



HAL
open science

Preverbal infants' sensitivity to grammatical dependencies

Mireille Babineau, Anne Christophe

► **To cite this version:**

Mireille Babineau, Anne Christophe. Preverbal infants' sensitivity to grammatical dependencies. *Infancy*, 2022, 27, pp.648 - 662. 10.1111/inf.12466 . hal-03873591

HAL Id: hal-03873591

<https://hal.science/hal-03873591>

Submitted on 27 Nov 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.

Preverbal infants' sensitivity to grammatical dependencies

Mireille Babineau a* and Anne Christophe b,c

a. Department of Psychology, University of Toronto, Canada

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

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

Manuscript accepted in *Infancy*Babineau, M. & Christophe, A. (2022). Preverbal infants' sensitivity to grammatical dependencies. *Infancy*, 27, 648-662. <https://doi.org/10.1111/inf.12466>**Abstract**

During their first months of life, infants can already distinguish function words (e.g., pronouns, determiners) from content words (e.g., verbs, nouns). Little research has explored preverbal infants' sensitivity to the relationships between these two-word categories. This preregistered study examines whether French-learning 8- and 11-month-olds track the grammatical dependencies between determiners and nouns and pronouns and verbs. Using the Visual Fixation Procedure, infants were presented with lists containing either grammatical (e.g., *tu manges* "you eat", *des biberons* "some bottles") or ungrammatical (e.g., *des manges* "some eat", *tu biberons* "you bottle") phrases. In Experiment 1 (N=59), the lists involved common nouns and verbs, while in Experiment 2 (N=28), only common verbs were used. Eleven-month-olds showed a clear preference for correct over incorrect co-occurrences in both experiments, while 8-month-olds showed a trend in the same direction. These results suggest that before their first birthday, infants' storage and access of words and word sequences is sufficiently sophisticated to include the means to track categorical dependencies. This early sensitivity to co-occurrence patterns may be greatly beneficial for constraining lexical access and later on for learning novel words' syntactic and semantic properties.

Keywords: infants, language acquisition, syntax, word categories, distributional learning**Introduction**

The early lexicon does not lack organization. Even before their first birthday, infants have gained substantial knowledge about word categories. They can distinguish function words (e.g., determiners, pronouns) from content words (e.g., nouns, verbs) during their first months of life

based on their acoustic, distributional, and phonological characteristics (e.g., Hochmann, Endress, & Mehler, 2010; Marino, Bernard, & Gervain, 2020; Shi, Morgan, & Allopenna, 1998; Shi & Werker, 2001; Shi, Werker, & Morgan, 1999). The distributional patterns of these word categories even guide infants' mapping strategies: they tend to associate referents (meaning) to infrequent syllables rather than to frequent syllables (Hochmann, Endress, & Mehler, 2010; Hochmann, 2013). Despite of the fact that content words are conceived as carrying the main semantic weight, while function words are thought as being mostly important for indicating grammatical structure, extensive work has shown that preverbal infants pay attention to both categories. On the one hand, function words represent a small group of items so frequent that preverbal infants recognize their phonological form (e.g., Shafer, Shucard, Shucard, & Gerken, 1998; Shi, Werker, & Cutler, 2006), and use them to segment co-occurring content words (Shi & Lepage, 2008). On the other hand, a large body of work has shown that around 11 months of age infants are also sensitive to the phonological forms of common nouns such as “bottle” and “ball” (French: Hallé & Boysson-Bardies, 1994; Hallé & de Boysson-Bardies, 1996; English: Vihman, Nakai, DePaolis, & Hallé, 2004; Dutch: Swingley, 2005; Hebrew: Segal, Keren-Portnoy, & Vihman, 2020). Even more remarkable findings have been found in the domain of lexical semantics: from about 6 months of age infants start to understand the meaning of frequent concrete content words (e.g., “mommy”, “hand”; Bergelson & Swingley, 2012; Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Tincoff & Jusczyk, 1999, 2012), and more abstract ones a few months later (e.g., “eat”; (Bergelson & Swingley, 2013, 2015). Since parents rarely produce words in isolation (e.g., “Hand!”), but rather include them in short phrases or sentences (e.g., “hold my hand”; Aslin, 1993; Brent & Siskind, 2001; van de Weijer, 1998), infants often encounter content words co-occurring with function words. This situation is advantageous, giving them the opportunity to learn how the different categories of words relate to one another.

Over the course of their second year of life, children develop more abstract syntactic knowledge enabling them to categorize novel words. This is possible because function words tend to be indicative of the word class of subsequent words (e.g., “the” is more likely to be followed by nouns than by verbs; e.g., Mintz, 2003). Experimental work has shown that at only 14 months of age infants can build syntactic expectations about the type of function words which is likely to co-occur with novel content words (Babineau, Shi, & Christophe, 2020; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Shi & Melançon, 2010). A novel word that co-occurs with determiners is expected to be used later on with another determiner, not a pronoun. Toddlers have also been shown to be sensitive to their native language's sub categorical dependencies (Van Heugten & Christophe, 2015) as well as to non-adjacent dependencies (Höhle, Schmitz, Santelmann, & Weissenborn, 2006; Santelmann & Jusczyk, 1998; van Heugten & Johnson, 2010). For instance, in van Heugten & Christophe (2015), French-learning 18-month-olds distinguished between grammatical and ungrammatical lists in which a determiner either matched or not the gender of the co-occurring familiar noun (e.g. grammatical: un_{MAS} biberon_{MAS}; ungrammatical: une_{FEM} biberon_{MAS}). From 18 months onward, not only are function words used to facilitate the recognition of familiar words (Cauvet et al., 2014; Gerken & McIntosh, 1993; Jones, Cabiddu, & Avila-Varela, 2020; Kedar, Casasola, & Lust, 2006; Van Heugten & Johnson, 2011; Zangl & Fernald, 2007), but they are also used to learn the meaning of novel nouns and verbs (He & Lidz, 2017; de Carvalho, He, Lidz, & Christophe, 2019). For instance, when presented with a video in which a novel animated animal is performing a novel action, French-learning 18-month-olds can use function words to

interpret a novel word such as *bamoule* as referring to the novel animal (if it is preceded by a determiner: *Regarde, c'est une bamoule!* “Look, it’s a bamoule”) or as referring to the novel action (if it is preceded by a pronoun: *Regarde, elle bamoule!* “Look, she’s bamouling!”; de Carvalho, He, Lidz, & Christophe, 2019). Preschoolers’ online processing of function words is even more efficient, enabling them to perform fast-mapping of novel words after only two repetitions (de Carvalho, Babineau, Trueswell, Waxman, & Christophe, 2019). Since function words convey information about the syntactic category of co-occurring words, they can also facilitate the processing of ambiguous words, such as homophones – words that share the same phonological form but have different meanings. Recent work has found that, in natural languages, pairs of homophones tend to differ in syntactic and/or broad semantic characteristics, more than would be expected by chance alone (e.g. the word “fly”, which designates a type of insect when used as a noun, and an action when used as a verb; the fly vs. they fly; Dautriche, Fibla, Fiévet, & Christophe, 2018). Toddlers make good use of the syntactic context of homophones, and experimental work has shown that they can easily learn that a novel noun homophonous with a known verb (a verb-homophone, e.g., *un manger* “an eat”) refers to a novel object (e.g., an animal) based on the co-occurring determiner (Dautriche et al., 2015; 2018). In sum, young children’s knowledge and reliance on function words offer a great advantage for constraining lexical access and facilitating lexical acquisition.

Before being able to use function words to build expectations about co-occurring content words’ syntactic and semantic class, infants must first notice the distributions of both word categories within their native language. Preverbal infants’ sensitivity to co-occurrence patterns is well-established. For instance, a large body of work has shown that after being exposed to an artificial language or an unknown natural language, 8-month-old infants (and even younger ones) discriminate between frequent strings and infrequent strings of syllables (e.g., Pelucchi, Hay, & Saffran, 2009; Saffran, 2020; Saffran, Aslin, & Newport, 1996). Around one year of age, infants have also been shown to build abstract representations of word categories after being exposed to a short artificial language by discriminating new grammatical strings from ungrammatical ones (Gómez & Gerken, 1999; Saffran et al., 2008). In the real world, with a language containing a much larger vocabulary following more complex rules and an exposure spreading across time, tracking co-occurrence patterns is undeniably a much more arduous task. However, around the time when infants have well-developed abilities to track the distribution of words in their input, they also have a handful of words in their lexicon. Based on the *semantic seed* hypothesis (Brusini, Seminck, Amsili, & Christophe, 2021; Christophe, Dautriche, de Carvalho, & Brusini, 2016; Gutman, Dautriche, Crabbé, & Christophe, 2014, see also Christodouloupoulos, Roth, & Fisher, 2016), preverbal infants might start tracking the contexts in which this small group of known words occur to build proto-syntactic categories. That is, by the end of their first year of life, infants might start organizing their early lexicon by grouping words into basic semantic categories (e.g., object, action; Carey, 2009) and track the contexts in which these known words appear (e.g., **this is a car**; **this is a ball**; **she is eating**, **she is sleeping**). They could then use this information to infer coarse semantics for new words they encounter in the same syntactic contexts: **this is a dax** \Rightarrow dax possibly refers to an object, while **she is daxing** \Rightarrow dax possibly refers to an action. This end point represents the mechanism known as *syntactic bootstrapping* (Gleitman, 1990; Fisher, Gertner, Scott, & Yuan, 2010). Even though the *semantic seed* hypothesis has been proven feasible based on computational modeling (Brusini, Seminck, Amsili, & Christophe, 2021) and recent experimental work with toddlers and preschoolers (Babineau, de Carvalho, Trueswell, & Christophe, 2021; Barbir,

Babineau, Fiévet, & Christophe, 2019), no research has yet explored the earliest stage of the mechanism during which preverbal infants develop their sensitivity to the relationship between function words and familiar content words. Based on past work showing that from 6 months of age infants use familiar content words to facilitate the segmentation of co-occurring words (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Mersad & Nazzi, 2012), we posit that these familiar words also make their immediate context more salient, which would heighten their sensitivity to grammatical dependencies. Given preverbal infants' ability to track linguistic regularities and their blossoming knowledge of lexical semantics, how much attention have they paid to the immediate context of their first words, such as “bottle” and “eat”?

The current study aims at shedding light on preverbal infants' sensitivity to the co-occurrence patterns between function words and common content words. We investigated whether French-learning 8- and 11-month-olds track and store common nouns and verbs along with the preceding function words. Using a Visual Fixation Procedure, infants were presented with lists containing either grammatical or ungrammatical phrases. In Experiment 1, the grammatical lists involved nouns preceded by a determiner and verbs preceded by a pronoun (e.g., *des biberons* “some bottles”, *tu manges* “you eat”) and the ungrammatical lists involved nouns preceded by a pronoun and verbs preceded by a determiner (e.g., *tu biberons* “you bottle”, *des manges* “some eats”). In Experiment 2, only common verbs were used, either in grammatical contexts (preceded by pronouns) or in ungrammatical contexts (preceded by determiners). If infants are indeed sensitive to noun and verb grammatical dependencies, a preference for the grammatical over the ungrammatical lists is expected in both experiments, based on previous work pitting grammatical and ungrammatical sentences using similar procedures (e.g., Santelmann & Jusczyk, 1998; Van Heugten & Christophe, 2015; Van Heugten & Johnson, 2010; Van Heugten & Shi, 2010). We hypothesize that infants from both age groups can succeed in this task based on their proven word recognition skills (e.g., Hallé & Boysson-Bardies, 1994; Shi & Lepage, 2008), and their notorious ability to track simple co-occurrence patterns (e.g. Pelucchi et al., 2009; Saffran et al, 1996).

Experiment 1

Method

The study reported in this paper, including the entire method, analyses, and criteria for exclusion of participants were pre-registered on the OSF (Open Science Framework) database before running the experiment. The preregistration can be accessed with the following link: https://osf.io/rcv5k/?view_only=0e69c74638014eb6b0649984d95ef5ac. The materials, collected data, and data analyses are also freely available to readers through this link.

Participants

A total of 28 French-learning 8-month-olds (15 female; age range=229-258 days; M=241) and 31 French-learning 11-month-olds (10 female; age range = 318-347 days; M=332) participated in the experiment. According to the parents, none of these children was known to have any language or hearing problem. An additional nine infants were tested, but excluded from the analyses due to extreme fussiness/crying (8-month-olds: 2; 11-month-olds: 4),

parental interference (8-month-olds: 1; 11-month-olds: 1), or technical issue (11-month-olds: 1).

The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. Since our study involves human subjects, it was carried out in accordance with the recommendations of our local ethics committee, which approved our protocol beforehand.

Stimuli

Twelve common words (eight monosyllabic and four multisyllabic) were used in this study. Half of these words are nouns and the other half verbs; nouns: *balles* “balls”, *bananes* “bananas”, *biberons* “bottle”, *mains* “hands”, *pieds* “feet”, *pommes* “apples”; verbs: *descends* “go down”, *donnes* “give”, *dors* “sleep”, *manges* “eat”, *regardes* “look”, *viens* “come” (see Table 1). Several of the chosen nouns were used in past studies studying preverbal infants’ phonological representation of common words (Hallé & Boysson-Bardies, 1994; Hallé & Boysson-Bardies, 1996), while the other nouns (and some of the chosen verbs) were used in studies investigating preverbal infants’ word mapping (e.g., Bergelson & Swingley, 2012; 2013; 2015). High-frequency French functors (*des* /de/ “some”, indefinite plural article; *tu* /ty/ “you” second person singular personal pronoun) preceded the content words, creating grammatical or ungrammatical phrases. A female native speaker of French (the last author) recorded the materials. To avoid recording ungrammatical utterances, potentially resulting in unnatural tokens, the stimuli were cross-spliced. For each item used in the study, three versions were recorded. One version contained the target word preceded by the correct function word (e.g., *tu manges* “you eat”). From this version, the content word was spliced. In the other two versions, two content words (one verb, one noun) starting with the same consonant-vowel sequence, were recorded preceded by the correct function word for their grammatical category (e.g., *des mangues* “some mangos”, *tu manques* “you miss”). From these two tokens, the function word was spliced out from the phrase, just before the onset of the content word. The ungrammatical and grammatical phrases were formed by concatenating the two function words with the content word, such that the grammatical phrase contained the correct function word (e.g., *tu manges*) and the ungrammatical one contained the incorrect function word (e.g., *des manges*). The two phrases were similar in duration and intonation (average correct phrase length: 947 ms; average incorrect phrase length: 929 ms).

Table 1. Stimuli for Experiment 1.

Grammatical	Ungrammatical
<i>des balles</i> “some balls”	<i>tu balles</i> “you ball”
<i>tu dors</i> “you sleep”	<i>des dors</i> “some sleep”
<i>tu viens</i> “you come”	<i>des viens</i> “some come”
<i>des pieds</i> “some feet”	<i>tu pieds</i> “you foot”
<i>des biberons</i> “some bottles”	<i>tu biberons</i> “you bottle”

<i>tu regardes</i> “you look”	<i>des regardes</i> “some look”
<i>des pommes</i> “some apples”	<i>tu pommes</i> “you apples”
<i>tu manges</i> “you eat”	<i>des manges</i> “some eat”
<i>tu descends</i> “you go down”	<i>des descends</i> “some go down”
<i>des mains</i> “some hands”	<i>tu mains</i> “you hand”
<i>tu donnes</i> “you give”	<i>des donnes</i> “some give”
<i>des bananes</i> “some bananas”	<i>tu bananes</i> “you bananas”

*Note that *des* ‘some’ is an indefinite plural article that would be dropped in English (e.g., *Il y a des vaches dans le pré* ‘There are *cows in the field’).

Eight lists were created, four of which contained the correct phrases and four of which contained incorrect phrases. Each list contained two unique tokens of each of the twelve short phrases. These phrases were ordered differently across each of the correct and incorrect lists, but the order was identical across the two conditions. Interstimulus pauses were approximately 750 msec long. All lists lasted 40.7 sec.

Visual stimuli included a colorful checkerboard for all trials. As an attention-getter to attract infants’ attention towards the screen between trials, we used a video of a butterfly perched on a leaf that is accompanied by a bubble sound.

Procedure

Two measures of infants’ word knowledge were obtained from the caregivers before or during their visit to the laboratory: the French version of the MacArthur–Bates Communicative Development Inventory (CDI; short form for 12-month-olds: Kern, Langue, Zesiger & Boone, 2010), and an item exposure survey asking parents to estimate how often they tend to use these items in their day-to-day life when they speak to their child using the following scale: 1=rarely/once a week; 2=few times a week; 3=everyday.

The Visual Fixation Procedure was used to test infants’ spontaneous listening preferences. Infants were individually tested in a sound-attenuated double-walled booth. During the experiment, they sat on their caregiver’s lap, approximately 70 cm from a 27-inch monitor. Stimuli were presented using the Habit software (Cohen, Atkinson, & Chaput, 2004). Two loudspeakers positioned on each side of the TV screen presented the audio stimuli. The experimenter monitored the infant’s behavior through a video camera that was set above the monitor in the booth and which was connected to a second monitor placed outside the booth. This set-up enabled the experimenter to observe infants’ behavior from an adjacent room without being aware of the stimuli presentation. The procedure started by displaying the attention getter on the screen. Once the infant oriented toward the screen, the experimenter initiated the first trial, and she pressed another key whenever a look toward the monitor occurred during a trial. Trials ended either after the infant had looked away from the screen for more than two consecutive seconds or after the maximum trial length was reached. A new trial started when the infant looked back towards the screen. The experiment finished when all eight

lists were presented. Note that the same lists were presented to all infants, but the order in which these lists appeared was pseudo-randomized so that no more than two consecutive trials of the same type (i.e. grammatical vs ungrammatical) were presented.

Results

The vocabulary checklist was filled out by the majority of caregivers (8-month-olds: 24/28; 11-month-olds: 25/31). Parental reports reveal that infants were not perceived as understanding most of the words used in the study (mean number of words scored as being comprehended by 8-month-olds: 1.17 of 12; range 0-5; mean number of words scored as being comprehended by 11-month-olds: 3.8 of 12; range 0-12). None of the parents of the 8-month-olds reported that their child was producing any of the words used in the experiment, but two target words were produced by a few 11-month-olds (*balle* “ball” was produced by 2 children, *banane* “banana” by 1 child). Previous work with preverbal infants has shown that word knowledge is mostly not apparent to parents (e.g., Bergelson & Swingley, 2012), so even though parents reported that most of the words used in our study are not comprehended, it does not mean that participants are clueless about the phonological form or the meaning of these common words. Furthermore, the survey confirmed that the majority of participants are exposed to the items on a regular basis at home (mean score for 8-month-olds: 2.23 out of 3; range: 1.58-2.5; for 11-month-olds: 2.31 out of 3; range: 1.92-2.67). On average, the 12 items received a score of 3 (everyday use) from 57% of the 11-month-olds’ caregivers (score of 1: 9%; score of 2: 34%), and from 53% of 8-month-olds’ caregivers (score of 1: 14 %; score of 2: 33%). Specific items were reported as rarely used by 8-month-olds’ caregivers, with 50% of parents reported rarely using the noun *balle* “ball” and the verb *descendre* “go down” (with the remaining 10 words being used frequently by the majority of caregivers)¹. Detailed information about the survey can be found on the OSF page.

The dependent variable for analysis was the average time that infants spend looking at the visual stimuli during each trial type (i.e. the average looking time for the four grammatical lists vs the average looking time for the four ungrammatical lists). An ANOVA² with Grammaticality (grammatical vs ungrammatical lists) as a within-participant factor, and Age (8 vs 11 months) as a between-participants factor, revealed a significant effect of the Grammaticality ($F(1,57)=8.44$, $p=0.005$, $\eta_p^2=0.129$), but no effect of age ($F(1,57)=1.73$, $p=0.2$, $\eta_p^2=0.029$), nor an interaction between age and grammaticality ($F(1,57)=0.21$, $p=0.65$, $\eta_p^2=0.004$). As planned in our pre-registration, we used paired-sample t -tests (two-tailed) for each age group and found a significant difference only with the 11-month-olds ($t(30)=-2.686$, $p=.012$; $d=.482$), but not with the 8-month-olds ($t(27)=1.545$, $p=0.134$, $d=0.29$). Nevertheless,

¹ We measured parents’ perception of input frequency, i.e. how often they think they are using these words when they talk to their child. This measure has its limits. For instance, the majority of participants attend a full-time care service (e.g. nanny, daycare). Hence, the item exposure survey is only giving us a glimpse into half of the language input of each child.

² As stated in our preregistration, we initially intended to test solely 11-month-olds. After obtaining successful results with this age group, we wrote an amendment indicating our intention to test 8-month-olds with the same task. Hence, based on the preregistration, only the paired-sample t -tests can be considered as planned analyses, and the ANOVA was decided upon later. We present the ANOVA first because a reviewer mentioned that it was weird to present it after the paired t -tests.

as shown in Figure 1, infants in both age groups looked toward the screen for a longer time during grammatical trials than during ungrammatical trials (8-month-olds: $M= 13.24$ ($SE=1.57$) vs. $M= 11.6$ ($SE= 1.2$); 11-month-olds: $M=11.548$ ($SE=1.03$) vs. $M= 9.3$ ($SE=0.84$)³. Most infants showed this pattern of looking time, i.e. 18 out of 28 8-month-olds (with 7 showing the opposite pattern and 3 not showing a clear preference⁴) and 20 out of 31 11-month-olds (with 6 showing the opposite pattern and 5 not showing a clear preference).

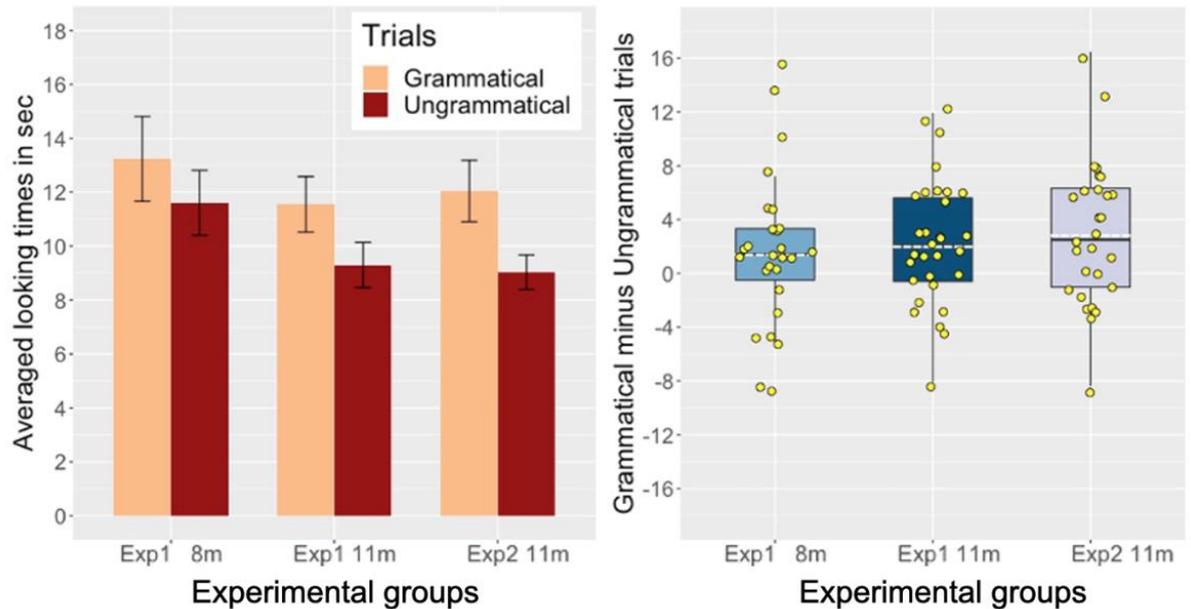


Figure 1. On the left side: Mean looking time in seconds during the two trial types, i.e. grammatical (in orange) and ungrammatical (dark red), for the three experimental groups (Experiment 1: 8-month-olds, Experiment 1: 11-month-olds, Experiment 2: 11-month-olds). Error bars represent the standard error of the mean. On the right side: Box plot of the difference between looking times during the grammatical and the ungrammatical test trials for the three conditions. The black line represents the median and the white dotted line represents the mean. Each yellow dot represents one participant. Longer looking times for grammatical over ungrammatical lists is indicated by positive values.

Discussion

By 11 months of age, infants display a robust preference for correctly used over incorrectly used function words, indicating that they have managed to track the co-occurrence

³ Following a reviewer's suggestion, we conducted an exploratory analysis to see whether the size of the lexicon, measured via parental reports, had an impact on infants' performance, but found no correlation between vocabulary size and infants' preference for grammatical lists, in either age group ($p>.5$).

⁴ Following the suggestion of a reviewer, we ran exploratory binomial tests on these numbers. Infants who had a difference of less than 1 sec between their averaged looking time during the grammatical vs ungrammatical trials were removed from this analysis, which showed that both 8-month-olds and 11-month-olds significantly preferred to listen to grammatical lists than to ungrammatical lists ($p=0.043$; $p=0.009$, respectively). When all infants were included (however small their difference in looking times), 8-month-olds were found to significantly prefer the grammatical lists ($p=0.036$), but 11-month-olds only showed a trend ($p=0.071$).

between determiners and common nouns and between pronouns and common verbs during this early stage of language acquisition. This preference might be emerging in 8-month-olds, who showed a trend in the same direction. Even though no interaction between Age and Grammaticality was found, some caution is in order when interpreting 8-month-olds' results as infants' preference for the grammatical lists was not reliable enough to reach significance with the planned *t* test.

Since both nouns and verbs were used in Experiment 1, it fails to reveal if these young learners know noun and verb contexts to the same extent. Infants could have succeeded at discriminating grammatical from ungrammatical lists on the sole basis of their deeper knowledge of the nouns used in the study, and/or of their syntactic contexts. Across languages, nouns tend to constitute the predominant form in early lexicons (Braginsky, Yurovsky, Marchman, & Frank, 2019; Gentner, 1982; Saah et al., 1996), especially frequent concrete nouns and those strongly associated with babies (e.g. “bottle”). By 11 months, most infants have successfully mapped a few nouns to their referents, but they have established fewer mappings for words from other form classes (such as verbs). For instance, based on the parental reports of the 11-month-olds who participated in Experiment 1, an average of 7.33 concrete nouns (range: 0-33) and of 4.38 verbs (range: 0-16) were comprehended. This asymmetry was not found in 8-month-olds, who had overall smaller receptive vocabulary, with an average of 1.25 concrete nouns (range: 0-7) and 1.33 verbs (range: 0-8)⁵. Since the “noun bias” is partly shaped by the linguistic and cultural environment, it might become more apparent as the child gains more exposure to his native language and as the size of his lexicon increases (e.g., Tardif, Gelman, & Xu, 1999). When it comes to noticing the syntactic contexts of these lexical items, it might be easier to track the contexts of common nouns than to track the contexts of common verbs. In French, nouns appear in more stable contexts (i.e., consistently co-occur with a preceding determiner in 71% of cases or with a preceding adjective in 26% of cases) than do verbs (i.e., co-occur with a preceding subject pronoun in 59% based on a corpus by Cécylre & Shi, 2005). Nonetheless, the co-occurrence of verbs with pronouns might still be noticeable early on, based on recent work showing French-learning 14-month-olds' success at building syntactic expectations for both novel nouns and novel verbs (Babineau, Shi & Christophe, 2020). Specifically, in that study, infants expected a novel verb that previously co-occurred with subject pronouns to co-occur with another subject pronoun, but not with a determiner. These results demonstrate that French-learning 14-month-olds are sensitive to the co-occurrence patterns of verbs with subject pronouns, which has clear benefits for the processing of novel words. Experiment 2 aims to test if younger infants (11-month-olds) are already tracking the co-occurrence patterns of common verbs with pronouns. If infants succeed at discriminating grammatical from ungrammatical lists, this will indicate that preverbal infants' verb knowledge and their ability to track their syntactic contexts is more sophisticated than previously thought.

Experiment 2

⁵ Proper nouns (such as infants' own names, “mommy” and “daddy”) as well as social nouns (such as *bonjour* “hello” and *au revoir* “bye”) were not counted as concrete nouns, but they were frequently known by the 8-month-olds.

Method

Participants

A total of 28 French-learning 11-month-olds (8 female; age range = 322-362 days, $M=338$) participated in the experiment. Parents reported no hearing problems. An additional nine infants were tested, but excluded from the analyses due to extreme fussiness/crying (6) or parental interference (3).

Stimuli

Twelve common consonant-initial verbs (nine monosyllabic, three bisyllabic) were used in Experiment 2: *cache* “hide”, *chante* “sing”, *dort* “sleep”, *descend* “go down”, *dessine* “draw”, *donne* “give”, *lave* “wash”, *mange* “eat”, *monte* “go up”, *regarde* “look”, *saute* “jump”, *vient* “come”. These words were chosen based on early production and comprehension norms in older children from the same population (children who participated in previous experiments at the lab), as well as for their associations with common activities and routines. However, many highly commonly known verbs were excluded because of the presence of frequent noun homophones (e.g., *couche* can mean “to lay” or “a diaper”, *lit* can mean “to read” or “a bed”, etc.). High-frequency French functors (the masculine determiner *un* and the feminine determiner *une* “a”, the pronouns *elle* “she” and *il* “he”) preceded the verbs, creating grammatical or ungrammatical phrases (see Table 2).

The stimuli were recorded by the same speaker as in Experiment 1. As in Experiment 1, the stimuli were cross-spliced. For each item used in the study, three versions were recorded. One version contained the target word preceded by the correct function word (e.g., *elle mange* “she eats”). From this version, the content word was spliced. In the other two versions, two content words (one verb, one noun) starting with the same consonant-vowel sequence, were recorded preceded by the correct function word for their grammatical category (e.g., *une mangue* “a mango”, *elle manque* “she misses”). From these two tokens, the function word was spliced out from the phrase, just before the onset of the content word (or after the onset consonant for words starting with a fricative). The ungrammatical and grammatical phrases were formed by concatenating the two function words with the content word, such that the grammatical phrase contained the correct function word (e.g., *elle mange*) and the ungrammatical one contained the incorrect function word (e.g., *une mange*). The two phrases were similar in duration and intonation.

As in Experiment 1, a total of eight lists were created, four with correct (grammatical) short phrases and four with incorrect (ungrammatical) short phrases. Each list contained two unique tokens of each of the twelve short phrases. These phrases were ordered differently across each of the four correct and incorrect lists. Interstimulus pauses were approximately 750 msec long. All lists lasted 43.6 sec.

Visual stimuli and the attention-getter were the same as in the previous pre-registered study.

Table 2. Stimuli for Experiment 2.

Experiment 2

<u>Grammatical</u>	<u>Ungrammatical</u>
<i>elle cache</i> ‘hide’	<i>une cache</i>
<i>il dort</i> ‘sleep’	<i>un dort</i>
<i>elle descend</i> ‘go down’	<i>une descend</i>
<i>elle lave</i> ‘wash’	<i>une lave</i>
<i>elle mange</i> ‘eat’	<i>une mange</i>
<i>il dessine</i> ‘draw’	<i>un dessine</i>
<i>il donne</i> ‘give’	<i>un donne</i>
<i>elle monte</i> ‘go up’	<i>une monte</i>
<i>il regarde</i> ‘look’	<i>un regarde</i>
<i>il saute</i> ‘jump’	<i>un saute</i>
<i>il vient</i> ‘come’	<i>un vient</i>
<i>elle chante</i> ‘sing’	<i>une chante</i>

Results

As in Experiment 1, two measures of infants’ word knowledge were obtained from their caregivers. However, these measures offer limited information due to the low numbers of questionnaires⁶: out of 28 participants, we gathered 12 French CDIs (short form) and 10 item surveys. As in Experiment 1, parental reports reveal that infants were not perceived as understanding most of the words used in the study (mean number of words scored as being comprehended: 2.83 of 12; range 0-12) and none of the infants produced the items. The survey confirmed that most participants are exposed to the items on a regular basis (mean score: 2.13 out of 3; range: 1.83-2.42). On average, the 12 items received a score of 3 (everyday use) from 52% of the 11-month-olds’ caregivers (score of 1: 24%; score of 2: 24%). Eight of the 12 verbs were reported as being used frequently (i.e. score of 2 or 3) by more than 70% of parents. For *descendre* “go down” and *monter* “go up” more than 60% of caregivers used them frequently, while below 20% of caregivers reported using *dessiner* “to draw” and *sauter* “to jump” (detailed results can be found on OSF). Despite the rare use of two out of twelve items, the survey confirmed that participants are exposed to the majority of target verbs on a regular basis.

Infants oriented toward the screen for an average of 12.05 sec during the grammatical trials and for an average of 9.024 sec during the ungrammatical trials, with 18 out of 28 infants showing this listening preference (7 infants showing the opposite preference and 3 infants not showing a clear preference; see Figure 1). A paired-sample *t* test (two-tailed) comparing the

⁶ During the testing phase of Experiment 2, the authors moved their office space. Sadly, we did not manage to recover the misplaced CDIs and item surveys.

average looking time during the grammatical vs the ungrammatical trials revealed a significant difference, $t(27) = 3$; $p = .006$; $d = .567$).⁷

Discussion

The results of Experiment 2 replicate the ones obtained in Experiment 1, confirming 11-month-olds' ability to track co-occurrence patterns involving pronouns and common verbs. Hence, although 11-month-olds might seem to know much more about nouns than about verbs based on their caregivers' report, they are nonetheless sensitive to common verbs' co-occurrence patterns.

General Discussion

The goal of the current set of experiments was to specify the scope of preverbal infants' sensitivity to grammatical dependencies involving familiar nouns and verbs. Using a Visual Fixation Procedure, 8- and 11-month-olds were presented with two types of lists containing either grammatical or ungrammatical phrases. Common nouns and verbs were used in Experiment 1, while only common verbs were used in Experiment 2. Eleven-month-olds showed a significant preference for grammatical lists in both experiments, while 8-month-olds showed a non-significant trend in the same direction. Future work should test the scope of 8-month-olds' ability to track adjacent grammatical dependencies. All things considered, our findings indicate that during the second part of their first year of life, infants can track the co-occurrence patterns naturally present in their native language and more specifically, they develop a precocious sensitivity to the relationship between determiners and nouns, as well as pronouns and verbs.

Infants' familiarity with the individual content words and their immediate contexts is likely the driving force of the effect observed in the current set of experiments. Even though both 8-month-olds and 11-month-olds can quickly pick up speech patterns and regularities in laboratory settings, two obvious advantages of the 11-month-olds is their larger lexicon and their lengthier exposure to their native language. A recent meta-analysis on word-form recognition has reported an increase in effect sizes with age and items' familiarity (Carbajal, Peperkamp, & Tsuji, 2021). Hence, 11-month-olds are better equipped to distinguish familiar words from unfamiliar ones, and as found here, they seem to also be better at distinguishing grammatical from ungrammatical contexts for such familiar word forms. Note, however, that we should remain cautious in reporting this latter advantage, since no interaction between Age and Grammaticality was found in Experiment 1. This indicates that the participating 8-month-olds and 11-month-olds did not behave differently during the task (both age groups showed a preference for grammatical over ungrammatical lists). As proposed by the *semantic seed* hypothesis (e.g., Christophe, Dautriche, de Carvalho, & Brusini, 2016), the early lexicon has an essential role in language acquisition: it enables infants to build proto-syntactic categories. We show that around 11 months of age infants are tracking the contexts in which their first words appear. This helps them to notice what type of function words tend to precede certain

⁷ An exploratory binomial test confirmed the significance of this preference ($p = 0.043$).

categories (e.g., “the” and “a” before object words such as “bottle”, “ball”, and “car”). This knowledge can be used just a few months later to build expectations about novel words’ syntactic (e.g., Babineau, Shi, & Christophe, 2020) and semantic properties (e.g., de Carvalho, He, Lidz, & Christophe, 2019).

Our findings are also in line with usage-based accounts of language acquisition (e.g., Arnon, 2010; Arnon & Christiansen, 2017; Abbot-Smith & Tomasello, 2006; Bannard & Lieven, 2012; McCauley & Christiansen, 2019; Tomasello, 2003): from early on, infants extract and represent multiword sequences alongside and in parallel with individual words. While van Heugten & Christophe (2015) has shown that French-learning 18-month-olds track gender marking between determiners and common nouns, we show that 11-month-olds (and possibly even younger infants) are also sensitive to simple co-occurrence patterns involving common nouns, as well as those involving common verbs. Our findings are building on recent work showing that 11-month-olds are sensitive to frequent sequences of words (e.g., the frequent sequence “clap your hands” vs the infrequent sequence “hold your hands”; Skarabela, Ota, O’Connor, & Arnon, 2021), and nonwords (e.g., the frequent string *dans la* [dãla] “in the” vs the infrequent string [dalã]; Ngon et al., 2013). It remains unknown if 11-month-olds’ success in our task could be linked to their blossoming sensitivity to the grammaticality of the phrases, and not purely on their well-developed ability to distinguish frequent sequences of words from infrequent ones. After all, just a few months later, at 14 months, they expect novel content words to co-occur with specific function words (e.g., Babineau, Shi & Christophe, 2020). More work is needed to uncover the abstractedness of preverbal infants’ representations as well as the sources of their sensitivity to grammatical and sub-grammatical dependencies.

In sum, our findings provide the earliest account of infants’ ability to track co-occurrence patterns between different word categories from their natural language input. Before one year of age, infants’ syntactic knowledge is rudimentary, but they have nonetheless noticed simple yet powerful syntactic regularities which can bootstrap their lexical acquisition.

References

- 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., 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., 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>
- 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.
- Bergelson, E., & Swingle, 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>
- 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](https://doi.org/10.1016/S0010-0277(01)00122-6)
- Brusini*, P., Seminck*, O., Amsili, P., & Christophe, A. (2021). 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.
- 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., Nespors, 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>
- Dautriche I., Swingley D. & Christophe A. (2015). Learning novel phonological neighbors: syntactic category matters. *Cognition*, 143, 77–86. <https://doi.org/10.1016/j.cognition.2015.06.003>
- 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>
- 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>
- 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>

- Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, 1, 3–55.
- 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](https://doi.org/10.1016/S0010-0277(99)00003-7)
- 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>
- 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>
- Hochmann, J.-R. (2013). Word frequency, function words and the second gavagai problem. *Cognition*, 128(1), 13–25. <https://doi.org/10.1016/j.cognition.2013.02.014>
- 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/s15473341l1d0204_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
- Marino, C., Bernard, C., & Gervain, J. (2020). Word frequency is a cue to lexical category for 8-month-old infants. *Current Biology*, 30(8), 1380–1386.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117. [https://doi.org/10.1016/S0010-0277\(03\)00140-9](https://doi.org/10.1016/S0010-0277(03)00140-9)
- Ngon, C., Martin, A., Dupoux, E., Cabrol, D., Dutat, M., & Peperkamp, S. (2013). (Non) words,(non) words,(non) words: evidence for a protolexicon during the first year of life. *Developmental Science*, 16(1), 24–34.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, 113(2), 244–247.
- Saffran, J. R. (2020). Statistical language learning in infancy. *Child development perspectives*, 14(1), 49–54.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- 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](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, R., 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](https://doi.org/10.1207/s15327078in1002_5)
- Shi, R., 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](https://doi.org/10.1016/S0010-0277(99)00047-5)
- 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>
- Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the “noun bias” in context: A comparison of English and Mandarin. *Child development*, *70*(3), 620-635.
- 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>
- van de Weijer, J. (1998). *Language input for word discovery*. Max Plank Institute for Psycholinguistics.
- van Heugten, M., & Christophe, A. (2015). Infants' acquisition of grammatical gender dependencies. *Infancy*, *20*(6), 675–683.