8-month-old infants' ability to process word order is shaped by the amount of exposure

Caterina Marino, Judit Gervain

8-month-old infants' ability to process word order is shaped by the amount of exposure

Caterina Marino[1] & Judit Gervain[1, 2]

[1]Integrative Neuroscience and Cognition Center (INCC UMR8002), Université de Paris & CNRS, Paris, France

[2]University of Padua, Padua, Italy

Corresponding author:

Judit Gervain

DPSS

Università degli Studi di Padova

via Venezia 8

35128

Padova (PD)

Italy

judit.gervain@unipd.it

Word count: 5656 words

Abstract

In the majority of languages, the functional distinction between functors and content words correlates with lower-level, perceptually observable properties. Functors are generally more frequent and prosodically more minimal than content words. Previous studies demonstrate that the frequency distribution and the different acoustic realization of frequent and infrequent words guide infants in discovering their native word order. However, whether and if yes, how the exact frequency ratio impacts infants' ability to recognize function and content words and their relative order has never been explored. Here we investigate this by testing whether with a small ratio between functors' and content words' frequency, 1:3 as opposed to the 1:9 ratio in previous studies, French 8-month-olds are able to establish the functor-initial word order typical of their native language (Experiment 1) and whether prosody (Experiment 2) and the amount of exposure (Experiment 3) modulate this ability. We observed that infants exhibited the predicted functor-initial preference only when they were exposed to a short familiarization phase, i.e. reduced exposure. This suggests that different amounts of information selectively trigger different processing mechanisms, and little exposure may favor the extraction of regularities.

Keywords: word order, word frequency, lexical categorization, artificial grammar, infants, rule extraction.

1. Introduction

Learning word order is one of the fundamental achievements that infants have to accomplish to comprehend and produce their first multiword utterances. Language typology generally distinguishes between two types of languages. In functor-final languages, like Turkish, Basque and Japanese (e.g. Japanese: *Tokyo* ***ni*** Tokyo to 'to Tokyo'), function words are placed in the final position of morphosyntactic units. By contrast, in functor-initial languages (e.g. Italian, French and English), functors are typically placed in the initial position (e.g. Italian: *a Venezia* (to Venice). The position of functors in grammatical phrases correlates with a large number of other word order phenomena. Thus functor-final languages have an Object-Verb (OV) order, use postpositions etc., while functor-initial languages have a VO order, prepositions etc. (Dryer 1992). Knowing the relative order of functors and content words in the native language can, therefore, provide young language learners with an important "bootstrapping" cue to break into syntax. Such a bootstrapping strategy has indeed been identified in infants as young as 8 months of age (Gervain et al, 2008). However, many aspects of the nature of this mechanism remain unknown. In the current study, we ask how the complexity of the input selectively triggers the frequency-based bootstrapping of word order.

1.1 The frequency-based bootstrapping of functors and content words and their relative order

The distinction between functors, which indicate grammatical structure (e.g. articles, pronouns, preposition etc.), and content words, which carry lexical meaning (e.g. nouns, verbs, adjectives etc.), is a linguistic universal (Chomsky, 1995; Fukui 1986; Abney, 1987). Theoretical considerations and experimental studies suggest that language would be more difficult to learn without the presence of functors. Indeed, child and adult participants learn artificial grammars without functors less well than artificial grammars that contain functors (e.g. Braine, 1966; Green, 1979; Morgan et al., 1981; although learning may be possible without functors, Reeder et al. 2013).

Relevantly for language acquisition, in addition to the functional difference between functors and content words, the two grammatical classes also differ in a number of other properties. On a series of phonological measures, functors have been found to be more reduced or to have weaker prosodic prominence compared to content words, (e.g. Nespor & Vogel, 1986; Selkirk, 1984; Morgan, Shi & Allopenna, 1996). This provides a perceptually readily available cue to distinguish the two classes and infants are indeed sensitive to these acoustic differences already at birth (Shi, Werker & Morgan, 1999). Furthermore, the two classes also differ in their frequency of occurrence: while functors are highly frequent, content words have mid-to-low frequency of occurrence (Kucera & Francis 1967, Gervain et al. 2008), which constitutes another perceptually accessible cue. Lastly, while content words come in large open classes, into which new items can be easily added (e.g. *tag*, *spam*, *app*), functors constitute closed lexical classes, into which new items cannot be added without major language changes.

Importantly, word frequency and the phonological make-up of words are low-level, perceptually available cue that even young infants are sensitive to. On the basis of these cues, infants may establish the categories of functors and content words and determine their relative order in the input, which then helps them bootstrap the basic word order of their native language.

In a set of artificial grammar-learning studies using the headturn preference procedure (HPP) it has been demonstrated that sensitivity to the relation between word frequency, phonology and word order is established already pre-lexically (e.g. Gervain et al., 2008; Bernard & Gervain, 2012; Gervain & Werker, 2013; Marino et al., 2020). In these studies, infants were familiarized with artificial grammars containing a strict alternation of frequent and infrequent words, mimicking functors and content words, respectively and were then tested on their looking time preferences when exposed to test items presenting opposite word orders. For familiarization, different grammars were created with the same basic underlying structure consisting of the concatenation a basic unit of four CV syllables, i.e. an AXBY structure. In this AXBY structure, A and B were frequent words (mimicking functors in real language), whereas X and Y were infrequent words (mimicking content words). Specifically, the A and B categories had one token each (*fi* and *ge*, respectively), while the X and Y categories contained 9 tokens each (X: *mu, fo*, etc; Y: *ka, bi*, etc). This resulted in a stream in which A & B items were 9 times more frequent than the X & Y items, i.e. the frequency ratio was 1:9. This basic unit was repeated 243 times (each possible _X_Y syllable combination was used 3 times), resulting in an approximately 4-minute-long familiarization stream. By ramping the amplitude of the initial and final 15 seconds of the stream, the phase was suppressed, i.e. the initial item was made imperceptible, rendering the underlying structure of the stream ambiguous between a frequent-initial and a frequent-final parse. During the tests, infants were presented with eight different test items, four with a frequent-initial structure (AXBY or BYAX) and four with a frequent final structure (XBYA or YAXB). Importantly, all test items appeared in the familiarization stream, i.e. there was no difference in familiarity between the frequent-initial and frequent-final test items. Any preference between the two necessarily resulted from the knowledge and processing mechanisms infants brought to the task.

8-month-olds exposed to languages with opposite word orders, functor-initial Italian and functor-final Japanese, exhibited opposite preferences in the task. Japanese infants preferred the test items that mirrored the order of their functor-final language (frequent items in final position), whereas Italian infants preferred functor-initial items (Gervain et al., 2008). Infants could thus use the frequency of words as a cue and select the relative order of frequent and infrequent words that characterizes their native language. Crucially, Japanese and Italian infants were familiarized with and tested on exactly the same material. The observed difference in their word order preference during test must, therefore, be attributable to knowledge they brought to the task.

While these results establish that infants are familiar with the relative order of frequent and infrequent words in their native language, whether they represent them as the actual grammatical categories of functors and content words, with all their characteristics or just groups of words with different frequencies remains, however, unanswered.

This issue was addressed in follow-up studies that demonstrated that infants expect the categories of frequent and infrequent words to have all the typical distinctive features of the categories of functors and content words, respectively. A recent study (Marino et al., 2020) demonstrated that after being exposed to the same artificial grammar as in Gervain et al. (2008), French 8-month-olds expected infrequent words to belong to open classes, like content words, readily accepting novel tokens in the X and Y positions, whereas they considered frequent items as belonging to closed classes, like functors, accepting no novel items, and used only the frequent words to compute word order, suggesting they are familiar with the grammatical function of functors (Marino et al., 2020). This suggests that infants treat the categories established on the basis of frequency to behave differently with respect to the extensibility of their classes, similarly to functors and content words in natural language.

Infants also expect frequent and infrequent words to have different phonologies, as functors and content words do. In natural language, prosodic and word frequency information are aligned at the phrasal level. In OV languages, the prosodic prominence is marked by increased pitch or intensity, which falls on the phrase-initial content word, whereas its phrase-final functor is non-prominent. In VO languages, by contrast, prosodic prominence, typically signalled by increased duration, falls on the final constituent of the phrase, which is typically the content word (Nespor et al., 2008). Infants are sensitive to these prosodic cues and integrate them with word frequency coherently. Using a similar artificial grammar as in Gervain et al. (2008), Gervain and Bernard (2012) showed that French monolinguals preferred frequent-initial test items (mimicking the functor-initial order of French) only when prosodic and word frequency cues were aligned, i.e. when infrequent words were prominent and frequent words non-prominent, but they exhibited no preference when the two cues were misaligned, i.e. when frequent words were prosodically prominent (lengthened). This suggests that already at 8 months, infants process these two cues simultaneously, expecting word frequency and prosodic cues to be aligned and this expectation follows the characteristic phonologies of content words, which are heavy, and functors, which are reduced, in natural language.

Such sensitivity to the prosodic properties of phrases might be particularly useful for bilingual infants exposed to a functor-initial and a functor-final language to keep the grammars of the two languages apart, as they can rely on the physical realization of prosodic prominence as a cue. In this regard, Gervain & Werker (2013) tested 7-month-old bilinguals exposed to various functor-final languages, in which prosodic prominence at the phrasal level is realized as increased pitch or intensity, and English, a functor-initial language in which phrasal level prominence is carried by increased duration, i.e. lengthening. Infants were exposed to an artificial grammar similar to that of Gervain et al. (2008), but with prosody overlaid on the stream. One group of bilinguals was exposed to a functor-final prosody, i.e. X & Y infrequent words higher in pitch, a second group was exposed to functor-final prosody, i.e. X & Y infrequent words were longer in duration. In both conditions, bilinguals exhibited the predicted preference for the word order that corresponded to their prosody, suggesting that they rely on prosody, in addition to word frequency, as a cue to word order.

By 17 months, infants also show evidence of being familiar with the function of content words, i.e. that they carry semantic meaning. When an artificial grammar similar to Gervain et al. (2008) was followed by a subsequent word-learning phase, infants only associated infrequent, but not frequent words with pictures of objects, suggesting they only treated the former as potential labels (Hochmann et al., 2010).

In sum, even before acquiring a substantial lexicon, infants establish two categories of words of the basis of frequency, which show all the characteristics of functors and content words, and they build a rudimentary representation of the relative order of these two categories in the language(s) they are acquiring.

1.2 The current study

The artificial grammars that were used in the studies mentioned above (Gervain et al., 2008; Bernard & Gervain, 2012; Gervan & Werker, 2013; Marino et al., 2020), shared a common basic structure. Specifically, frequent words and infrequent words were presented with a 1:9 frequency ratio. In natural languages, most functors indeed have a much higher token frequency than most content words (Kucera & Francis, 1967; Gervain et al., 2008). For instance, the 20-30 most frequent words in corpora of infant-directed speech across different languages, e.g. Italian, Hungarian, French, Japanese etc., are all functors (Gervain et al., 2008). Indeed, word frequencies in natural languages follow a Zipfian or power law (also known as 1/f) distribution, with a few words having very high frequencies – these are typically functors –, and most other words having mid to low frequency (Zipf, 1935; Ferrer i Cancho & Sole, 2003; Gervain et al., 2008). In corpora of infant-directed speech in French (Morgenstern and Parisse, 2007; Leroy et al., 2009; Morgenstern and Sekali, 2009), the native language of our infant participants, we found (Gervain et al., 2013) that the frequency ratios between individual functors and content words ranged roughly between 1:7 and 1:3000, themselves following a Zipfian distribution as a consequence of the Zipfian distribution of word frequencies. The 1:9 ratio is an approximation of this. However, exactly what ratio is necessary to trigger different categorization in infants has never been explored.

Yet, experimental evidence in other artificial grammar learning tasks (Gomez, 2000; Endress & Bonatti, 2007; Radulescu, Wijnen, Avrutin; 2020; Valian & Coulson, 1988) suggests that input variability, complexity and frequency ratios play a role in selectively triggering different processing mechanisms. It is, therefore, important to understand under what input conditions the extraction of word order regularities take place. This is the question we address in the current study.

In a study on the learning of nonadjacent dependencies a_b in structures of the type aXb, Gomez (2002) showed that for both adults and infants, the size of the set from which the middle elements (X) were drawn affected the leaning. When participants were exposed to small set sizes for X, they showed decreased learning of the non-adjacent dependency, while performance was high for large sets of X. This suggests that materials in which some elements have high variability facilitate the extraction of the invariant patterns. Similarly, adult participants exposed to artificial grammars that differ in the number of markers (functors) and content words showed that the grammars with more frequent markers are easier to learn than grammars in which markers are less frequent (Valian & Coulson, 1988).

Similar patterns are observed during the acquisition of the native language. Infants seem to be sensitive to functors with the highest frequency of occurrence first. They start to segment functors from the speech stream by 6 to 8 months of age (Höhle and Weissenborn, 2003; Shi, Marquis & Gauthier, 2006), and show a preference for high frequency functors (Shi, Cutler, Werker, & Cruickshank, 2006). By 8-11 months, they use highly frequent, but not less frequent functors to segment adjacent content words (Shi, Cutler et al. 2006). This suggests that by being more perceptually salient, high frequency functors might facilitate the establishment of linguistic regularities and the learning of the relation between lexical items.

Another important factor that has been shown to impact learning in artificial grammar studies is the overall amount of exposure to the input. Shorter exposure typically leads to the extraction of structural regularities and their generalization (Endress & Bonatti, 2006; Radulescu, Wijnen, Avrutin, 2020), while longer exposure favors item-based learning and memorization. A possible explanation for these results is that short exposure doesn't allow sufficient time/opportunity for the memorization of individual items. Under these conditions the best strategy to capture the largest amount of information from the input is the extraction of invariant structural regularities. By contrast, long exposure offers the necessary time window to memorize individual items or other specific features of the input. A similar explanation, sometimes referred to as the "Less is More Hypothesis" (Goldowsky & Newport, 1993), has also been proposed to account for the difference between infants' spontaneous, effortless, implicit and fast acquisition of their first language(s) and adults' slow, effortful and conscious learning of a second language. Infants, having limited memory capacity, cannot store enough information to learn about individual items in the input and thus extract the broadest possible generalizations possible to capture the input, while adults have sufficient attentional, memory and cognitive resources to learn specific, item-based information, and are thus less likely to extract rules (Kam & Newport, 2009). Thus, whether limited by the quantity of exposure available or by memory capacity, the ability to process reduced amounts of information leads to the extraction of regularities and invariances that can capture most of the input.

In the light of the above, in this work we have investigated whether infants are able to categorize frequent words as functors and use them to bootstrap their native word order if the frequency ratio between frequent and infrequent words is higher than what is typically found in natural languages (1:3). This reduces the complexity/variability of the input. We thus expect that under these conditions (Experiment 1) infants may no longer show a word order preference. During the 4-minute familiarization, using the 1:3 ratio, the 9 possible combinations of X and Y are repeated 27 times each (as compared to only 3 repetitions with the 1:9 ratio) and the lexicon of the artificial grammar only consists of 2 frequent and 6 infrequent words (as compared to 2 frequent and 18 infrequent words with the 1:9 ratio), providing overexposure, i.e. more opportunity for item-based learning and hindering the extraction of structural regularities, i.e. the computation of word order. Using Shannon entropy (H) as a formal measure of complexity, in the original 1:9 ratio grammars, the probability p of any AXBY unit (x) is p=1/81=0.0123 and $H = -\sum p(\mathrm{x}) \log_2 p(\mathrm{x}) = 6.34$ bits. By contrast, in the 1:3 grammar, p=1/9=0.111 and $H = -\sum p(\mathrm{x}) \log_2 p(\mathrm{x}) = 3.17$ bits. Entropy is thus twice as high in the original 1:9 ratio grammars than in the current 1:3 ratio one. Variations between artificial grammars corresponding to entropy values of 0.5-1 bits have been documented to trigger different performance in adults (Radulescu et al. 2020) and infants (Gerken 2006). We thus expect the twofold reduction in entropy of more than 3 bits from the 1:9 to the 1:3 ratio grammars to reduce infants' preferences.

This is all the more likely since in the test phase, infants are faced with test sequences that all appeared in the stream. Both frequent-initial and frequent-final test sequences actually occur in the familiarization stream and are, therefore, familiar to infants. Any preference for one of the two patterns derives from how infants parse and represent the stream, not from the familiar vs. novel nature of the test items.

Given this prediction for infants not to extract the native word order pattern with a frequency ratio of 1:3 with the original 4-minute long familiarization, we performed two further experiments to investigate whether a word order preference can nevertheless be restored if the input characteristics are modified (while maintaining the same entropy, as the underlying structure of the grammar and the 1:3 ratio and thus the probability of occurrence of the AXBY units to not change). In Experiment 2, we thus maintained the 1:3 frequency ratio, but we added native-like prosody, i.e. lengthening of the infrequent word, which has been shown to support word order extraction in a similar paradigm (Bernard & Gervain, 2012). If prosody as a cue to word order can counter the effects of overexposure and low complexity, then infants in this condition may recover their word order preference.

In Experiments 1 and 2, the familiarization is kept the same length, 4 minutes, as in Gervain et al. (2008). This has two consequences: (i) it reduces the variability/complexity of the input (the lexicon of the grammar) and (ii) leads to more repetitions, i.e. more exposure to individual items. To disentangle whether the reduced set size of the lexicon or the amount of exposure (length) influences infants' ability to extract word order, in Experiment 3, we decreased the length of the familiarization by repeating each of the 9 possible AXBY combinations only 3 times, i.e. the number of repetitions per possible AXBY combination was the same as in Gervain et al. (2008). This yielded to a considerably reduced 27-second-long familiarization. With the reduction of the amount of input triggering structural generalizations, we now predicted that infants would recover their word order preference.

Importantly, here we test French-exposed infants. This population has been found to show a frequent-initial word order preference in a previous study (Marino et al., 2020) using exactly the same paradigm as here, similarly to our previous work with infants exposed to other functor-initial languages (e.g. Gervain et al., 2008; Gervain & Werker, 2013). We thus did not need to run a group of infants exposed to the 1:9 grammar.

2. Experiment 1

In this experiment, we tested whether the three-fold reduction of the set size of infrequent items, resulting in a reduction of the complexity of the lexicon and an increase in the redundancy of the input still allows infants to extract the frequent-initial word order characteristic of their native language, French, as is the case for a higher frequency ratio between frequent and infrequent items (Gervain et al., 2008).

2.1 Methods

2.1.1 Participants

Based on Gervain et al. (2008) and Marino et al. (2020), using similar artificial grammar learning paradigms as the current one, a power calculation to estimate the sample size was performed. In the two previous studies, an effect size (Cohen's *d*) of $d = .524$ was obtained. By applying this effect size and a power of 0.7, a sample size calculation performed using G*Power [Faul et al., 2009] resulted in a desired sample size of 19. Therefore, we aimed for a final sample size (after rejection) of at least 19 infants in this and all subsequent experiments.

Twenty-seven (12 girls) 8-month-old (mean age 8 months and 8 days, range 7.5- 9 months) French infants took part in Experiment 1. Among these 27 infants, 6 were not included in the final data analysis for fussiness and crying. A final sample of 21 participants was entered into the analysis.

Parents of all participating infants gave written informed consent prior to participation. All experiments were approved by the ethics boards of the institutions involved (CERES of the Université Paris Descartes) CER-Paris Descartes, approval nr 2016/32.

2.1.2 Materials

The artificial grammar was created based on the one used in Gervain et al. (2008). The basic grammatical structure consisted of a four-syllable-long sequence (AXBY), where each unit was realized as a consonant-vowel (CV) syllable. In this structure, A and B units are frequent (mimicking function words), whereas X and Y are infrequent (mimicking content words). This is because the A and B categories contain one token each (A: *fi*; B: *ge*), while the X and Y categories have three tokens (X: fo, ru, de; Y: bi, mu, do), making the individual X and Y tokens three times less frequent than the A and B tokens. Importantly, the CV units respected French phonotactics, but were all non-words in the French infant vocabulary.

The familiarization stream was synthesized using a text-to-speech synthesis software MBROLA (Dutoit, 1996). The familiarization consisted of a 3-minute 53-second long speech stream where frequent and infrequent words were concatenated without pauses. The AXBY structure was repeated 243 times, and it was synthesized with a constant pitch of 200Hz (corresponding to the fundamental frequency of female voices) and a phoneme duration of 120 ms. This stream thus provided no prosodic cues.

The initial and final 15 s of the familiarization stream were ramped in amplitude, masking information about the exact beginning and end of the stream. As a result, the structure of the streams was ambiguous between a frequent word-initial and a frequent word-final parse (e.g. …*ge*bi*fi*de-*ge*mu*fi*ru-*ge*do*fi*de-*ge*mu*fi*fo… or …*ge*-bi*fi*de*ge*-mu*fi*ru*ge*-do*fi*de*ge*-mu*fi*fo…). During the test phases, 8 items, 4 frequent-initial and 4 frequent-final, all flat in prosody, were presented (Figure 1).

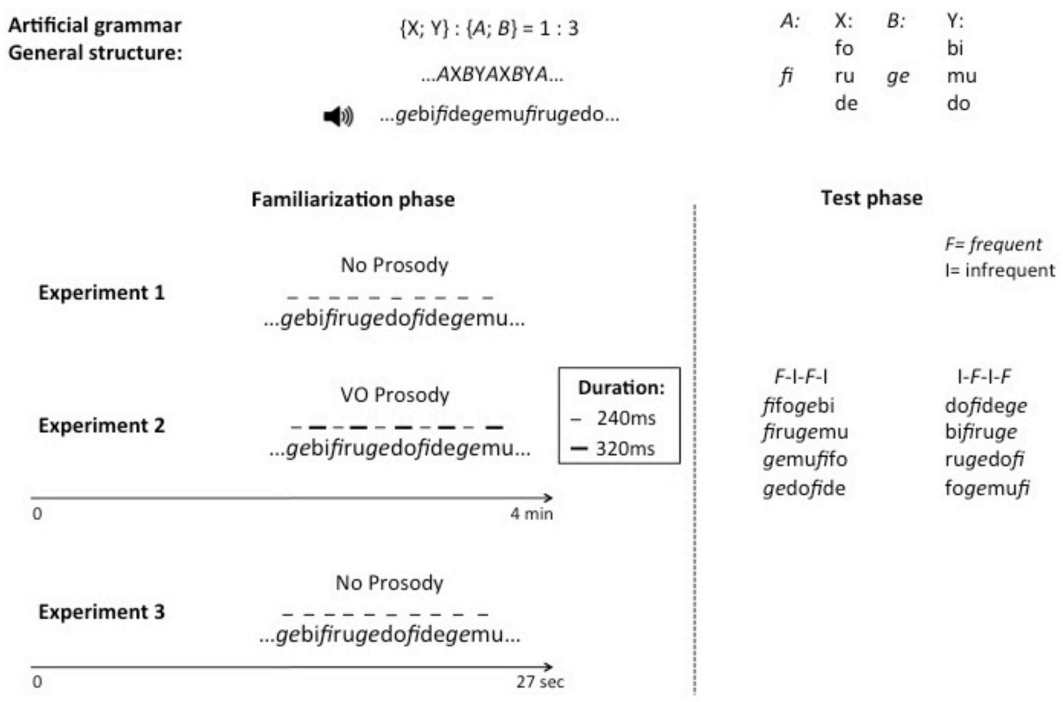

Figure 1: Artificial grammar tasks used in Experiment 1, 2 & 3

2.1.3 Procedure

Infants were tested individually using the Headturn Preference Paradigm (HPP) (Nelson, Jusczyk et al., 1995; Saffran, Johnson et al., 1999), measuring infants' looking behavior to assess preferences between different types of auditory stimuli (Figure 2). Infants were seated on a caregiver's lap, sitting on a chair in the middle of a quiet testing booth, with three side screens. On the screens, videos of looming circles imitating blinking lights were played (one on each side) as attention getters. Below each side screen, loudspeakers were placed to present the sound stimuli. The caregiver listened to masking music to avoid influencing the infant's response. Each experimental session included a familiarization phase (with one of the three streams) and a test phase. During the familiarization phase, a continuous familiarization stream was presented independently of infants' looking behaviour. Simultaneously, The visual attention getters, unlike the sounds, were presented contingently upon infants' looking behavior (see below). At the end of the familiarization phase, infants immediately went on to the test phase. During the test phase, both the sound and the visual stimulus were contingent upon infants' looking behavior. Before each test trial, the central attention getter was presented on the front screen. Once infants reliably fixated on it, the central attention getter was turned off and one of the side attention getters was turned on (sides were randomized and counterbalanced within and across infants). A sound stimulus started to play from the loudspeaker placed below the corresponding side screen once the infant reliably fixated on the blinking side screen, as indicated by a head turn of at least 30^{o} to that side. When the infants looked away for more than a predefined look away criterion (2 s) or until the end of the sound file (21 s), the test trial ended and a new trial was then presented. Each participant heard eight test items: four in each condition (*F*-I-*F*-I or I-*F*-I-*F*). In a test trial, a given test item was repeated identically 15 times with a silence interval of 500msec between repetitions resulting in a maximum trial duration of approximately 21sec. Stimuli were pseudo-randomized for each participant and the order of presentation was counterbalanced between participants. During the study, a blinded experimenter observed infants' looking behavior and controlled the stimulus presentation software (PsyScope version X B55 run on a Mac OS X, version 10.10.5). Experimental sessions were recorded, and the videos were coded offline by a blind coder to measure infants' looking times. Additionally, a second blind coder analyzed a set of randomly selected videos, representing 10% of all videos. The correlation between the two coders was r = .88. After the offline coding, looking times across all trials of the same condition were averaged for each participant.

Figure 2: Experimental box and set up

2.2 Results

The average looking times to *F*-I-*F*-I and I-*F*-I-*F* items are shown in Figure 3. A paired samples t-tests (with equal variance not assumed) showed no preference for either of the two word orders (*F*-I-*F*-I: looking times: M = 5.92s; SD = 2.63; I-*F*-I-*F*: looking times: M 6.03= s; SD = 2.77; t (20) = 0.150; p **=** 0.881**;** d = 0.25; power (1-β) = 0.19. To estimate the degree of confidence in this null finding, we also calculated the Bayes factor for this comparison. BF_{10}=0.23, indicating that the null hypothesis is more likely than the alternative hypothesis, as this value is below the 0.33 threshold conventionally associated with substantial support for the null hypothesis (Lee & Wagenmakers, 2014).

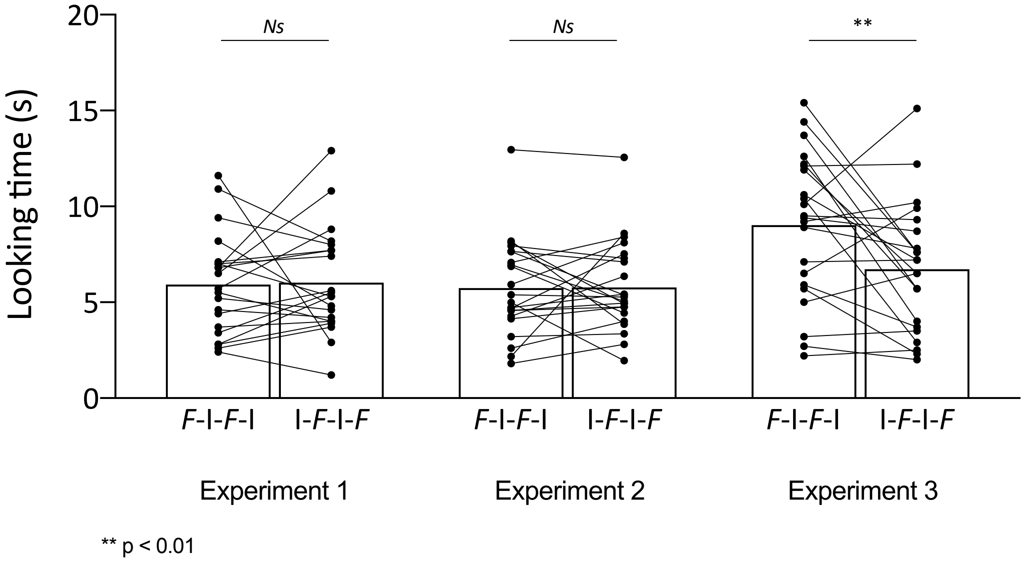

Figure 3: Results of Experiments 1, 2 & 3. Looking times for Experiments 1, 2 & 3. The x-axis shows the different experimental groups. The y-axis shows looking time (s). Bars represent group means, connected dots represent individual participants' looking times in the two experimental conditions.

2.3 Discussion

The results revealed no word order preference, suggesting, as predicted, that lower complexity input and highly redundant exposure hinder rule extraction. A frequency ratio of 1:3, resulting in a lexicon of 8 words in total and 27 repetitions of the basic structure do not allow infants to parse the familiarization stream according to the native word order. While this study doesn't establish an absolute threshold for the frequency ratio (and/or lexicon set size and/or amount of exposure) necessary to trigger word order extraction, arguably the complexity of the input in the current experiment is sufficiently low for 8-month-olds to learn about at least some of the individual items in the familiarization stream. This then allows them to recognize both frequent-initial and frequent-final test items as familiar, as indeed, in the current task, there is no difference in novelty between the two test item types, unlike in many other artificial grammar tasks (e.g. Marcus et al., 1999).

3. Experiment 2

Can the processing mechanisms triggered by low complexity be overridden by cues that we know infants rely on to establish word order? In Experiment 2, we tested one such cue, prosody (Bernard & Gervain, 2012; Gervain & Werker, 2013). We overlaid a duration contrast based prosodic pattern typical of French on the familiarization stream to address this question.

3.1 Methods

3.1.1 Participants

Twenty-nine (13 girls) 8-month-old (mean age 8 months and 12 days, range 7.5- 9 months) French infants took part in Experiment 2. Among these 29 infants, 7 were not included in the final data analysis for fussiness and crying. A final sample of 22 participants was entered into the analysis.

Parents of all participating infants gave written informed consent prior to participation. All experiments were approved by the ethics boards of the institutions involved (CERES of the Université Paris Descartes) CER-Paris Descartes, approval nr 2016/32.

3.1.2 Materials

The familiarization was similar to the one in Experiment 1, but native prosody (duration-based prosody typical of functor-initial languages) was added following Bernard and Gervain (2012). A constant pitch of 200Hz (corresponding to the fundamental frequency of female voices) was used for all words (both frequent and infrequent). The durational contrast marking prosodic prominence was then created by making the prominent infrequent words longer (320 ms) than the non-prominent frequent words (240 ms). This resulted in a familiarization with a total duration of 4 minutes and 32 seconds.

3.1.3 Procedure

The procedure was identical to Experiment 1, except for the duration of the familiarization stream.

3.2 Results

The average looking times to *F*-I-*F*-I and I-*F*-I-*F* items are shown in Figure 3. A paired samples t-tests (with equal variance not assumed) showed no preference for either of the two word orders (*F*-I-*F*-I: looking times: M = 5.75s; SD = 2.54; I-*F*-I-*F*: looking times: M = 5.77s; SD = 2.35; t (21) = 0.04; p = 0.98; d = 0.37; power (1-β) = 0.38). The Bayes factor for this comparison was BF_{10}=0.22, indicating substantial support for the null hypothesis (Lee & Wagenmakers, 2014).

3.3 Discussion

Like in Experiment 1, infants showed no word order preference despite the addition of the native-like prosodic cue to the familiarization stream. Prosody may be a useful cue to word order when infants are applying processing mechanisms that aim to extract structural regularities (Bernard & Gervain, 2012; Gervain & Werker, 2013), but seem not to be relevant or strong enough once more item-based learning takes place.

4. Experiment 3

In both previous experiments, the change in set size from 9 to 3 for the infrequent categories resulted in a simultaneous reduction in complexity and an increase in redundancy. As predicted, this prevented infants' from projecting their native word order onto the stream. But the two changes can be dissociated, allowing us to tease apart whether it is the reduced complexity/variability or the overexposure that is responsible for this. In Experiment 3, we therefore reduced the number of repetitions of the AXBY basic unit to the same level as in the original Gervain et al. (2008), i.e. 3 repetitions, resulting in a shorter familiarization and less exposure. We reasoned that this might re-establish word order preferences as infants can no longer memorize items sufficiently and will instead extract regularities.

4.1 Methods

4.1.1 Participants

Twenty-six (14 girls) 8-month-old (mean age 8 months and 11 days, (range 7.5-9 months) French infants took part in Experiment 3. Among these 26 infants, 4 were not included in the final data analysis for fussiness and crying. A final sample of 22 participants was entered in the analysis.

Parents of all participating infants gave written informed consent prior to participation. All experiments were approved by the ethics boards of the institutions involved (CERES of the Université Paris Descartes) CER-Paris Descartes, approval nr 2016/32.

4.1.2 Materials

The general structure of the familiarization was identical to Experiment 1, except that we shortened the familiarization by decreasing the number of repetitions. Each of the nine possible AXBY combinations was repeated 3 times, for a total duration of 27 seconds.

4.1.3 Procedure

The procedure was identical to Experiment 1, except for the duration of the familiarization stream.

4.2 Results

The average looking times to *F*-I-*F*-I and I-*F*-I-*F* items are shown in Figure 3.Infants in Experiment 3 showed the predicted frequent word initial preference, as evidenced by a paired sample t-test (*F*-I-*F*-I: looking times: M = 9.04 s; SD = 3.82; I-*F*-I-*F*: looking times: M = 6.73s; SD = 3.36; t (21) = 2.96; $p = 0.0075$; d = 0.64; power (1-β) = 0.82). The Bayes factor for this comparison was BF_{10}=4.18, providing moderate support for the alternative hypothesis (Lee & Wagenmakers, 2014).

To directly test whether infants behaved differently in Experiments 1, 2 and 3, we conducted a repeated-measures ANOVA with Experiment (1/2/3) as a between-subjects factor and Test Item Type (*F*-I-*F*-I/I-*F*-I-*F*) as a within-subjects factor. We obtained a significant main effect of Experiment [F (2,62) = 4.79, p = 0.012] due to greater overall looking times in Experiment 3 than in the two other experiments (Sheffe's post hoc tests: Experiments 2 vs. 3 p = 0.024; Experiments 1 vs. 3 p = 0.051; Experiments 1 vs. 2 p = 0.96). We also found a significant Experiment × Test Item Type interaction [F (1,62) = 4.003, p = 0.023] due to significantly longer looking times to frequent-initial (*F*-I-*F*-I) than frequent-final (I-*F*-I-*F*) items in Experiment 3 (Scheffe's post hoc p = 0.005), but not in Experiment 1 and 2 (ns.). Furthermore, looking times to frequent-initial items were lower in Experiments 1 and 2 than in Experiment 3 (Scheffe's post hocs, respectively, p = 0.0001 and p < 0.0001). In a Bayesian repeated measures ANOVA, the Bayes factor associated with the model including Experiment (1/2/3) as a between-subjects factor and Test Item Type (*F*-I-*F*-I/I-*F*-I-*F*) as a within-subjects factor as compared to the null model is BF_{10}=8.71, indicating considerable support for this model. This is the model with the highest BF. The inclusion probabilities for the factor Experiment and the Experiment x Test Item Type interaction are BF_{incl}=6.18 and BF_{incl}=3.62, respectively, providing considerable support for these factors, in accordance with their significance in the classical ANOVA. For the factor Test Item Type, BF_{incl}=1.58, providing weak support in accordance with the lack of significance for this factors in the classical ANOVA.

4.3 Discussion

As predicted, reduced exposure in Experiment 3 restored infants' word order preference. This suggests that the null preference found in Experiment 1 is not linked to the low frequency ratio per se, but rather to the large overall exposure to the artificial grammar.

In the comparison of the three experiments, we also observed the predicted difference. Only in Experiment 3 did we find greater looking times to frequent-initial than to frequent-final items. We also observed greater overall looking times in Experiment 3 than in the other two experiments, most likely as a consequence of the much shorter familiarization time in Experiment 3, leaving infants less tired and more attentive during test.

5. General Discussion

In this work, we have investigated whether infants are sensitive to the magnitude of the frequency difference between words to categorize them as functors and content words and extract their relative order (Experiment 1) and whether adding prosody (Experiment 2) and modulate the amount of exposure (Experiment 3) shapes this ability. We have found that a low frequency ratio is not in itself a hindrance to the frequency-based bootstrapping of word order, as long as exposure is sufficiently limited. By contrast, prosody cannot counter the effects of overexposure.

Specifically, in Experiment 1 and 2, the amount of exposure was kept the same as in previous studies (e.g. Gervain et al., 2008), but the decrease in the set size of the lexicon from 20 to 8 elements resulted in less variable, less complex and more redundant input. This might have overexposed infants to the material to a point where each word was perceived as frequent or familiar, resulting in no preference at test. Indeed, once redundancy was reduced in Experiment 3 by shortening the familiarization, recognition at test was no longer possible and the extraction of invariant structure took place instead. We interpret these findings in the framework of the Less is More Hypothesis (Newport 1990) according to which immature working memory and processing abilities force infants to attend to general regularities rather than item-based or specific information in the input.

These results are in line with previous studies suggesting that infants at this age only use the most frequent functors in their native language to segment out content words. Our results also mesh well with evidence in the literature showing that the length of familiarization and the set size of the categories used in artificial grammars affect participants' learning performance (Endress & Bonatti, 2006; Gomez, 2002; Radulescu et al., 2019). Whether the two different learning outcomes triggered by short and long exposure, i.e. rule extraction and item-based rote learning, are the products of two different mechanisms (Endress & Bonatti, 2006) or a single mechanism that produces different behaviors as a function of the characteristics of the input (Radulescu et al., 2019) remains debated.

Our findings also have implications for the architecture and learnability of natural languages. Input to young infants is typically not limited. Although there is important cultural variation in the amount of speech heard by an infant, the majority of babies receive large amounts of speech input, and getting larger quantities typically leads to better language development outcomes (Soderstrom, 2007). In the light of this, our results imply that for infants to be able to use functors to extract grammatical structure, at least some functors need to be of high frequency so that the invariant structural patterns they signal pop out from the high variability of the speech signal despite considerable exposure.

Importantly, unlike in many other artificial grammar learning studies, our task does not measure infants' learning of the artificial language. All test items are taken from the familiarization stream, so all are familiar and "grammatical". There are no "violation" items like in many other studies. What our experiments gauge instead is the knowledge infants bring to the task on the basis of the large quantities of exposure they have already received of their native language. Our task is thus a prism that offers insight into existing native language biases, providing a crucial link between our laboratory task and infants' natural experience with their native language. The "Less is More Hypothesis" readily explains under what conditions infants are able to use these native language biases when facing novel linguistic input, pointing towards high entropy and short exposure time as key factors.

6. Conclusion

Overall, this study was one of the first to explore whether and how the frequency ratio of frequent and infrequent words impacts infants' ability to recognize functors and content words and their relative word order. Our results are in line with previous studies showing that pre-lexical infants, as young as 8 months, are able to use word frequency on their way to the acquisition of grammar, even under strongly reduced input variability conditions. These results fit well with *bootstrapping theories of language acquisition* (e.g. Morgan & Demuth, 1996), arguing that learners are able to extract abstract, structural, and hence directly unobservable properties of the target language, from perceptually available cues present in the input that correlate with the underlying structure. Since at 8 months, infants do not have a sizable lexicon yet, this knowledge is general and not linked to specific lexical items.

Acknowledgements

This work was supported by the ERC Consolidator Grant "BabyRhythm" 773202.

Author Contributions

CM and JG developed the study concept. All authors contributed to the study design. Testing, data collection, analysis and interpretation were performed by CM under the supervision of JG. CM drafted the manuscript and JG provided critical revisions. All authors approved the final version of the manuscript for submission.

References

Abney, S. (1987). The English Noun Phrase in its Sentential Aspect. Unpublished PhD, MIT, Cambridge, MA.

Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: what level of representation?. *Frontiers in Psychology*, *3*, 451.

Braine, M. D. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology, 72 (4)*, 532-540. http://dx.doi.org/10.1037/h0023763

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33-B44.

Chomsky, N. (1995). The Minimalist Program. Cambridge: MIT Press.

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, 81-138.

Dutoit, T. (1996). *An Introduction to Text-to-Speech Synthesis,* Boston: Kluwer Academic Publishers. *Springer Netherlands; 4. 10.1007/978-94-011-5730-8*

Endress, A. D., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*(2), 247-299.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149-1160.

Fukui, N. (1986). A Theory of Category Projection and its Applications. Unpublished PhD, MIT, Cambridge, MA.

Gervain, J., & Werker, J. F. (2013). Prosody cues word order in 7-month-old bilingual infants. *Nature communications*, *4*(1), 1-6.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive psychology*, *57*(1), 56-74.

Goldowsky, B. N., & Newport, E. L. (1993). Limitations on the Acquisition of Morphology: The Less Is More Hypothesis. In *The proceedings of the twenty-fourth annual child language research forum* (p. 124). Center for the Study of Language (CSLI).

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431-436.

Green, T. R. G. (1979) The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, *18*, 481- 496. https://doi.org/10.1016/S0022-5371(79)90264-0

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.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, *6*(2), 122–127.

Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*(1), 30-66.

Kucera, H., Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown university press.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.

Marino, C., Bernard, C., & Gervain, J. (2020). Word Frequency Is a Cue to Lexical Category for 8-Month-Old Infants. *Current Biology*.

Morgan, J. L. & Demuth, K. (1996). Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 263–283.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, 263-283.

Morgan, J. M. & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, 20 (1): 67-85

Nelson, D. G., Jusczyk, P., Mandel, D., Myers, J., Turk, A., Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development*, *18*, 111- 116. 10.1016/0163-6383(95)90012-8

Nespor, M. & Vogel, I. (1986). Prosodic Phonology, Vol. 28 Dordrecht: Foris, 1st edition.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H. & Donati, C. (2008). Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio* 7, 1–28.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Radulescu, S., Wijnen, F., & Avrutin, S. (2020). Patterns Bit by Bit. An Entropy Model for Rule Induction. *Language Learning and Development*, *16*(2), 109-140.

Saffran, J. R., Johnson, E. K., Aslin, R. N., Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52. https://doi.org/10.1016/S0010-0277(98)00075-4

Selkirk, E. (1984) Phonology and syntax: The relation between sound and structure. Cambridge, MA: The MIT Press.

Shi, R., Cutler, A., Werker, J., Cruickshank, M. (2006a). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of Acoustic Society of America,* 11, 9(6).

Shi, R., Marquis, A., Gauthier, B., Bamman, D., Magnitskaia, T., & Zaller, C. (2006). Segmentation and representation of function words in preverbal French-learning infants. *Proceedings of the 30th Annual Boston University Conference on Language Development*, *2*, 549–560.

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

Shi, R., Werker, J., & Cutler, A. (2006b). Recognition and representation of function words in English-learning infants. *Infancy*, 10, 187–198.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501-532.

Valian, V., Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *Journal of Memory and Language*, *27*(1), 71-86.