtain more compelling evidence in support of the semantic seed hypothesis, the learnability of other morphosyntactic regularities needs to be tested. For instance, noun class systems are widespread across the world's languages, with semantic features of nouns

(natural gender, animacy, shape) often being associated with the class (Aikhenvald, 2000; Denny, 1976; Dixon, 1986). Inspection of cross-linguistic universals reveals that across the world's languages, certain conceptual-categories distinctions are often marked in morphosyntax (e.g., object/action, animate/inanimate), while others are not (e.g., electronic/non electronic; Strickland, 2017). It has been argued that this is because morphosyntactic regularities in human languages would more likely be associated with conceptual distinctions that are clearly salient to infants and young children, such as *core* conceptual categories (Spelke & Kinzler, 2007). Hence, novel morphosyntactic regularities should be learnable as long as they are marking *core knowledge* categories.

Animacy, as a conceptual property, is a highly salient property that is detected early in life by infants (e.g., Molina et al., 2004; Rakison & Poulin-Dubois, 2001; Rostad et al., 2012). Past work has shown that toddlers use animacy-based characteristics of familiar content words to guide their inferences about the meaning of novel words. Ferguson et al. (2014) found that 19-month-olds looked more at a stuffed animal than at an artifact when hearing a sentence in which the novel noun "dax" is the subject of the animate-selecting construction "the dax is crying". Hence, infants are able to use a verb's selectional restrictions (here, that only animates can cry) to infer whether a novel noun is animate or inanimate. Similarly, Yuan et al. (2011) found that 2-year-olds who heard the novel verb "stipe" co-occurring with known animal names such as "pig" associated it with animacy. At test, when asked "Which one can stipe?", they looked more at a picture of a familiar animal (e.g. a cat) than at an inanimate object (e.g. a spoon). Here, children were shown to use a semantic seed (a group of familiar animal names) to infer the novel verb's selectional restrictions (that only animates can "stipe"). What is missing from these studies, however, is a demonstration of whether and how infants can use familiar content words to learn about new function words and their role as morphosyntactic markers of animacy.

These issues were addressed in a recent study with 20-month-old infants. Barbir et al. (2019) tested French-learning infants' ability to learn a novel noun class system grounded in semantic properties that are not syntactically encoded in their native language, i.e. animate/inanimate, but are correlated with the specific syntactic properties of noun

phrases in many languages (e.g., Algonquian languages, Bantu languages)[3]. Participants had to learn that a novel determiner "ko" was restricted to occur before animate agents, while the novel determiner "ka" appeared only before inanimate objects (e.g., *ko chien*, *ka poussette*, "ko dog", "ka stroller"). As in Babineau et al. (2021), the novel function words were introduced using a short training video, in which an actress played with toy animals and objects while uttering sentences using the novel determiners 'ko' and 'ka' appropriately before animate or inanimate common nouns (but otherwise using French syntax, prosody, and lexicon). Since participants were younger, a longer exposure was offered, and families were instructed to show the training video at home on three successive days before coming to the lab. During the test phase in the lab, infants watched two pictures side-by-side (one depicting an animate agent and the other inanimate) and listened to sentences containing novel words preceded by either "ko" or by "ka". Infants looked more towards the novel animal when hearing novel words embedded in sentences such as "look at *ko bamoule*!" than when hearing sentences such as "look at *ka pirdale*!" (see Figure 3). These findings provide strong evidence in support of the semantic seed hypothesis. They are even more remarkable when the size of the semantic seed provided to the participating infants in this task is taken into account: 6 animates with "ko" and 6 inanimates with "ka".

The computational and experimental work reviewed above strongly supports the semantic seed hypothesis, thereby confirming young children's ability to generalize from a small set of exemplars, which is highly beneficial for both syntax and vocabulary development. This process of generalization does not manifest itself only in categorization but also in the ability to rapidly build representations of predictive contexts, with "ko" and "ka" predicting this newly cued category distinction between animates and inanimates. Infants here can be said to show an urge to understand their linguistic input and generate expectations/predictions based on their limited state of knowledge.

[3] Several languages, such as Japanese, Thai, and Vietnamese, also have a classifier system used to signal the animate properties of co-occurring words (e.g., Adams & Conklin, 1973).

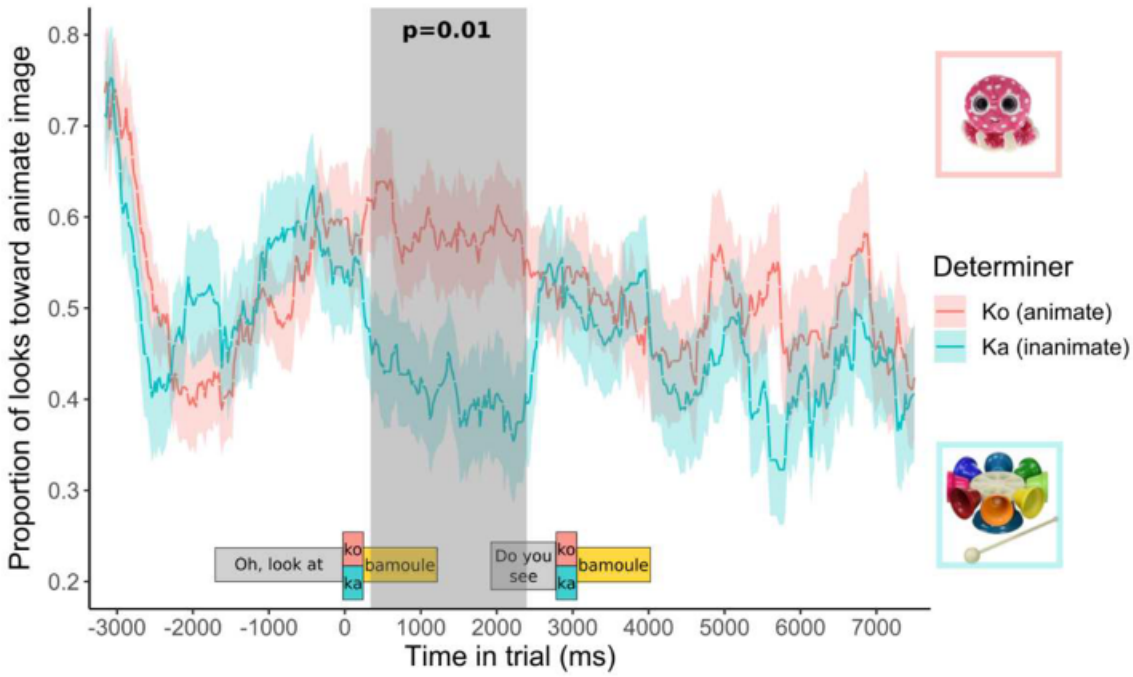

Figure 3. Figure taken from Barbir et al. (2019). Proportion of looks to the novel animate image, when "ko" (pink line) and "ka" (blue line) preceded the novel words at test. A cluster-based permutation analysis revealed that infants looked significantly longer toward the animate agent when hearing sentences with "ko" than when hearing sentences with "ka" at test, between 300ms and 2440ms (p = .01).

4. Bootstrapping and linguistic predictions

We have seen above how young children can learn new syntactic contexts as well as how they successfully use them to create novel expectations and constrain their interpretation of novel words. They can, therefore, be considered as predictive learners (see Chang et al., 2016). What children seem to be learning here, and incredibly fast, are statistical regularities in their input. Distributional learning is a key, yet simple, mechanism present from birth: Even newborns have been shown to segment the speech stream (i.e., finding the boundaries between words) by tracking distributional cues (e.g., Fló et al., 2019). However, the regularities that structure input to infants do more than signal the boundaries of words -they can be used to narrow down the number of possibilities for the meaning of words as well. Processing syntactic contexts thus allows infants to *predict* certain things about their language. When infants listen to and process linguistic input,

certain expectations arise even before complete bottom-up information becomes available (see Lau, et al., 2013 for a discussion). These representations may be, among other things, phonological (e.g., faster selection of vowel-initial words following the English determiner *an,* Gambi et al., 2018) or semantic (e.g., expectation of an edible object following the word "eat", Mani & Heuttig, 2012; or of a causative meaning for a verb occurring in a transitive construction, Lee & Naigles, 2008).

To understand the significance of the idea of prediction, we need to take a broader look beyond the field of language acquisition to the idea of predictive processing in cognitive and brain science. Predictive processing has been proposed to underlie the way our brain operates in a variety of domains (e.g., Clark, 2013). According to the predictive processing framework, we perceive and make sense of the world by constantly predicting what will happen next in a probabilistic fashion using our current best model of the world around us. When these predictions fail, lower networks, for example, those of perception, "trickle up", affecting increasingly higher levels of cognition and planning - serving at each level to update our model of the world for future predictions. Viewed through the lens of language acquisition, it has been proposed that children learn language through prediction and error-based learning (e.g., Chang, Dell & Bock, 2006; Dell & Chang, 2014; Elman, 1990; Zettersten, 2019). Children's best model of their linguistic environment leads them to create predictions about everything from the phonological form of the next word in a sentence (e.g., 3-5 year olds already start to look more at the referent of a word beginning with a vowel after hearing the determiner *an* than after the determiner *a*, though this skill continues to improve until adulthood, Gambi et al., 2018) to its meaning (e.g., Gambi et al., 2018), syntactic category (e.g., Cauvet et al., 2014; Shi & Melançon, 2010), and grammatical gender (e.g., Lew-Williams & Fernald, 2007, 2010; Van Heugten & Johnson, 2011). When these predictions fail, infants experience an error response - the strength of which correlates with children's vocabulary scores (Fernald, et al., 2006; Marchman & Fernald, 2008; Reuter et al., 2018; Ylinen, et al., 2017). This suggests that stronger predictive processing skills enable infants to better make use of these word learning situations (see e.g., Reuter et al., 2019; though the reverse explanation - that more robust linguistic experience supports predictive processing - will be discussed below). In adults, stronger violation of expectations leads to stronger abstract structural knowledge of an

artificial language (Bovolenta & Marsden, 2020), though this has not yet been tested in children or infants.

Rabagliati et al. (2016) cast doubt on whether, at the time of writing their article, evidence existed to support the role of prediction in language learning, or whether the then-current literature should have been interpreted as only showing skilled language users' ability to use prediction (having already learnt to predict). They presented two conditions that would need to be met in order to show this effect: (1) that children make the types of prediction that are necessary for learning, and (2) that prediction is a cause and not (only) an effect of language learning. They believed that these requirements, which support the claim that prediction might propel language acquisition, were yet to be met. We would argue that the years that have passed since Rabagliati et al.'s (2016) publication have changed this picture. The studies presented in the last paragraph seem to already satisfy these conditions: Gambi et al., (2018) showed semantic prediction (e.g., "Can you see one...ball/two...ice creams?") in 2- to-5-year-olds - which would be one type of prediction relevant for learning, and Ylinen et al. (2017) showed that error signals are present in 12-month-olds when prediction about the end of a common word fails. Another recent study (Gambi et al., 2021) found that 2- to-5-year-olds were able to make graded predictions (not only predicting the most likely ending for a sentence, but also considering other, less likely, options) and that this ability improved with age. Interestingly, children's ability to make these graded predictions, as well as their processing speed, predicted their vocabulary scores when tested again after 7 months. What remains to be shown is children's ability to adapt their predictions, and that prediction is directly related to language learning.

A recent study examined whether young children could rapidly adapt their syntactic predictions and use their adapted predictions to infer the intended meaning of homophones. Havron et al. (2021) exposed 3-to-4-year-olds to a syntactic context that can precede either nouns or verbs in French with the goal of manipulating the expected syntactic category. In French, *la petite* can be followed either by a noun (*la petite pomme*, "the little apple") or by a verb (*la petite mange*, 'the little [one] is eating'). In the first phase of the experiment, this syntactic context was made highly predictive of different categories for different groups of participants. Children in the noun group heard sentences where *la petite* was

followed by familiar nouns and children in the verb group heard sentences where *la petite* was followed by familiar verbs. At test, children were exposed to familiar noun-verb homophones following *la petite* (e.g., *la petite ferme* can be interpreted as either "the little farm" or "the little [one] closes"). Children for whom *la petite* preceded familiar verbs in the first phase of the experiment were more likely to interpret the homophone as a verb than children for whom it preceded familiar nouns in the training phase (see also Havron, Scaff, Carbajal, Linzen & Christophe, 2020 for an experiment with 5- to 6-year olds which looked at the resolution of another type of global ambiguity - attachment ambiguity).

A second study directly examined whether 3-to-4-year-olds could update the model of their linguistic environment to generate predictions about the meaning of *novel words*. Havron, et al. (2019) used the same preceding context, *la petite*, but this time, at test, it was followed by a novel word (e.g., *dase*) that could be interpreted as either a noun or a verb in French (e.g., *Tu vois ? La petite dase !* "Do you see? The little dase/the little [one] dases!"). Children for whom *la petite* preceded familiar verbs in the first phase of the experiment were more likely to infer that a novel word appearing in this context referred to an action than children for whom it preceded familiar nouns. Thus, young children were shown to rapidly update their models of what category of words should follow certain syntactic contexts and use their updated model to generate predictions immediately - to guide their learning of novel words. In this experiment, prediction structured and guided preschoolers' inferences about a novel word's meaning. This study was replicated twice with French and English-speaking children (see Swanson & Frank, 2021; Yu et al., *in preparation*). Given that there was no evidence of a difference in effect size in the two experiments from Havron and colleagues (i.e., the one looking at homophones and the other at novel words), the authors interpreted these findings as pointing to a unity of the processes active in language processing and learning, as hypothesized in the predictive processing framework. That is, children use their best model of linguistic input to predict what speakers would say next, they rapidly update this model to fit any new evidence they encounter, and they deploy their updated model to process familiar words and learn new words, with no clear distinction between learning and processing (see Havron et al., 2021). To summarize, these studies show how syntactic contexts allow children to predict certain things about their

language - relating to the idea of syntactic bootstrapping, but adding to it the dimension of children's ability to adaptively change their predictive models of syntax.

Looking at syntactic bootstrapping through the lens of the predictive processing framework can also help gain insight into the question of how children choose which linguistic levels to rely on for making predictions (phonology, lexicon, syntax) despite the fact that their knowledge of each level is incomplete. How does the predictive processing framework shed light on this problem?

In a recent study (Beretti et al., 2020), researchers explored children's ability to adapt their predictions at the lexical and the syntactic levels. Two groups of 4- to-5-year old French children were presented with sentences that are ungrammatical without context such as *elle balle*–"she ball(s)". The critical manipulation in this experiment was the contextual support provided to the seemingly ungrammatical sentences. In one group of children (syntactically-reliable-speaker group), the ungrammatical sentences were presented with videos supporting novel meanings for the familiar word forms (e.g., as if the phrase "She balls" referred to a novel action being performed by a girl). In the other group of children (semantically-reliable-speaker group), the ungrammatical sentences were presented with videos depicting the familiar meanings of the words (e.g. a girl holding a ball while hearing *elle balle* "she balls", which would be interpreted as an error in the use of a function word). This manipulation prompted children to update their syntactic predictions. At test, children from both groups were presented with two videos, and asked to choose the one that best illustrated the sentence they heard. For example, w hile listening to "ungrammatical sentences" (e.g., *une dort*, "a sleeps"), participants saw two videos side-by-side: one presenting the reasonable choice if the sentence was believed to be syntactically correct (e.g., *une dort* –"a sleeps" describes a novel object) and another video presenting the reasonable choice if the sentence was believed to be semantically correct despite the ungrammatical use of function words (e.g., "a sleeps" describes a girl sleeping). Children who had been led to expect function words to be unreliable cues (i.e., those in the semantically-reliable-speaker group exposed to semantically correct sentences involving erroneous function words) pointed more to the event of the girl sleeping than children who had been led to expect the meanings of familiar content words to be unreliable

cues (i.e., those in the syntactically-reliable-speaker group exposed to syntactically correct sentences involving new homophones (see Figure 4).

Figure 4. Example of a test trial from Beretti et al., (2020). On the right is the video containing a novel object, a plausible interpretation of the sentence *une dort*, "a sleeps", if children interpret the test sentences as being syntactically correct and thus referring to a novel meaning for a word they already know, a homophone of the verb "sleep"). On the left is the video containing a familiar action (e.g., a woman sleeping), a plausible interpretation of the sentence if children interpret the test sentences as being semantically correct (i.e., "sleep" in the sentence "a sleep" means to sleep) despite the unreliability of function words in the context of the experiment. Note that in the induction phase, there was only one viable interpretation open to the participants: a novel word for the syntactically-reliable-speaker group (i.e., homophones group), and a familiar object or action for the semantically-reliable-speaker group - it is only at test that these two options were both presented to the participants, with words that were not presented during the induction phase.

These results showed that children can not only adjust their syntactic predictions as in Havron and colleagues (2019, 2021), but also meta-adjust whether they should be learning about new lexical items (i.e., novel homophones) based on their knowledge of syntax, or learning about syntactic contexts based on their knowledge of the lexicon (here, either stopping to rely on syntactic contexts or reversing their expectations, starting to expect a verb to follow a determiner). This is consistent with Bayesian optimization predictive inferences. According to this idea, a sentence is interpreted by taking into account a listener's prior expectations and combining them with new evidence to generate adjusted predictions (in Bayesian terms the *posterior*). To give another example, if an English-speaking child, for example, expects the words "the baby" to be followed by a noun 63% of the time (e.g., "the baby elephant"), and by a verb 37% of the time (e.g., "the baby jumps"), but they have recently heard four examples of "the baby VERBs", they

would adapt their model to expect “the baby” to be followed by a noun only 60% or 50% or 30% of the time (the exact number will depend not only on the predicted probability, but also on their confidence in this prediction). Four examples of “the baby NOUNs”, the more frequent option, would be expected to cause smaller changes to the model as was indeed found in Yu et al., (*in preparation*).

As mentioned above, the decision of whether and which level to adapt (here, syntactic cues or lexical information) is a meta-Bayesian decision, that is, choosing between two models which are themselves Bayesian. This idea was suggested by the noisy channel models of language processing (Clayard et al., 2008; Gibson, et al., 2013; Levy, 2008). Following noisy-channel models, the listener estimates which information, i.e., syntactic vs. lexical, is more likely to be unreliable, and adjusts accordingly to their (lexical or syntactic) expectations in the face of a prediction error. This unreliability, or noise, can either come from the input they perceive now (Is it reliable?) or their own representations (Are they robust? Are they certain?). In Beretti et al., (2020), children were led to trust their syntactic knowledge more in one condition than in the other, showing that children do not have one default route (always use syntax to learn words, or always use familiar words to learn about syntax), but can choose between these options by relying on contextual cues to judge where the error signal is most likely to result from. However promising, these findings offer limited support to the idea that prediction is directly related to language learning, as this experiment was done with 4- to-5-year-olds, and the one from Havron et al., (2019, 2021) with 3- to-4-year-olds. To support the role of predictive processing in the first phases of language learning, and reliably connect this with the semantic seed hypothesis, we would need to extend these findings to infants, first conceptually and then practically.

Conceptually, we would expect young infants to deploy the same Bayesian adaptation processes by determining the level of noise at different linguistic levels depending on the context they are in. That is, the more noisy or uncertain a predictive model is, the less likely infants are to rely on it. The syntactic bootstrapping and the semantic seed literature demonstrate the presence of such an ability to use two different levels of abstraction for prediction. For example, infants who have managed to learn that

certain syntactic contexts are likely to predict nouns or object labels (e.g., “this is a” + OBJECT) would be able to learn other noun contexts when familiar words repeatedly appear in them (e.g., “here is the”+ OBJECT). They would use these familiar contexts to infer that novel words occurring in these contexts are also labels for objects (e.g., “this is a dase” -> *dase* is an object too). Thus, they would be able to use their rudimentary semantic categories to build syntactic predictions, and their rudimentary noun category to add lexical items to their lexicon more efficiently. This does not yet show, however, that their choice of which category to adapt depends on the level of noise in their representation or input.

The idea that the level of noise perceived by infants determines which sources of information they would rely on is supported by findings showing that some syntactic contexts trigger stronger predictions than others. For example, in some languages, determiners may be more reliable cues to the syntactic category of an adjacent novel word than are subject pronouns: In German, the determiner *ein* is followed by a noun in 71% of cases, while the personal pronoun *sie* is followed by a verb in 31% of cases (based on a child-directed corpus reported in Höhle et al., 2004). Similar distributional patterns were found in French (Cecyre & Shi, 2005). Accordingly, two previous studies have found that while 14-month-olds can build expectations about the type of function words which can co-occur with novel nouns (based on their past occurrence after determiners), they cannot build such expectations for novel verbs (based on their occurrence after pronouns; Höhle et al., 2004; Shi & Melançon, 2010). When additional cues are available, infants can successfully predict the syntactic category of novel words, whether nouns or verbs. A recent study found that French-learning 14-month-old infants can build expectations about novel verbs co-occurring with pronouns when familiar words are present in the familiarization phase (Babineau et al., 2020). In this case, the experimental manipulation strengthened the reliability of pronouns as cues to verbs for the task at hand. Thus, infants might avoid relying on noisy or unreliable sources of information, but strengthening the reliability of the available cues (e.g. distributional cues) does trigger syntactic prediction. Another piece of evidence comes from semantic predictions. Borovsky et al (2016) found that 24-month-olds performed better on items with higher semantic density (items that have a more robust lexical model) in two real-time comprehension tasks (lexical recognition and

sentence processing), while lower density items experienced more interference. That is, more weight was given to new information when the predictive model was less robust.

One aspect of predictive processing which was overlooked in these studies (Berreti et al., 2020; Havron et al., 2019, 2020a, 2020b, 2021) is the anticipatory aspect of prediction - using prediction to speed along processing, and expecting, for example, the end of a sentence before it is completed by the speaker. The same is true for other classical work dealing with error-based learning in language acquisition (Branigan & Messenger, 2016; Peter et al., 2015). The theory, as discussed here, does not require that representations be pre-activated and early anticipation be formed, only that representations will be updated after the novel data has been comprehended by the child and incorporated by them in their syntactic predictive model. However, the classical predictive processing studies in language acquisition (Mani & Huettig, 2012; Borovsky et al., 2014), as well as most recent studies (e.g., Gambi et al., 2021), do focus on anticipation. What, if anything, does real-time prediction add to this theory? Language comprehension and learning happens in real time (see e.g., Christiansen & Chater, 2016; de Carvalho, et al., 2019; Nappa et al., 2009; Trueswell, 2008; Trueswell et al., 1999), and while infants might be slow to respond, or even slow to *look* at a correct image, one of the goals of predictive processing is to reduce this gap. Seamless communication requires fast action by both the recipient and the sender of the message. The role of prediction is not only to make the recipient's inference about intended meaning more likely to be correct, but also to make it more efficient, in terms of resources and time. That would mean that both types of studies, those looking at how predictive contexts make comprehension better and those looking at how children adapt their predictions, contribute to understanding how predictive processing benefits language comprehension and acquisition. A welcome addition will be to conduct studies which incorporate both.

To summarize, children have been shown to use predictive processing to adapt their syntactic expectations in both language processing and learning. They were also shown to be flexibly adapting to the learning context by giving more weight to syntax or their lexicon based on the source of noise in their input. This aligns nicely with studies showing that the

robustness of children's predictive model affects their predictions: more weight seems to be given to new information when the predictive model is less robust.

5. Challenges to the predictive processing approach in language acquisition

One challenge to the predictive processing approach in language acquisition would seem to come from a literature showing that young children sometimes fail at revising their prediction: *the kindergarten-path effect* (e.g., Huang & Arnold, 2019; Trueswell et al., 1999). If children learn from prediction errors based on local context, how could learning occur if they fail to perceive their errors?

A predictive processing account of cognition does allow for our predictions to be so strong that we do not record an error when one occurs. To give an example, one might create such a strong prediction that a pacifier on the floor is a spider, based on partial evidence, that they would leave the room without noticing their error - not examining the truth of their prediction. Thus, one way to answer the challenge posed by the kindergarten path effect would be to look into why, under the same circumstances, some children are able to revise their expectations, while others do not seem to do so. Recent work might shed light on this issue. Reuter et al. (2019) used predictive contexts with a visual world paradigm. In the more constraining contexts, 3- to 5-year-old children saw, for example, a truck and a novel object which does not look like something someone could drive, and heard: "Vroom! You can drive the feb!". In less constraining contexts they heard, for example: "Where is the feb?" while viewing the same objects (i.e., the truck vs the novel object). The results showed that children who initially predicted a familiar referent and then redirected their attention to a novel referent showed greater accuracy at learning the label for the new object. Gambi et al (2021), used a similar design, and tested 2- to-4-year-olds and adults. Interestingly, like in the kindergarten-path effect, both age groups sometimes still chose the familiar object (e.g., the truck) despite the novel label (e.g., feb), and did this more often when the context was more constraining (e.g., "You can drive the feb"). Gambi et al (2021) found that when children did select the novel referent during the learning phase, children demonstrated above-chance retention of the association between the novel word and this novel referent. These children were capable of revising and updating their expectations based on the failed predictions on most trials. Taken together,

one possible explanation for these results is that children with stronger priors failed to update their expectations, but those who did not have such strong priors not only updated their priors, but also showed greater accuracy when they redirected their attention after a wrong prediction rather than directly look at the novel object.

A second challenge to the predictive processing account of language learning is that all studies described in this article were either conducted in a laboratory setting or were computational models relying on clear and clean input (transcribed text). The real world is noisy in many different ways: parents differ in how they use language, the same parents use language differently in different situations, events that occur around the child are distracting, and language development does not happen in an isolated way, but rather relies on many extralinguistic factors such as reference, sight and touch. If researchers cannot find traces of predictive processing (or, for that matter, syntactic bootstrapping) in the *real world*, these mechanisms might be very limited in their ability to explain how language is actually learned. For now, it is not that such studies in the real world have failed, but that - to our knowledge - they were never conducted.

A third challenge is the languages studied. While syntactic bootstrapping has been studied in a few languages other than English and French (e.g., in Mandarin Chinese: Lee & Naigles, 2005; in Turkish: Göksun, et al., 2008; and in Icelandic: Nowenstein, et al., 2020), and prediction of sentence ending has been studied in German (Mani & Heuttig, 2012), we know of no prediction adaptation studies in children in languages other than English and French. We plan to test prediction adaptation in Hebrew and Arabic ourselves. We believe it is also extremely important to test syntactic (or more generally, structural) bootstrapping, prediction of utterance ending and adaptation in Signed languages (see e.g., Gleitman et al., 2019). Without taking a broader look at whether and how these mechanisms play a role in diverse languages and situations, we cannot know to what extent current findings are generalizable, and whether they are restricted to some very specific European languages.

A last challenge to the predictive processing account of language learning is participants' age. If language learning relies on prediction mechanisms, these mechanisms should be present, and strong enough to compute syntactic regularities from very early

on. The semantic seed account hypothesizes that sometime during the second half of their first year of life, infants would have already learned a few seed words, and used those to begin to learn the most frequent and reliable contexts that predict them. Even though recent work has demonstrated infants' sensitivity to the context of these early known seed words (Babineau & Christophe, 2022) and their ability to build syntactic expectations based on these frequent and reliable co-occurrence patterns (e.g., Babineau, Shi, & Christophe, 2020), some recent studies have failed to find evidence for prediction *adaptation* in children younger than three years of age (e.g., Havron et al., 2021a,b; Gambi, 2018).

A few studies have found that young children can use function words predictively, by anticipating the likely features of the target word before its onset. In Lukyanenko and Fisher (2016), 2.5-year-olds used agreeing verbs to predict number features of an upcoming noun (e.g., children listening to the sentence "Where are the cookies?" started to look toward a picture with two cookies before they heard the noun onset). In Melançon and Shi (2015), French-learning 2.5-year-olds used French gender-marked determiners predictively; children started to look toward a familiar target picture before they heard the target-noun (e.g., le joli ballon 'the pretty ball'). In line with these results, Gambi and colleagues (2018) found that 3- to 5-year-olds start to reliably use the determiner *an* to predict whether the following word will begin with a vowel. However, they found that two-year-olds were unable to do so. This latter finding mirrors what Havron et al. (2021) reported, i.e. that 3-4-year-old children were able to adjust their syntactic predictions and use those to disambiguate a homophone, but two-year-olds showed no evidence of this. Similarly, a recent study examined French 18-month-old infants' syntactic adaptation using the syntactic frame *la petite* which can precede either nouns or verbs, the same one that was used to demonstrate syntactic adaptation in 3- to-4-year-olds in two studies (Havron, et al., 2019; 2021). Unlike the other studies mentioned above (e.g., Havron et al., 2019, 2021; Gambi et al., 2018) which used a looking-while-listening paradigm, this study used a habituation paradigm. Habituation not only places less processing burden on infants (who do not need to process and compare two videos or images at the same time), but also allows much more exposure. In Havron et al. (2021), infants were exposed to *la petite* sentences followed by either familiar nouns or familiar verbs mixed in with a novel word appearing after *la petite* (*la petite crale*). At test, infants heard the same novel word in noun

contexts in one block, and in verb contexts in another. If infants indeed rapidly created and deployed expectations during the habituation phase, they should be surprised and look longer when the novel word appears in contexts that are incongruent with the syntactic category of the familiar words appearing after *la petite in the habituation stage*. For example, if *la petite* was predictive of nouns during habituation, and given that the novel word also appeared after *la petite* during the habituation phase, infants should have categorized it as a noun and should have been surprised to hear it in a verb context at test. However, 18-month-olds failed to show signs of syntactic adaptation in this task.

Taken together, these findings do point to a developmental time-course whereby linguistic prediction (anticipation *and* adaptation) becomes more efficient as children mature and acquire more experience with language. However, Zhang and colleagues (2019), using non-linguistic stimuli (sounds such as a xylophone or trill predicting an image), showed identical prediction effects in adults and six-month-old infants. Reuter and colleagues (2018) also found adaptation of non-linguistic predictions in 12- to-24-month-olds. In their experiment, a central cue reliably predicted the side of the screen where the target would appear. In the middle of the experiment the target side changed. They found that infants with larger vocabularies were more efficient in updating their predictions to the new target side. This leads us to present three (non-mutually exclusive) possibilities for the null findings obtained with young infants in syntactic adaptation and linguistic prediction experiments.

First, task demands may interfere with infants' ability to display their underlying capabilities. They are slower, have smaller short-term memory capacities, and do not concentrate on the task as much as older children (and older children do not do as well as adults, e.g. in Beretti et al., 2020; see Havron, et al., 2020 for a discussion). Task demands are not only a technical, in-lab, confound. The daily demands of perceiving live input accurately, processing it, and generalizing from it would also weigh more heavily on younger infants. Indeed, Huettig and Janse (2016) tested adults' linguistic predictions (the rapidity to which they looked at an image of a noun that agrees with the determiner in gender) and found that higher working-memory abilities and faster processing speed predicted adults' anticipatory eye movements. Similarly, faster looks to target pictures

were found in children, as their processing speed abilities increase (e.g. Fernald, Swingley, & Pinto, 2001; see also Bergelson, 2020). Other cognitive capacities which are still underdeveloped in young children, including cognitive control and executive functions (see e.g., Woodard, Pozzan & Trueswell, 2016), might also account for this difficulty dealing with the experimental tasks that were used in the past and maybe with real-life predictions as well.

Second, adaptation happens when the current predictive model of language is not sufficient (i.e., results in errors) for the context at hand. Thus, another possibility is that infants, but not children, fail in syntactic adaptation experiments because they fail to evaluate the reliability of their knowledge, a necessary component for adaptation to occur. While the ability to verbally report on the reliability of one's own linguistic knowledge is a late achievement that emerges at around 4 to 5 years of age (e.g., Gleitman, Gleitman, 1979; Marazita & Merriman, 2004), an expanding literature suggests that the ability to evaluate one's own representations and processes is separate from the ability to consciously access and report on such evaluations ("core metacognition"; see e.g., Goupil & Kouider, 2019). For example, there is evidence that infants are able to evaluate the reliability of their knowledge or decisions years before they can provide verbal reports of their knowledge states (e.g., Balcomb & Gerken, 2008; Goupil & Kouider, 2016; Goupil, et al., 2016). More recently, Dautriche et al., (2021) also found evidence that even 2-year-olds can evaluate the reliability of their lexical knowledge. In a modified version of the looking-while-listening task (Golinkoff et al., 1987), infants looked longer towards the image they chose as the target when they had reasons to be sure about that word's meaning. While this suggests that the rudimentary self-assessment of linguistic knowledge may be available early on in development, it remains an open question whether such an ability is fully-developed in early ages, extends to other levels of linguistic knowledge (i.e., syntax), and can be *used* by infants as a cue to update their knowledge.

Third, it might be that linguistic prediction occurs only when linguistic knowledge is robust and stable enough to allow predictions, meaning that in all the cases of null results discussed above, the structures tested might not have been known well enough by the infants tested. This possibility is in line with our suggestion above: infants rely on very

frequent and highly-predictive syntactic contexts to predict a syntactic category, but do not do so when contexts are more variable (for example, young children are able to predict that nouns will follow determiners before they can predict that verbs are likely to follow pronouns) - and would thus confirm predictions derived from the predictive processing account of language learning.

Previous research found a positive correlation between children's vocabulary and their prediction skills (e.g., Borovsky & Creel, 2014; Borovsky et al., 2012; Lew-Williams & Fernald, 2007; Mani & Huettig, 2012; Ylinen et al., 2017), but to our knowledge the correlation between syntactic knowledge and prediction skills has not yet been examined. Nevertheless, since vocabulary and syntactic knowledge are highly correlated (see e.g., Levine et al., 2020), and, as described above, probably also synergistically depend on each other, it is very likely that children with more robust syntactic knowledge show better prediction skills. Additionally, language production seems to support linguistic prediction in adults (here, in the sense of anticipation of a word, not in the sense of adaptation to changes, Lelonkiewicz et al., 2021). It is therefore possible that adaptation of predictions is more difficult for infants who cannot yet produce such sentences themselves.

Therefore, even if prediction is a core characteristic of the way our brain works, is present from birth, and does not improve with age in the same way that abilities like short-term memory or processing speed do - linguistic predictions could improve, become stronger and faster, and affect more complex levels of linguistic processing as children acquire more linguistic experience. The developmental route of the strength of adaptation would not be linear: when infants have no active predictive model of a certain category or area of language (e.g., 6-month-olds might be unable to predict whether a verb is transitive or intransitive) they attempt no prediction. When they have acquired enough experience to attempt a prediction, but this prediction is based on just enough but little experience and rudimentary priors, they will show strong changes in their predictions in the face of new evidence. When their model is robust, they will show weaker adaptation (as any new piece of evidence carries less weight). This theoretical proposal has not (yet) been supported by existing evidence, however, it makes several concrete predictions that should be tested before we might consider discarding it. First, we should be able to find prediction

adaptation in infants younger than 2 years of age if we test them on highly frequent and reliable syntactic contexts. Second, infants' ability to predict should not be independent from the type and level of predictions the task requires them to make. For example, very young infants already distinguish between maternal speech that has been interrupted at clausal boundaries, and matched speech that has been interrupted at within-clause locations (see e.g., Kemler Nelson, et al., 1989), implying that infants could be predicting clause boundaries based on prosodic cues. Phrasal-boundary prosody would therefore be hypothesized to show prediction earlier than, for example, syntactic contexts. Third, the more language experience an infant has, the earlier they would attempt a prediction, and the more robust their models will be. To see this in children, though, we would have to control for children's age and cognitive abilities (e.g., a child with stronger memory skills is likely to succeed better in our tasks in the lab than a child with lower memory skills, we would thus need to test different groups of children with the same memory skills, but differing in their linguistic experience).

Rabagliati et al. (2016) maintained that it was not yet possible to disentangle two possibilities. One, that children, as we have claimed in this paper, are predicting to learn (learning language through linguistic predictions), and the second, that children are learning to predict - that is, prediction is an ability that emerges in expert language users, but is not yet developed enough in children (and especially not infants). The recent studies we reviewed in this paper represent a major advancement in understanding whether and how children use prediction to learn, and how they learn through predictions. These studies led us to formulate a theory of how infants synergistically learn to predict and predict to learn, relying on, and adapting, their current representations of their native language(s) at multiple levels in a cascading bottom-up and top-down manner.

6. Conclusions and future directions

The findings presented in this paper underline the tremendous sophistication of early lexical and syntactic acquisition. As proposed by the *semantic seed* hypothesis, these studies show that by the end of their first year of life, infants can use their blossoming lexicon to learn more about the syntax of their native language (e.g., Babineau &

Christophe, 2022). This enables them to use newly-learned syntactic contexts to guide their inferences about the meaning of novel words (syntactic bootstrapping; Babineau et al., 2021; Barbir et al., 2019). These synergies between lexical and syntactic acquisition are facilitated by children's distributional/statistical learning skills, notably their ability to keep track of the regularities in their input and their ability to generalize from a small set of exemplars.

Children's linguistic representations are shown to be flexible, such that with increasing experiences with the world, children can not only update the phonological representations of familiar words (e.g., Babineau, Legrand, & Shi, 2021) or word-to-referent mappings (e.g., Berreti et al., 2020; Dautriche et al., 2018; de Carvalho et al., 2021), but also their syntactic expectations (e.g., Havron et al., 2019; 2021). As demonstrated by Havron et al., (2019; 2021), the predictiveness of a syntactic context can change when the child is presented with convincing evidence from recent input, such as that *la petite* is more frequently followed by a verb than by a noun (whereas it is more often followed by a noun in French child-directed input). However, when the child is operating with limited experience (with rudimentary lexical and syntactic knowledge), adapting to conflicting information might not yet be possible. This might partly explain why children below the age of 3 have failed to adapt the predictiveness of familiar syntactic contexts in laboratory settings (e.g., Havron et al., 2020).

Taken together, we believe that the route we outline is a plausible theory for early language acquisition. From an initial group of semantic seeds, infants could use their predictive processing skills to learn a handful of syntactic contexts. These syntactic contexts would then help them infer the meaning of novel words, creating a larger and more complex lexicon. As the lexicon grows, so does the reservoir of syntactic contexts infants know and use to predict. At first, infants only rely on very frequent syntactic contexts that make very reliable predictions (such as determiners in French), but as their language develops, they learn to also utilize less reliable syntactic contexts (such as personal pronouns in French, and contexts like *la petite*). Throughout their development, children will also be faced with learning contexts in which they will either need to adapt their syntactic representations and learn something new about syntax from familiar words

(e.g. "ko *rabbit*"), or add new words to their lexicon by relying on the syntactic contexts they have mastered (e.g., "*she* blicks"). What factors affect this choice is still an open question. One factor we have discussed at length is the reliability of the source of information (i.e., whether the speaker uses syntax vs. semantic reliably), but it is likely that other factors also affect this decision. For example, we would hypothesize that structures which have been more reliable predictors in the past (e.g., determiners as predictors of nouns would be more reliable than pronouns as predictors of verbs) would be less likely to be updated in the face of new evidence. Other factors may be: (1) the reliability of the speaker (e.g., a native speaker might be a more reliable source of information than a non-native speaker, and an adult might be a more reliable source of information than a child; see e.g., Dautriche et al., 2020, Koenig & Woodward, 2010 for evidence that children factor in a speaker's reliability when updating their lexicon), (2) the consistency of the new information (whether it was presented only in the course of the experiment or also at home), or (3) the number of speakers who use it (the more speakers, the more likely the information is correct and generalizable). Future studies should examine these and other factors and their relative weight.

Most of the studies we reviewed in this paper were performed on infants and children learning Indo-European languages, especially English and French. Consequently, this paper does not differ from the field of language acquisition at large, which is lacking in studies on diverse languages and populations (Kidd & Garcia, 2021). We believe one way to move forward is to study other languages, which differ from Indo-European languages in their syntactic and morphological properties (see Havron et al, in press for a recommendation of how our field could do so efficiently).

Perhaps the most important question we leave open is whether young infants can be shown to adapt their predictions, and how and when they develop the ability to anticipate what others will say next. As we expanded on above, it is likely that infants would be slower to adapt than children and adults, given their more limited cognitive resources. Other, non-mutually exclusive options are that infants cannot yet gauge the reliability of their own representations or the level of noise in the environment, or that they need to develop linguistic experience in order to perform adaptations to their predictions. We

would hypothesize that this would be especially true for higher-level predictions such as deciding when to adapt syntactic representations, and when to decide to use syntactic representations to learn new words - based on what might be called extra-linguistic factors such as speaker reliability. However, more research is needed before we can conclude that infants can or cannot adapt, or whether their prediction adaptation relies on factors such as the amount and nature of their linguistic experience. The current review paper thus highlights new avenues to investigate the role of prediction in language learning, before and beyond the syntactic bootstrapper.

References

Adams, L., & Conklin, N. (1973). Toward a theory of natural classification. *Papers from the Annual Regional Meeting of the Chicago Linguistics Society*, 9, 1–10.

Aikhenvald, A. Y. (2000). Classifiers: A typology of noun categorization devices. In *A typology of noun categorization devices*. Oxford, OUP.

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. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. In *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 305–315). Dordrecht: Springer.

Babineau, M., & Christophe, A. (2019). Preverbal infants' sensitivity to grammatical dependencies. In *Talk presented at the 44th Boston University Conference on Language Development*. Boston - MA.

Babineau, M., de Carvalho, A., Trueswell, J., & Christophe, A. (2021). Familiar words can serve as a semantic seed for syntactic bootstrapping. *Developmental Science*, *24*(1), 1–12. https://doi.org/10.1111/desc.13010

Babineau, M., Legrand, C., & Shi, R. (2021). Variable forms in French-learning toddlers' lexical representations. *Developmental Psychology*. https://doi.org/10.1037/dev0001157

Babineau, M., Shi, R., & Christophe, A. (2020). 14-month-olds exploit verbs' syntactic contexts to build expectations about novel words. *Infancy*, infa.12354. https://doi.org/10.1111/infa.12354

Balcomb, F. K., & Gerken, L. A. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, *11*(5), 750–760. https://doi.org/10.1111/j.1467-7687.2008.00725.x

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. https://doi.org/10.1016/j.tics.2007.05.005

Barbir, M., Babineau, M., Fievet, A.-C., & Christophe, A. (2019). Infants quickly use newly-learned grammar to guide acquisition of novel words. In *The Way We Learn (Doctoral Dissertation)* (pp. 14–39). Paris, France.

Beretti, M., Havron, N., & Christophe, A. (2020). Four- and 5-year-old children adapt to the reliability of conflicting sources of information to learn novel words. *Journal of Experimental Child Psychology*, *200*, 104927. https://doi.org/10.1016/j.jecp.2020.104927

Bergelson, E. (2020). The Comprehension Boost in Early Word Learning: Older Infants Are Better Learners. *Child Development Perspectives*, *14*(3), 142–149. https://doi.org/10.1111/cdep.12373

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 of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391–397. https://doi.org/10.1016/j.cognition.2013.02.011

Bergelson, E., & Swingley, D. (2015). Early Word Comprehension in Infants: Replication and Extension. *Language Learning and Development*, *11*(4), 369–380. https://doi.org/10.1080/15475441.2014.979387

Bernal, S., Lidz, J., Millotte, S., & Christophe, A. (2007). Syntax Constrains the Acquisition of Verb Meaning. *Language Learning and Development*, *3*(4), 325–341. https://doi.org/10.1080/15475440701542609

Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, *50*(5), 1600–1613. https://doi.org/10.1037/a0035591

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child Development, 87*(6), 1893-1908. DOI: 10.1111/cdev.12554

Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436. https://doi.org/10.1016/j.jecp.2012.01.005

Borovsky, A., Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language, 73*, 1-14. https://doi.org/10.1016/j.jml.2014.02.001

Bovolenta, G., & Marsden, E. (2020). Expectation Violation Enhances the Development of New Abstract Syntactic Representations: Evidence from an Artificial Language Learning Study. *OSF Preprints*, *April 20*. https://doi.org/10.31219/osf.io/zyegf

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), 33–44. https://doi.org/10.1016/S0010-0277(01)00122-6

Brusini*, P., Seminck*, O., Amsili, P., & Christophe, A. (*in revision*). The Acquisition of Noun and Verb Categories by Bootstrapping from a few Known Words: A Computational Model. *Frontiers in Psychology*.

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

Cécyre, M. J., Shi, R. (2005). Speech rate in maternal speech to French-learning infants. In *Paper presented at the 15th meeting of the Canadian Society for Brain, Behaviour and Cognitive Science.* Montreal, Quebec, Canada.

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 7th Workshop on Cognitive Aspects of Computational Language Learning* (pp. 38–43). Berlin, Germany.

Christophe, A., Dautriche, I., de Carvalho, A., & Brusini, P. (2016). Bootstrapping the syntactic bootstrapper . In J. Scott & D. Waughtal (Eds.), *Proceedings of the 40th*

Annual Boston University Conference on Language Development (pp. 75–88). Somerville, MA: Cascadilla Press.

Christophe, A., Nespor, M., Guasti, M. T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Developmental Science*, *6*(2), 211–220. https://doi.org/10.1111/1467-7687.00273

Chrupała, G., & Alishahi, A. (2010). Online entropy-based model of lexical category acquisition. In *CoNLL 2010 - Fourteenth Conference on Computational Natural Language Learning, Proceedings of the Conference* (pp. 182–191).

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809. https://doi.org/10.1016/j.cognition.2008.04.004

Dautriche, I., Goupil, L., Smith, K., & Rabagliati, H. (2021). Two-year-olds' eye movements reflect confidence in their understanding of words. *PsyArXiv*, *June 22*. https://doi.org/https://doi.org/10.31234/osf.io/pd6xh

Dautriche, Isabelle, Cristia, A., Brusini, P., Yuan, S., Fisher, C., & Christophe, A. (2014). Toddlers Default to Canonical Surface-to-Meaning Mapping When Learning Verbs. *Child Development*, *85*(3), 1168–1180. https://doi.org/10.1111/cdev.12164

Dautriche, Isabelle, Fibla, L., Fievet, A. C., & Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive Psychology*, *104*(April), 83–105. https://doi.org/10.1016/j.cogpsych.2018.04.001

Dautriche, Isabelle, Goupil, L., Smith, K., & Rabagliati, H. (2020). Knowing how you know: Toddlers reevaluate words learned from an unreliable speaker. *Open Mind*, *5*, 1–19. https://doi.org/10.1162/opmi_a_00038

de Carvalho, A., Babineau, M., Trueswell, J. C., Waxman, S. R., & Christophe, A. (2019). Studying the Real-Time Interpretation of Novel Noun and Verb Meanings in Young Children. *Frontiers in Psychology*, *10:274*(February), 1–16. https://doi.org/10.3389/fpsyg.2019.00274

de Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and Function Words Cue the Acquisition of Word Meanings in 18-Month-Old Infants. *Psychological Science*, *30*(3). https://doi.org/10.1177/0956797618814131

de Carvalho, A. (2017). *The role of phrasal prosody and function words in the acquisition of word meanings*. Ecole Normale Supérieure, PSL Research University, Paris - France. Retrieved from https://tel.archives-ouvertes.fr/tel-01943786

de Carvalho, Alex, Dautriche, I., Fiévet, A., & Christophe, A. (2021). Toddlers exploit referential and syntactic cues to flexibly adapt their interpretation of novel verb meanings. *Journal of Experimental Child Psychology*, *203*, 105017. https://doi.org/10.1016/j.jecp.2020.105017

Dell, G. S., & Chang, F. (2014). The p-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634). https://doi.org/10.1098/rstb.2012.0394

Den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, *30*(9), 3210–3219. https://doi.org/10.1523/JNEUROSCI.4458-09.2010

Denny, J. P. (1976). What are noun classifiers good for? In *Papers from the 12th Regional Meeting, Chicago Linguistic Society* (pp. 122–132). Chicago, IL.

Dixon, R. M. (1986). Noun classes and noun classification in typological perspective. In *Noun classes and categorization* (pp. 105–112).

Dryer, M. (1992). The greenbergian word order correlations. *Language*, *68*(1), 81–138.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Feldman, N. H., Myers, E. B., White, K. S., Griffiths, T. L., & Morgan, J. L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition*, *127*(3), 427–438. https://doi.org/10.1016/j.cognition.2013.02.007

Ferguson, B., Graf, E., & Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: evidence from 15- and 19-month-olds. *Cognition*, *131*(1), 139–146. https://doi.org/10.1016/j.cognition.2013.12.014

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When Half a Word Is Enough: Infants Can Recognize Spoken Words Using Partial Phonetic Information. *Child Development*, *72*(4), 1003–1015. https://doi.org/10.1111/1467-8624.00331

Ferry, A. L., Fló, A., Brusini, P., Cattarossi, L., Macagno, F., Nespor, M., & Mehler, J. (2016). On the edge of language acquisition: Inherent constraints on encoding multisyllabic sequences in the neonate brain. *Developmental Science*, *19*(3), 488–503. https://doi.org/10.1111/desc.12323

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 143-149. https://doi.org/10.1002/wcs.17

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, 92, 333-375. https://doi.org/10.1016/0024-3841(94)90346-8

Fisher, C., Jin, K., & Scott, R. M. (2020). The Developmental Origins of Syntactic Bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77. https://doi.org/10.1111/tops.12447

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., & Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental Science*, *22*(4), 1–16. https://doi.org/10.1111/desc.12802

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

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

Gambi, C., Jindal, P., Sharpe, S., Pickering, M. J., & Rabagliati, H. (2021). The relation between preschoolers' vocabulary development and their ability to predict and recognize words. *Child Development, 92*(3), 1048-1066. https://doi.org/10.1111/cdev.13465

Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not in children. *Cognition, 211*, 104650.https://doi.org/10.1016/j.cognition.2021.104650

Gerken, L., & Bollt, A. (2008). Three exemplars allow at least some linguistic generalizations: Implications for generalization mechanisms and constraints. Language Learning and Development, 4(3), 228-248.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the*

National Academy of Sciences, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Gleitman, H., & Gleitman, L. R. (2020). Language Use and Language Judgment. In *Sentence First, Arguments Afterward* (pp. 114–138). Oxford University Press. https://doi.org/10.1093/oso/9780199828098.003.0005

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*, 3–55.

Gleitman, L., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1038/scientificamerican0491-138

Gleitman, L., Senghas, A., Flaherty, M., Coppola, M., & Goldin-Meadow, S. (2019). The emergence of the formal category "symmetry" in a new sign language. *Proceedings of the National Academy of Sciences, 116*(24), 201819872. https://doi.org/10.1073/pnas.1819872116

Göksun, T., Küntay, A. C., & Naigles, L. R. (2008). Turkish children use morphosyntactic bootstrapping in interpreting verb meaning. *Journal of Child Language, 35*(2), 291-323. https://doi.org/10.1017/S0305000907008471

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45. https://doi.org/10.1017/S030500090001271X

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135. https://doi.org/10.1016/S0010-0277(99)00003-7

Goupil, L., & Kouider, S. (2016). Behavioral and Neural Indices of Metacognitive Sensitivity in Preverbal Infants. *Current Biology*, *26*(22), 3038–3045. https://doi.org/10.1016/j.cub.2016.09.004

Goupil, L. & Kouider, S. (2019). Developing a reflective mind: From core metacognition to explicit self-reflection. *Current Directions in Psychological Science*, 28(4), 403-408. https://doi.org/10.1177/0963721419848672

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(13), 3492–3496. https://doi.org/10.1073/pnas.1515129113

Gutman, A., Dautriche, I., Crabbé, B., & Christophe, A. (2015). Bootstrapping the Syntactic Bootstrapper: Probabilistic Labeling of Prosodic Phrases. *Language Acquisition*, *22*(3), 285–309. https://doi.org/10.1080/10489223.2014.971956

Hallé, P. A., Durand, C., & de Boysson-Bardies, B. (2008). Do 11-month-old French infants process articles? *Language and Speech*, *51*(Pt 1-2), 23–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18561542

Havron, N., Babineau, M., Fievet, A.-C., de Carvalho, A., & Christophe, A. (2021). Syntactic prediction adaptation accounts for language processing and language learning Naomi. *Language Learning,* 71(4), 1194-1221.

Havron, N., Babineau, M., & Christophe, A. (2021). 18-Month-Olds Fail To Use Recent Experience To Infer the Syntactic Category of Novel Words. *Developmental Science*, *24*(2), 1–12. https://doi.org/10.1111/desc.13030

Havron, N., de Carvalho, A., Fiévet, A.-C., & Christophe, A. (2019). Three- to Four-Year-Old Children Rapidly Adapt Their Predictions and Use Them to Learn Novel Word Meanings. *Child Development*, *90*(1). https://doi.org/10.1111/cdev.13113

Havron, N., Scaff, C., Carbajal, M. J., Linzen, T., Barrault, A., & Christophe, A. (2020). Priming syntactic ambiguity resolution in children and adults. *Language, Cognition and Neuroscience, 3*5(10), 1445-1455. https://doi.org/10.1080/23273798.2020.1797130

Havron, N., Scaff, S., Hitczenko, K., & Cristia, A. (in press) Community-set goals are needed to increase diversity in language acquisition research: A commentary on Kidd and Garcia (2022). First Language. https://osf.io/u7w6t

He, A. X., & Lidz, J. (2017). Verb Learning in 14- and 18-Month-Old English-Learning Infants. *Language Learning and Development*, *13*(3), 335–356. https://doi.org/10.1080/15475441.2017.1285238

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition, 115*(3), 444-457. https://doi.org/10.1016/j.cognition.2010.03.006

Hohle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The Recognition of Discontinuous Verbal Dependencies by German 19-Month-Olds: Evidence for Lexical and Structural Influences on Children's Early Processing Capacities. *Language Learning and Development*, *2*(4), 277–300. https://doi.org/10.1207/s15473341lld0204_3

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional Elements in Infants' Speech Processing: The Role of Determiners in the Syntactic

Categorization of Lexical Elements. *Infancy*, *5*(3), 341–353. https://doi.org/10.1207/s15327078in0503_5

Huang, Y. T., & Hollister, E. (2019). Developmental parsing and linguistic knowledge: Reexamining the role of cognitive control in the kindergarten path effect. *Journal of experimental child psychology, 184*, 210-219. https://doi.org/10.1016/j.jecp.2019.04.005

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Jin, K., & Fisher, C. (2014). Early evidence for syntactic bootstrapping : 15-month-olds use sentence structure in verb learning. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38h Annual Boston University Conference on Language Development*. Boston - MA: Cascadilla Press. Retrieved from http://www.bu.edu/bucld/supplementvol38/

Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word Form Extraction by 6-month-olds. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0083546

Kedar, Y., Casasola, M., Lust, B., & Parmet, 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

Kemler Nelson, D. G. , Hirsh-Pasek, K., Jusczyk, P. W., & Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*(1), 55-68. https://doi.org/10.1017/S030500090001343X

Kidd, E., & Garcia, R. (2021, September 6). How diverse is child language acquisition?. https://doi.org/10.31234/osf.io/jpeyq

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: Possible mechanisms. *Developmental Psychology*, *46*(4), 815–826. https://doi.org/10.1037/a0019664

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A Look around at What Lies Ahead: Prediction and Predictability in Language Processing. In *Predictions in the Brain*

(Vol. 15, pp. 190–207). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195395518.003.0065

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. Dissociating (2013) N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience, 25*, 484–50. 10.1162/jocn_a_00328

Lee, J. N., & Naigles, L. R. (2005). The input to verb learning in Mandarin Chinese: a role for syntactic bootstrapping. *Developmental Psychology, 41*(3), 529. https://doi.org/10.1037/0012-1649.41.3.529

Lelonkiewicz, J. R., Rabagliati, H., & Pickering, M. J. (2021). The role of language production in making predictions during comprehension. *Quarterly Journal of Experimental Psychology*, 174702182110284. https://doi.org/10.1177/17470218211028438

Levine, D., Pace, A., Luo, R., Hirsh-Pasek, K., Michnick Golinkoff, R., de Villiers, J., … Sweig Wilson, M. (2020). Evaluating socioeconomic gaps in preschoolers' vocabulary, syntax and language process skills with the Quick Interactive Language Screener (QUILS). *Early Childhood Research Quarterly*, *50*, 114–128. https://doi.org/10.1016/j.ecresq.2018.11.006

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lew-Williams, C., & Fernald, A. (2007). Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198. https://doi.org/10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake-But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843–847. https://doi.org/10.1037/a0029284

Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language*, *51*(3), 458–472. https://doi.org/10.1016/j.jml.2004.06.008

Marchetto, E., & Bonatti, L. L. (2013). Words and possible words in early language acquisition. *Cognitive Psychology*, *67*(3), 130–150. https://doi.org/10.1016/j.cogpsych.2013.08.001

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Martin, A., Peperkamp, S., & Dupoux, E. (2013). Learning Phonemes With a Proto-Lexicon. *Cognitive Science*, *37*(1), 103–124. https://doi.org/10.1111/j.1551-6709.2012.01267.x

Messenger, K., Yuan, S., & Fisher, C. (2015). Learning Verb Syntax via Listening: New Evidence From 22-Month-Olds. *Language Learning and Development*, *11*(4), 356–368. https://doi.org/10.1080/15475441.2014.978331

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

Molina, M., Van de Walle, G. A., Condry, K., & Spelke, E. S. (2004). The animate-inanimate distinction in infancy: Developing sensitivity to constraints on human actions. *Journal of Cognition and Development*, *5,* 399–426. https://doi.org/10.1207/s15327647jcd0504_1

Nappa, R., Wessel, A., McEldoon, K. L., Gleitman, L. R., & Trueswell, J. C. (2009). Use of speaker's gaze and syntax in verb learning. *Language Language and Development*, 5(4). https://doi.org/10.1080/15475440903167528

Newport, E. L. (2016). Statistical language learning: computational, maturational, and linguistic constraints. *Language and Cognition, 8*(3), 447–461. https://doi.org/10.1017/langcog.2016.20

Nowenstein, I., Sigurjónsdóttir, S., Yang, C., Ingason, A., & Wallenberg, J. (2020). The Meaning of Case: Morphosyntactic Bootstrapping and Icelandic Datives. In *Proceedings of the 44th annual Boston University conference on language development* (pp. 402-415). Boston, MA: Cascadilla Press.

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*(03), 455–484. https://doi.org/10.1017/S0305000910000127

Parise, E., & Csibra, G. (2012). Electrophysiological Evidence for the Understanding of Maternal Speech by 9-Month-Old Infants. *Psychological Science*, *23*(7), 728–733. https://doi.org/10.1177/0956797612438734

Parisien, C., Fazly, A., & Stevenson, S. (2008). An incremental Bayesian model for learning syntactic categories. In *CoNLL 2008 - Proceedings of the Twelfth Conference on Computational Natural Language Learning* (pp. 89–96). https://doi.org/10.3115/1596324.1596340

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Rabagliati, H., Gambi, C., & Pickering, M. J. (2016). Learning to predict or predicting to learn? *Language, Cognition and Neuroscience, 31*(1), 94–105. https://doi.org/10.1080/23273798.2015.1077979

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological Bulletin*, *127*, 209–228. https://doi.org/10.1037/0033-2909.127.2.209

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425–469. https://doi.org/10.1207/s15516709cog2204_2

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology*, *55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, *176*(March), 215–219. https://doi.org/10.1016/j.cognition.2018.03.006

Rostad, K., Yott, J., & Poulin-Dubois, D. (2012). Development of Categorization in Infancy: Advancing Forward to the Animate/Inanimate Level. *Infant Behavior and Development*, *35*, 584-595. doi: 10.1016/j.infbeh.2012.05.005

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical Learning by 8-Month-Old Infants. *Science, 274*(5294), 1926–1928. doi: 10.1126/science.274.5294.1926

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language, 35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479–500. https://doi.org/10.1016/j.cognition.2007.10.010

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134. https://doi.org/10.1016/S0010-0277(98)00060-2

Shafer, V., Shucard, D., Shucard, J., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, *41*, 874–886.

Shi, R, & Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants: Paper. *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

Shi, R, Morgan, J. L., & Allopenna, P. D. (1998). Phonological and acoustic bases for earliest grammatical category assignment: a cross-linguistic perspective. *Journal of Child Language*, *25*, 169–201. https://doi.org/10.1017/S0305000997003395

Shi, Rushen. (2014). Functional Morphemes and Early Language Acquisition. *Child Development Perspectives*, *8*(1), 6–11. https://doi.org/10.1111/cdep.12052

Shi, Rushen, Cutler, A., Werker, J. F., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *The Journal of the Acoustical Society of America*, *119*(6), EL61-L67. https://doi.org/10.1121/1.2198947

Shi, Rushen, & Werker, J. F. (2001). Six-Month-Old Infants' Preference for Lexical Words. *Psychological Science*, *12*(1), 70–75. https://doi.org/10.1111/1467-9280.00312

Shi, Rushen, & Werker, J. F. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science*, *6*(5), 484–488. https://doi.org/10.1111/1467-7687.00305

Shi, Rushen, Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198. DOI: 10.1207/s15327078in1002_5

Shi, Rushen, Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11–B21. https://doi.org/10.1016/S0010-0277(99)00047-5

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*(1), 1–32. https://doi.org/10.1016/j.cogpsych.2006.04.002

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Strickland, B. (2017). Language Reflects "Core" Cognition: A New Theory About the Origin of Cross-Linguistic Regularities. *Cognitive Science*, *41*(1), 70–101. https://doi.org/10.1111/cogs.12332

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Syrnyk, C., & Meints, K. (2017). Bye-bye mummy – Word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, *35*(2), 202–217. https://doi.org/10.1111/bjdp.12157

Swanson, E., Frank, M. C., & Degen, J. (2021). Syntactic adaptation and word learning in French and English. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175.

Tincoff, R., & Jusczyk, P. W. (2012). Six-Month-Olds Comprehend Words That Refer to Parts of the Body. *Infancy*, *17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*(2), 89-134. https://doi.org/10.1016/S0010-0277(99)00032-3

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. In I. Sekerina, E. Fernandez, & H. Clahsen (Eds.), Developmental Psycholinguistics: On-line methods in children's language processing (pp. 73–96). John Benjamins.

van de Weijer, J. (1998). *Language input for word discovery*. Max Plank Institute for Psycholinguistics.

van Heugten, M., & Johnson, E. K. (2010). Linking infants' distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, *63*(2), 197–209. https://doi.org/10.1016/j.jml.2010.04.001

van Heugten, M., & Johnson, E. K. (2011). Gender-marked determiners help Dutch learners ' word recognition when gender information itself does not *, *38*(August 2006), 87–100. https://doi.org/10.1017/S0305000909990146

Wang, H., Höhle, B., Ketrez, N., Küntay, A. C., & Mintz, T. H. (2011). Cross-linguistic distributional analyses with frequent frames: the cases of german and turkish. *Proceedings of 35th …*, (January), 628–640. Retrieved from http://www-bcf.usc.edu/~tmintz/pdf/BU35_final_HaoWang.pdf

Waxman, S. R. (1999). Specifying the scope of 13-month-olds' expectations for novel words. *Cognition*, *70*(3). https://doi.org/10.1016/S0010-0277(99)00017-7

Waxman, S. R., & Booth, A. E. (2001). Seeing Pink Elephants: Fourteen-Month-Olds' Interpretations of Novel Nouns and Adjectives. *Cognitive Psychology*, *43*(3), 217–242. https://doi.org/10.1006/cogp.2001.0764

Waxman, S. R., Lidz, J. L., Braun, I. E., & Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, *59*(1), 67–95. https://doi.org/10.1016/j.cogpsych.2009.02.001

Weisleder, A., & Waxman, S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language*, *37*(05), 1089–1108. https://doi.org/10.1017/S0305000909990067

Woodard K., Pozzan L., Trueswell J.C. (2016) Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology. 141*, 187–209. https://doi.org/10.1016/j.jecp.2015.08.005

Ylinen, S., Bosseler, A., Junttila, K., & Huotilainen, M. (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Developmental Science*, *20*(6). https://doi.org/10.1111/desc.12472

Yu, Y., Havron, N. & Fisher, C. (in preparation). Syntactic adaptation and word learning in 3- to 4-year-olds.

Yuan, S., & Fisher, C. (2009). "Really? She Blicked the Baby?" : Two-Year-Olds Learn Combinatorial Facts About Verbs by Listening. *Psychological Science*, *20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x

Yuan, S., Fisher, C., Kandhadai, P., & Fernald, A. (2011). You Can Stipe the Pig and Nerk the Fork : Learning to Use Verbs to Predict Nouns. In *Proceedings of the 35th annual Boston University conference on language development* (pp. 665-677). Boston, MA: Cascadilla Press.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the Nouns : Simple Structural Cues to Verb Meaning. *Child Development*, *83*(4), 1382–1399. https://doi.org/10.1111/j.1467-8624.2012.01783.x

Zettersten, M. (2019). Learning by predicting: How predictive processing informs language development. *Patterns in Language and Linguistics*, 255-288.

Zhang, F., Jaffe-Dax, S., Wilson, R. C., & Emberson, L. L. (2019). Prediction in infants and adults: A pupillometry study. *Developmental Science*, *22*(4), 1–9. https://doi.org/10.1111/desc.12780