

REAL-TIME COMPREHENSION OF MULTI-WORD PHRASES

Do Children Use Multi-Word Information in Real-Time Sentence Comprehension?

Abstract

Meaning in language emerges from multiple words, and children are sensitive to multi-word frequency from infancy. While children successfully use cues from single words to generate linguistic predictions, it is less clear whether and how they use multi-word sequences to guide real-time language processing and whether they form predictions on the basis of multi-word information, or pairwise associations. We address these questions in two visual-world eye-tracking experiments with 5- to 8-year-old children. In Experiment 1, we asked whether children generate more robust predictions for the sentence-final object of highly frequent sequences (e.g. *throw the ball*) compared to less frequent sequences (e.g. *throw the book*). We further examined if gaze patterns reflect event knowledge or phrasal frequency by comparing processing of phrases that have the same event structure but differ in multi-word content (e.g., *Brush your teeth* vs. *Brush her teeth*). In the second study, we employed a training paradigm to ask if children are capable of generating predictions from novel multi-word associations, while controlling for overall frequency of the sequences. While the results of Experiment 1 suggested that children primarily relied on event associations to generate real-time predictions, those of Experiment 2 showed that the same children were able to use recurring novel multi-word sequences to generate real-time linguistic predictions. Together, these findings suggest that children can draw on multi-word information to generate linguistic predictions, in a context-dependent fashion, and highlight the need to account for the influence of multi-word sequences in models of language processing.

Keywords: language processing; multi-word units; prediction; visual world paradigm; eye-tracking

1. Introduction

Language learners are faced with numerous challenges during spoken language processing: multiple words unfold rapidly, in a transient signal, while meaning is built incrementally across an utterance. Predictive processing is one cognitive mechanism that is posited to help learners succeed at real-time language processing (Huettig, 2015; Pickering & Gambi, 2018; Pickering & Garrod, 2013). Put simply, prediction during language processing facilitates comprehension by allowing the listener to pre-activate likely linguistic content based on what has already been heard. Listeners use a variety of linguistic cues to support prediction in real-time (see Huettig, Rommers, & Meyer, 2011 for a review), and predictive processing skills emerge relatively early in development: children from at least age two can use semantic information from spoken words to restrict the domain of upcoming reference (Gambi, Gorrie, Pickering, & Rabagliati, 2018; Gambi, Jindal, Sharpe, Pickering, & Rabagliati, 2020; Mani & Huettig, 2012), and from 20 months of age, they can exploit prosodic information to unpack the syntactic structure of utterances and predict the upcoming noun/verb (de Carvalho, Dautriche, & Christophe, 2016; de Carvalho, Dautriche, Lin, & Christophe, 2017). Predictive processing is also thought to support language learning processes, with performance on a variety of verbal and nonverbal predictive processing tasks associating with current and future vocabulary skills (Ellis, Gonzalez, & Deák, 2014; Gambi et al., 2020; Reuter, Emberson, Romberg, & Lew-Williams, 2018). Prediction, therefore, seems to play an important role in facilitating language processing and learning from early in life; hence, understanding the drivers of these predictive processing skills is necessary to support a full account of language learning and growth. The current project seeks to expand this body of work by exploring whether and how children use their knowledge of multi-word sequences to engage in predictive processing.

Language processing relies on the incremental emergence of meaning as words unfold across a sentence. Crucially, sentence meaning is not based on individual words, but on the combination of multiple words, and the information gleaned from such combinations. Studies have shown that adults and children are sensitive to information in multi-word sequences during linguistic production and comprehension. For instance, adults are faster to process higher (vs. lower) frequency multi-word phrases (e.g., Arnon & Snider, 2010; Tremblay, Derwing, Libben, & Westbury, 2011), even in their second language (Hernández, Costa, & Arnon, 2016; Siyanova-Chanturia, Conklin, & van Heuven, 2011), and are impacted by multi-word frequency in judging phrase complexity and inferring meaning (Reali & Christiansen, 2007). Similarly, young children show higher production accuracy in more frequent phrases (Arnon & Clark, 2011; Bannard & Matthews, 2008), and school-aged children are faster to read more frequent sequences (Jiang, Jiang, & Siyanova-Chanturia, 2020). Sensitivity to multi-word frequency emerges early: 11- to 12-month-old infants prefer listening to trigrams that are frequent in infant-directed input (e.g. “clap your hands”) as opposed to less frequent trigrams (e.g. "take your hands"; Skarabela, Ota, O’Connor, & Arnon, 2021). Beyond this early sensitivity to the *frequency* of multi-word sequences, studies also showed that toddlers are sensitive to the *prosodic features* of such sequences: they utilize them for inferring the meaning of multi-word sequences and for predicting upcoming linguistic content (de Carvalho et al., 2016, 2017). Together, these findings highlight the significance of multi-word sequences for language processing and their role as building blocks in language (Arnon & Christiansen, 2017; Arnon & Cohen Priva, 2013; Arnon, McCauley, & Christiansen, 2017; Arnon & Snider, 2010). Importantly, they suggest that children are sensitive to multi-word information from a young age. Yet it is not clear whether (and when) children use

multi-word information (known and novel) to generate predictions during real-time language processing – a question that we address in this paper.

In fact, much of the extant evidence for prediction during language processing in children is based on studies that explore pairwise associations among words, such as when children generate a prediction for an upcoming noun on the basis of an earlier lexical or morpho-syntactic cue (Fernald, Thorpe, & Marchman, 2010; Gambi, Pickering, & Rabagliati, 2016; Huang & Snedeker, 2013; Lew-Williams & Fernald, 2007; Lukyanenko & Fisher, 2016; Zhou, Crain, & Zhan, 2014). For example, young children can use gendered articles (Lew-Williams & Fernald, 2007, 2010), selectionally-restrictive verbs (Fernald, Zangl, Portillo, & Marchman, 2008; Mani & Huettig, 2012), color adjectives (Fernald et al., 2010), and determiners (Gambi et al., 2018) to anticipate an upcoming noun. While these studies are immensely valuable in demonstrating the wide range of (individual) linguistic cues which drive predictive processing, meaning in language emerges from combinations of multiple words. However, only a few studies have measured children's ability to integrate information across multiple words to guide predictive processing (Borovsky & Creel, 2014; Borovsky, Elman, & Fernald, 2012; Gambi, Jindal, Sharpe, Pickering, & Rabagliati, 2020).

For example, Borovsky and colleagues (2012) found that 3- to 10-year-old children use an informative agent+action combination (e.g., *The pirate chases the --*) to predict the identity of a highly expected sentence-final object (SHIP). Importantly, children looked at the target object that was most related to the agent+action combination (e.g. *pirate+chases*) more than at items that were only related to the agent (TREASURE to pirate) or the action (CAT to chase). That is, children's expectations were based on incremental integration of the agent and the action. A subsequent study suggested that children aged 5-10 years could also generate real-time predictions for recently learned events which were expressed in novel combinations of known agents and

actions (Borovsky, Sweeney, Elman, & Fernald, 2014). For example, after hearing a story that included multiple novel events (e.g. a monkey that rides a bus and eats candy), in a subsequent eye-tracked sentence comprehension task, children generated more anticipatory looks towards the target BUS in the sentence context “*The monkey rides ...*” than towards other distractor images that were Action-Related (i.e. *car*), Agent-Related (i.e. *candy*), and Unrelated (i.e. *apple*) even though the “*monkey riding a bus*” association is not predictable from the child’s prior linguistic and world experience (Borovsky et al., 2014).

Hence, children seem to generate predictions towards a sentence-final target (e.g. *ship*) based on their knowledge of the associations between the elements that precede it (e.g. *pirate-chase*), but it is not clear whether the distributional properties and frequency of the entire multi-word sequence (e.g. *pirate-chase-ship*) inform these predictions (Borovsky et al., 2012). Similarly, children’s ability to predict the target in novel agent+action combinations may have been the result of computing pairwise statistics (Borovsky et al., 2014). In this latter study children were exposed to novel events that involved two agents (e.g. a monkey and a dog), who completed two actions (e.g. riding and eating) on different objects (e.g. bus, car, candy, apple). Because children were then subsequently tested on the associations that appeared in the story (for example, in one version of the learning phase, the monkey agent was paired with riding a bus, and eating an apple), it is not possible to distinguish between a prediction based on paired-lexical associations and one based on multi-word information (where the entire trigram is represented). In sum, the evidence we have to date does not allow us to ascertain whether and how children rely on the frequency of **known and newly learned** multi-word information for predictive processing. The two experiments reported here directly extend and expand previous work and address the gap in the literature by examining whether 5- to 8-year-olds generate expectations in real time processing based on multi-

word sequences or whether they rely only on pairwise associations. Importantly, our definition of multi-word sequences *does not* refer to lexicalized wholes (in the age range we are testing, children have probably already analysed sequences into individual words), or to abstract frames with slots. Instead, we use the term multi-word sequence to refer to the associations formed between multiple individual words resulting in a representation of the entire sequence alongside its parts.

In the first experiment, we asked if children rely on the frequency and familiarity of Frequent Frames (multi-word sequences that are frequently present in child-directed language input) to anticipate upcoming nouns using a visual world task. Using a design similar to that of Arnon & Clark (2011), we present children with trials in which they see a target object (e.g. *teeth*), and a distractor (e.g. *hair*) and then hear a stimulus sentence (e.g. *brush your teeth*). Participants are exposed to four experimental conditions in which we manipulate the frequency of the object at the end of the frame, changing the *frequency* of the entire frame (high, low: e.g., High frequency: *Brush your teeth* vs. Lower frequency: *Brush your hair*), and the item in the middle of the frame changing its *familiarity* or *type* (familiar, modified: e.g., Familiar: *Brush your teeth* vs. Modified: *Brush her teeth*). Our question here focuses on children's sensitivity to the associations between the three words as opposed to the event associations (represented by the verb-object association), and our main aim is to determine whether children's prediction of the target object is impacted by the frequency of the entire linguistic frame in which the object is presented. We specifically compare gaze to the target and distractor across the experimental conditions. If children rely on multi-word frequency, then we should expect to see more accurate prediction of the target object in the high frequency frames compared to the low frequency frames; children should be more accurate in generating such predictions following familiar frames as opposed to modified ones. However, if children only rely on event associations for prediction (e.g., the association between

brush and *teeth*), then we should not see any differences between the familiar and modified frames and should observe differences only between the high and low frequency frames.

In the second experiment, **run on the same children**, we employ a training paradigm to ask if children are capable of rapidly forming novel multi-word associations and using these associations in predictive processing. **This study allows us to better distinguish the effect of pairwise associations from that of multiword frequency (by carefully controlling for both in the training data) while also asking whether such associations are created after a relatively short exposure and then used in real-time processing.** Using a design similar to that of Borovsky and colleagues (2014), we expose children to stories in which we vary the frequency of the agent-action-object trigram (e.g., monkey-ride-bus) while controlling the frequency of the agent-action bigram (monkey-ride) and the action-object bigram (ride-bus), by presenting events that highlight both the trigram and bigram associations. For example, some portions of the story will present the trigram relation (monkey-riding-bus), while others will increase the frequency of the component/alternate bigrams (riding-car). Children are then tested in a sentence recognition task, in which they are asked to identify the object that goes with the sentence. Here too, we compare children's gaze to the target and distractor images as the sentence unfolds. We hypothesize that children would rely on the agent-action-object relations they learned from the stories and generate anticipatory looks towards the target of the sentence before it is mentioned. The design of this experiment allows us to isolate the effect of multi-word frequency from that of event frequency in a way that is hard to do in natural language and that can shed light on children's representation of combinatorial multi-word information.

Identifying whether and how children rely on the frequency of multi-word sequences in real-time language comprehension has several theoretical implications. First, these findings would

enrich our understanding of the kind of information that guides predictive processing in children. While some proposals framed by evidence derived from adult processing studies suggest that predictive processing involves multiple potential mechanisms (that rely on production, association, probabilistic knowledge, and more, see Pickering & Garrod, 2013), others have argued that these accounts may not translate directly to younger children with alternating evidence for and against (pairwise) associative mechanisms in prediction (Mani, Daum, & Huettig, 2016; Mani & Huettig, 2013), suggesting that children may recruit statistical information about word sequences to guide predictive processing in a context-dependent fashion. If children track the frequency of both words and larger sequences, then models of processing should account for multi-word representations across development. Second, finding that processing is guided (among other factors) by multi-word frequency would support the role of larger units as building blocks in language (Arnon & Clark, 2011; Arnon et al., 2017; Bannard & Matthews, 2008), and would be consistent with single-system models of language where all linguistic experience, whether words or larger phrases, is processed and represented using a single cognitive mechanism that compiles frequency at multiple levels of experience (e.g., Bybee, 1998; Elman, 2009; McClelland et al., 2010). Alternatively, if the outcomes of both studies suggest that children generate predictions based only on individual or pair-wise associations, this would cast doubt on the role of larger sequences in children's online processing. Either way, our results will add valuable and much-needed evidence to the work examining the cues and information used in real-time processing during development.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Thirty-seven children between the ages of 5 and 8 years old participated in the study (see details on sample size justification in the next section 2.1.2). Families were recruited from a recruitment registry at [institution masked for review], which includes a list of families with children from the surrounding community who expressed interest in participating in developmental research. Families with children within the desired age range were invited to take part in a study at the [institution masked for review] Language and Cognitive Development Laboratory. Inclusionary criteria were: (1) Native English speaking, and not hearing more than 20 hours a week at home of a language other than English, (2) Normal or corrected-to-normal vision and normal hearing, (3) Has not received treatment for any speech, language, reading or cognitive concerns. Two participants failed to meet these criteria: one parent reported a reading-related issue, and another reported a diagnosis of ADHD. One participant was not included in the analyses due to an equipment issue that led to the failure to collect eye-tracking data. A total of 34 participants were therefore included in the final sample. All child participants' parents provided written consent, and the experimental procedures were reviewed and approved by the local institutional review board at [institution masked for review] under the following IRB protocol # 2016.18919 "Development of Language Processing Skills."

2.1.2. Sample Size Justification

Our target sample size (N=32) was based on previous predictive processing studies (our recruitment strategy led us to overshoot slightly and we ended up with a slightly larger sample, N=37). To ensure that our sample size was adequately powered to detect effects in our study, we have carried out several post-hoc power analyses using the simR package in R for linear mixed effects modeling (Green & MacLeod, 2016). These analyses indeed showed that our experiments were adequately powered. The simulation approach in this package is appropriate for a variety of

experimental designs using linear mixed effects modeling, as we have used. We used this package to determine whether our sample size (N=34 participants included in the final analyses) was sufficient with at least power = 80% to detect a (medium) fixed effect size of 0.20 in both experiments. For experiment 1, this analysis indicated that we had at least 99% power to detect a main fixed effect of this size, and 96% power to detect an interaction. For experiment 2, the analysis similarly indicated 100% power to detect a (medium) fixed effect of interest area of size = 0.2.

2.1.3. Stimuli

2.1.3.1. Sentence Selection

Three word sequences of the form “verb-(pronoun/article/preposition)-object” were selected for the study using a large subset of the CHILDES database (MacWhinney, 2000). We used all the speech directed to children under the age of six who are learning American English (around six million words). We extracted all the trigrams starting with a set of early acquired and highly frequent verbs (based on the MacArthur-Bates Communicative Development Inventory; Fenson et al., 2007): *brush*, *close*, *drink*, *open*, *throw*, *want*, *wash*. For each verb, we then selected two sentence frames: one that contained a more frequent object completion (e.g. *close your eyes*) and one that contained a less frequent object completion (e.g. *close your mouth*, see Appendix A for the full list of sentence frames). All objects were highly frequent in the corpus (mean frequency of 578 per million words) and all frames were plausible. For example, “*Brush your teeth*” is more frequent than “*Brush your hair*”, but both are feasible sentences that could be spoken to young children as verified by their appearance within the CHILDES corpus. Both High and Low frequency frames were assigned to the Familiar Frame condition. We then created sentences for the Modified Frame condition by changing the intervening element (pronoun, article, or

preposition) in each of the items classified as Familiar Frames to another item that would still result in a plausible sentence/event (e.g. *Brush his hair/teeth*), but one that rarely appeared in the same child-directed corpus (mean frequency of 1.5 per million). Modified Frames were created to directly compare against the Familiar frames, as a test of whether processing in the High and Low frequency condition is driven by the multi-word sequence that precedes the target noun, or simply by the association between the object and the event denoted by the verb.

Together, this arrangement yielded four experimental conditions that varied by Frame Type (Familiar or Modified) and Frame Frequency (High or Low): (1) Familiar Frame – High Frequency (FF-High; e.g. *Close your eyes*), (2) Familiar Frame – Low Frequency (FF-Low; e.g. *Close your mouth*), (3) Modified Frame – High Frequency (MF-High; *Close his eyes*), and (4) Modified Frame – Low Frequency (MF-Low; *Close his mouth*). Sentences falling under the Familiar Frame – High Frequency and Modified Frame – High Frequency conditions had the same mean Object Frequency and Verb-Object Frequency, as seen in Table 1, but had different Frame Frequency. Similarly, sentences classified as Familiar Frame – Low Frequency and those classified as Modified Frame – Low Frequency shared the same mean Object Frequency and Verb-Object Frequency, but differed in the Frame Frequency. Importantly, as seen in Table 1 and Appendix A, the modified frames (both High and Low) have similar low frequencies, that contrast with the frequencies of the familiar frames. The goal behind creating these modified frames from the original familiar frames was to try to tease apart the effects of the verb alone on the prediction of the object as opposed to the verb+pronoun. See Appendix A for a list of all sentences in all conditions along with associated frequency counts for each sentence type.

Table 1

Means and Standard Deviations for all Frequency Information for the Stimuli in Experiment 1

Type	Frame Frequency	Object Frequency	Verb-Object Frequency
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
FF-High	22.785 (13.98)	476.57 (288.16)	34.141 (18.825)
FF-Low	3.071 (2.07)	680.356 (433.528)	12.046 (9.405)
MF-High	1.308 (1.171)	476.57 (288.16)	34.141 (18.825)
MF-Low	1.475 (1.295)	680.356 (433.528)	12.046 (9.405)

Note. Counts are extracted from a mega-corpus (~6 million words) of American English in naturalistic interactions with typically developing children between the ages of 2-6 year. Numbers represent the frequency per 1 million words. Frame Frequency refers to the frequency of the entire three-word frame (e.g., “Brush your teeth”), Object Frequency is the number of times the noun in the frame appeared on its own, and Verb-Object Frequency counts the occurrence of the verb and object with any intervening item (how often “Brush X teeth” occurs).

2.1.3.2. Visual Stimuli

Colorful, photorealistic images of all experimental sentence-final objects were selected. For each image, any extraneous background information was removed, and the remaining image was centered on a white background of 400 x 400 pixels. Several additional images were either used to direct the child’s attention to the center of the screen (e.g. a small flower, or star), or to provide longer breaks and encouragement during the study (e.g. a scene containing the characters from Finding Nemo).

2.1.3.3. Auditory Stimuli

All spoken items were recorded in a child-directed voice by a native American-English speaking female. Auditory stimuli were sampled on a mono channel at 44100 kHz sampling rate. Praat software (Boersma & Weenink, 2012) was used to align the critical onsets across stimuli. Sentences were adjusted to an overall duration of 1655 ms, with the following durations for each word: verb – 693 ms, pronoun/article/preposition – 379 ms, noun – 583 ms. The intensity of all

stimuli was adjusted to a mean of 70 dB. The speaker also recorded several additional encouraging phrases to be used to maintain child attention during the experiment.

2.1.4. Procedure

Children were seated in a stationary chair in front of a 17-inch LCD monitor with attached eye-tracking camera. The monitor was attached to a remote arm mount so that the height and distance of the camera could be adjusted specifically to the child. The eye-tracking camera was positioned between 580-620 mm from the child's eyes. Auditory stimuli were presented via child-size headphones and the experimenter did not hear the sentences. This set up also insured that the auditory stimuli did not influence the experimenter's judgment of the child's responses.

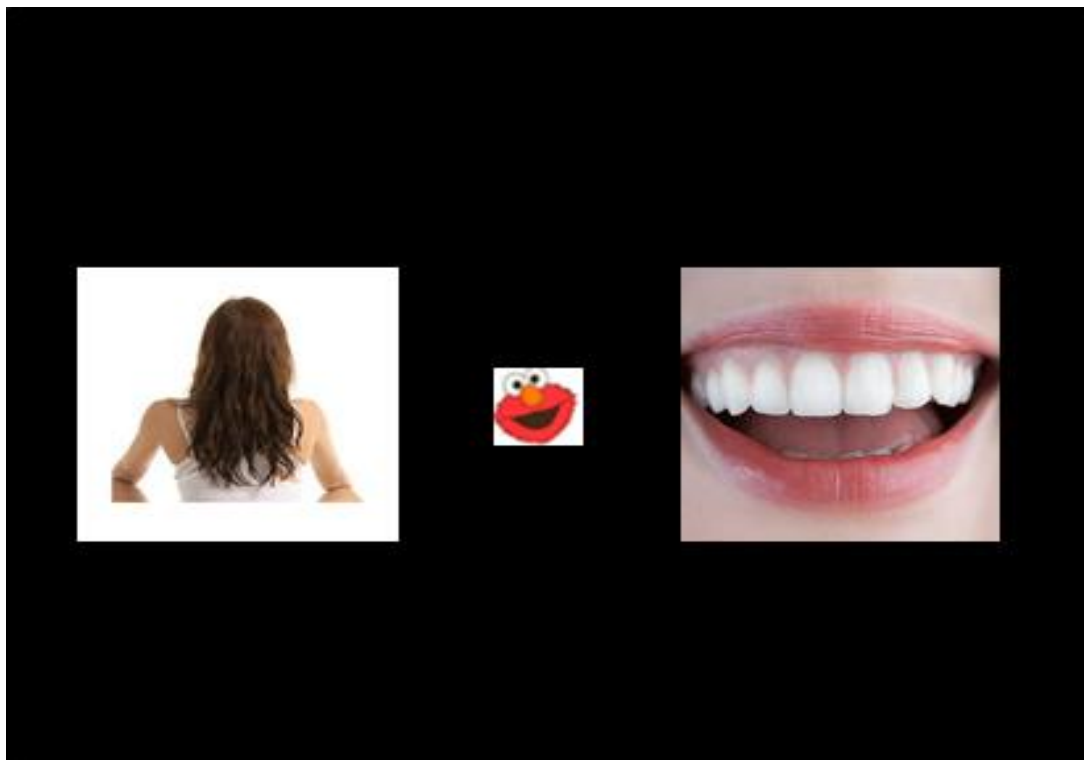
Before the start of the experiment, children were told they would hear some sentences and would be asked to point to images that were depicted on the screen that went along with these sentences. They were given a single example trial where they heard a sentence and pointed to a picture. No children required additional explanation after this step, and the eye-tracker was then calibrated using a standard five-point calibration routine. Children generally viewed these calibration stimuli without additional instruction, but the experimenter did occasionally remind the child to direct their gaze towards the calibration images.

Next, the experimental trials began. Each trial started with the presentation of a drift-check image (a 20-point black-and-white bull's-eye image), to ensure the child was attending to the screen before the onset of the experimental trial and that the calibration was maintained throughout the course of the experiment. Next, the Target (e.g. *teeth*) and Distractor (e.g. *hair*) images appeared in silence for 2000 ms, followed by the appearance of a gaze-contingent 100 x 100 pixel center image as the sound "Look!" was played. This central image remained on the screen until the tracker detected that the child fixated within the 100 x 100 pixel bounds of the image for 100

ms. Then, the sentence stimulus was spoken (e.g. *Brush your teeth*), and the images remained on the screen until the child pointed towards one of the images and the experimenter used the computer mouse to click on the child's selection. See Fig. 1 for an illustration of an experimental trial.

Figure 1

An Example of an Image Used During the Visual-World Paradigm Sentence Recognition Task



Each image appeared four times across the study, twice as the Target and twice as the Distractor. Each image appeared on the left and right sides with equal frequency across trials. With 14 Target-Distractor image pairs, this yielded 28 total experimental trials (14 in the Familiar Frame condition and 14 in the Modified Frame condition) that were presented in four blocks of seven trials each. Trials were equally distributed across all conditions in the study, and presented in pseudo-random order, such that no single condition appeared in more than two sequential trials.

Each participant therefore contributed equally to all conditions of the study, yielding a completely within-subjects design. Between each block, children saw filler trials that contained images of smiling animals or popular cartoon characters. Children usually completed the entire experimental procedure in less than 10 minutes.

2.1.4.1. Eye-Movement Recording

Children's eye-movements were recorded at 500 Hz as they viewed the experimental trials using an SR-Research Eyelink 1000+ eye tracker. Areas of interest were pre-defined as the two 400 x 400 pixel regions comprising the Target and Distractor images. Data were collated into 50 ms bins for offline analysis.

2.1.4.2. Eye-Movement Data Analysis

Our primary analyses address whether and when children used the first two words of the multi-word frames (verb + pronoun/article/preposition) to generate fixations towards the Target object. Our analyses sought to determine whether recognition of the Target object varied as a function of Frame Frequency (High and Low) and Frame Type (Familiar and Modified), and specifically, whether these fixations were anticipatory in nature. As in prior work (Borovsky et al., 2012; Kamide et al., 2003), we consider conditions where looks to the Target object exceed those to the Distractor before the onset of the spoken target word as strong evidence for anticipatory activation of the target item. Despite the straightforward nature of this criterion for prediction, there is substantial variability in the eye-tracking field regarding precisely how predictive fixations should be measured. Here, we take two approaches that quantify eye-movement behavior over relatively broad and fine-scale time-windows, as described below. First, we assess predictive effects using a broad time-window analysis spanning the time-window from verb onset to noun onset (i.e. the anticipatory period before the onset of the spoken target noun). In this broad time-window, we

measured overall accuracy of prediction of the Target as a proportion of fixations towards the Target divided by the total looks towards the Target and Distractor. This accuracy measure was selected as it is commonly used in studies of lexical development (see Swingley, 2011 for a review). This broad time-window analytical approach has been commonly implemented in the developmental eye-tracking literature, and summarizes looking behavior over a relatively long time-window spanning several hundred milliseconds. However, this approach can overlook more rapid fixation phenomena that are captured by the high sampling rate of modern eye-tracking equipment. To investigate these rapid processes, we use a non-parametric cluster test procedure that can characterize the timing and duration of a preference to view the Target image while controlling for multiple comparisons (see Groppe, Urbach, & Kutas, 2011; Maris & Oostenveld, 2007 for expanded description and tutorials on this method). This cluster-based permutation procedure has been already implemented in the developmental eye-tracking literature (Borovsky, Ellis, Evans, & Elman, 2016; Von Holzen & Mani, 2012), and we use it here to identify time-windows across the sentence where fixations towards the Target significantly exceed those to the Distractor object. If the Target divergence point begins before the onset of the spoken target, this is taken as evidence for predictive processing.

The basic procedure for this non-parametric cluster-based permutation test is as follows: First, the procedure is used to identify “clusters” – defined as temporally adjacent time points in the dataset where the statistical difference (by t-test) is significantly different (at $\alpha = 0.05$ threshold). The t-statistics of these clusters are summed to create a cluster-t statistic. These cluster-t statistics are then compared to randomly permuted t-statistic distribution of the same dataset (following the permutation approach outlined in Barr, Jackson, & Phillips, 2014, Appendix).

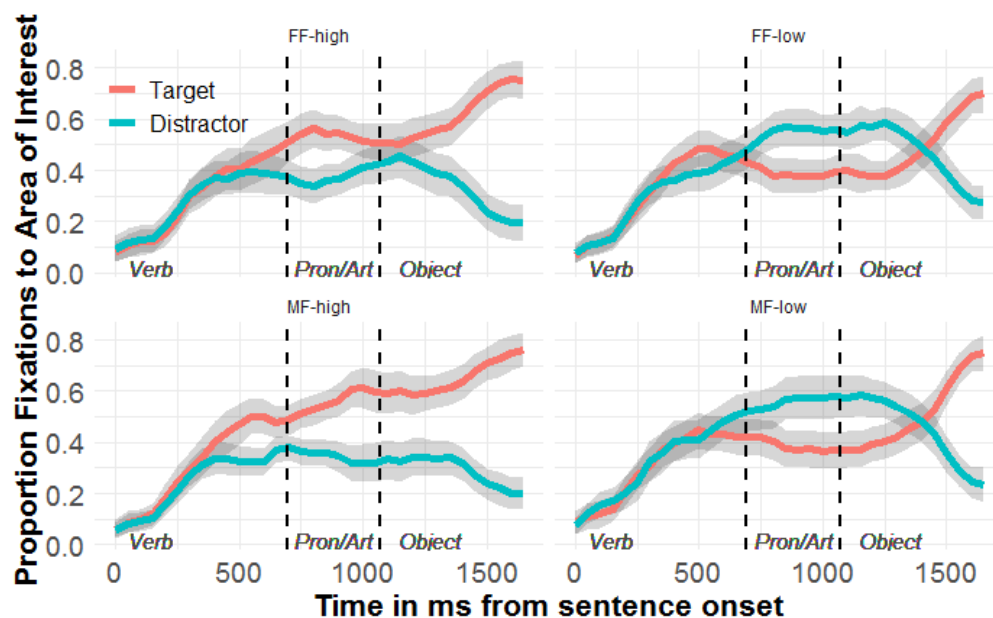
Finally, we wished to verify that the trials we included in the analyses were those in which children were attending to the auditory and visual stimuli. First, we removed trials in which children selected the incorrect target picture. This resulted in the removal of 31 trials (out of 952 total, or 3.26% of total trials) with incorrect target selections, leaving 921 total correct trials to be entered into subsequent analyses. Second, we sought to identify trials in which children were not visually attentive to the screen during the time-window of analysis. Thus, we removed trials in which children did not view the target or distractor picture for more than 25% of the analyzed time-window. This led to removal of 10% of remaining trials (92 out of 921 correct trials, with 829 trials remaining in the final analyses) from the broad time-window analysis which was limited to the anticipatory time-window. For the finer-grained cluster analysis involving the entire sentence time-window, this track loss threshold led to the removal of 8.9% of remaining trials (82 out of 921 correct trials, with 839 remaining in the final analyses).

2.2. Results

Fig. 2 illustrates the time-course of fixations towards the Target and Distractor images across the spoken sentences in the four experimental conditions (FF-High, MF-High, FF-Low, MF-Low). This figure illustrates an expected increase and divergence in fixations towards the Target relative to the Distractor as the sentence unfolds.

Figure 2

Time-Course of Proportion Fixations to Target and Distractor Images Across Experimental Conditions



Note. Error cloud represents Confidence Intervals (CIs). Lines mark word onsets.

2.2.1. Experimental Effects on Target Fixations During a Pre-determined Anticipatory Time-Window

To explore whether we find broad evidence for predictive looking and if this interacted with Frame Type or Frame Frequency, we calculated mean accuracy for each experimental condition as a proportion of fixations towards the Target relative to the Distractor over the entire anticipatory time-window from verb onset to noun onset (Fig. 3; data for individual children can be seen in the line graphs in Fig. 4). Using the lme4 library (Bates, Maechler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2016), we fitted a linear mixed-effects regression model to test for differences in accuracy as a function of Frame Type and Frame Frequency. We additionally included fixed effect factors of age (in months) and trial number (centered) in the model. For random effects, we began with a full model that included by-item and by-subject random intercepts, along with a by-subject random slope of experimental condition, however, this model did not converge, and so we iteratively simplified the random effects by removing those that explained the least variance, starting with removing the slopes, until the model did converge

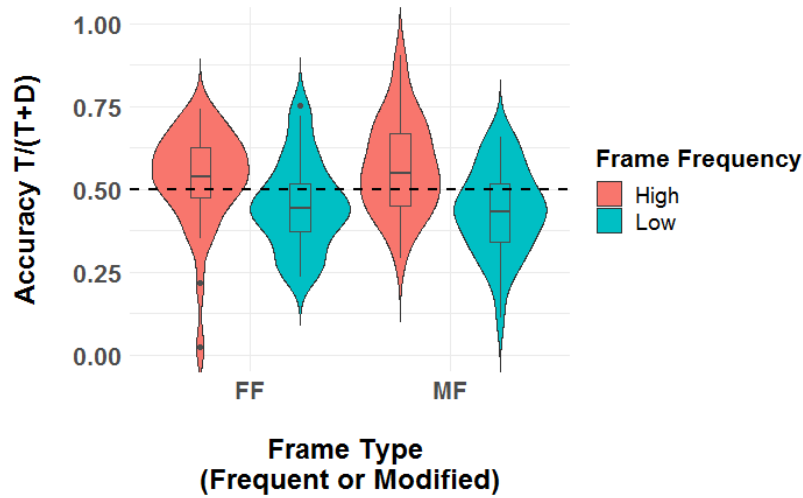
(final model: $Accuracy \sim Frame\ Type * Frame\ Frequency + Trial\ Order + Age + (1 / Items)$); see Table 2; a more detailed version of Table 2 that includes random effects is available in the supplementary materials, section 7.1). We also compared this final model to a model with the same fixed effect structure and no specified random effects, and there was a marginally significant improvement in fit ($LR = 3.59, p = .058$).

As is evident from Fig. 3 and Table 2, we found a significant effect of Frame Frequency on performance: children view objects that occur in high frequency frames during the anticipatory time-window to a greater extent than objects that occur in low frequency frames. However, we did not find evidence for a statistically significant effect of Frame Type on the prediction of the Target object. We also did not observe evidence for a significant interaction between Frame Type and Frame Frequency; see Fig. 1 in the supplementary materials section 7.2 in which we collapse Frame Types into one group and present proportion of fixation to the target and distractor only as a function of the frequency of the frame, i.e., high vs. low). These results suggest that these effects are likely to be driven by the association between the verb and the sentence-final object, rather than the sentence frame per se. Further, we did not find a significant effect of Age on performance, but we found a significant effect of Trial Order, with a negative beta-estimate indicating that performance declined over the course of the experiment, suggesting that some fatigue effects may have influenced performance across the study. Importantly, this result suggests that children did not adapt their performance to generate more predictive fixations across the experimental trials¹.

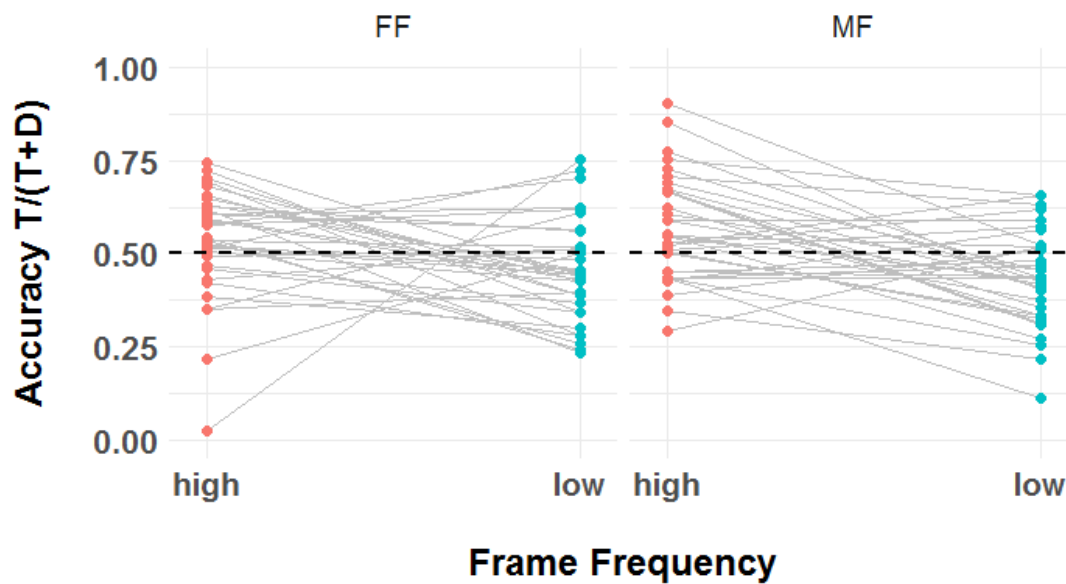
¹ One experimental item, “Drink of water” did not contain a verb (see Appendix A). Following the recommendation of an anonymous reviewer, we re-ran our analyses after removing all “drink” trials. The statistical patterns remained the same, with the exception of the *Trial Order* effect that became no longer significant ($p = .102$). Importantly, the main effect of *Frame Frequency* remained significant even after removing these trials. The statistical output of these analyses is provided in Table 1 in the supplementary materials section 7.3.

Figure 3

Fixation Accuracy during the Anticipatory Time Window by Condition Across all Participants in Experiment 1

**Figure 4**

Change in Accuracy during the Anticipatory Time Window Across Frame Frequency (High/Low) in Each of the Frame Types (FF, MF) for all Participants in Experiment 1



Note. The dotted horizontal line at 0.5 indicates equivalent looking towards the target and distractor.

Table 2*Regression Model for Experiment 1*

Effect/Slope	Estimate	CI	t-value	p-value
(Intercept)	0.496	0.461 – 0.531	27.672	<.001
Frame Type	-0.006	-0.077 – 0.064	-0.179	.861
Frame Frequency	0.122	0.072 – 0.172	4.778	<.001
Type x Frequency	-0.056	-0.156 – 0.043	-1.110	.268
Age	0.004	-0.020 – 0.029	0.342	.733
Trial Order	-0.031	-0.059 – -0.003	-2.202	.029

Note. The model summary including random effects is listed in supplementary materials, section 7.1.

2.2.2. Finer-Grained Time-Course Analysis

Analysis of fixations over broad time-windows can sometimes mask effects that occur over much shorter timescales. Therefore, our next analytical goal was to explore whether the timing of Target divergence varied as a function of experimental condition using a cluster-based permutation analysis, as described above. In this analysis, we initially calculated a dependent measure of Target looking relative to that of the Distractor item as a log-gaze proportion ratio in each time-window that defines the bias of looking towards the Target or Distractor on a scale from negative to positive infinity, centered around 0 (see Arai, van Gompel, & Scheepers, 2007; Borovsky et al., 2016 for a similar approach and additional explanation). These values were then used to generate fine-grained comparisons of fixations across experimental conditions. We initially compared log-gaze ratios as a function of Frame Frequency in the Familiar and Modified conditions separately to determine whether and when frequency influenced looking as a function of Frame Frequency in each frame condition. Next, we compared log-gaze ratios in High and Low Frame Frequency conditions as a function of Frame Type (Modified or Familiar) separately to explore whether there were timing

differences in looks towards that Target as a function of Frame Type in each Frame Frequency condition. The results of these cluster analyses are outlined in Table 3 and support the findings of the broad time-window analyses. When comparing differences between High and Low frequency in Familiar and Modified frames separately, significant differences emerge starting in the anticipatory time-window and continue until approximately 100 ms before sentence offset, when participants begin to look away from the Target object. No cluster-based differences emerge when comparing Familiar and Modified frames in each of the Frequency conditions, supporting prior analysis that did not find any main or interaction effects involving Frame Type.

Table 3

Timing Differences Identified by Cluster-Based Permutation Analyses in Experiment 1

Comparison	Time-window
Familiar Frame	750-1550 ms
High vs. Low	($t = 47.03, p < .0001$)
Modified Frame	800-1550 ms
High vs. Low	($t = 57.05, p < .0001$)
High Frequency	No clusters identified
Familiar vs. Modified	
Low Frequency	No clusters identified
Familiar vs. Modified	

Note. Timing is measured from the onset of the spoken sentence, which was 1655 ms in duration. The time-window where looking at the Target significantly exceeded that to the distractor is reported in ms, followed by the Cluster t statistic and Monte Carlo p value for that window, in parentheses. The onset of the sentence final target was at 1070 ms post sentence onset. **P-values are Bonferroni-corrected.**

2.3. Discussion

Using the visual-world paradigm and a design inspired by prior language production work (Arnon & Clark, 2011; Bannard & Matthews, 2008), we sought to examine whether and how children utilize multi-word information during online speech processing, and whether they form linguistic predictions based on multi-word associations. Specifically, we examined whether children are sensitive to the associations between multiple words in a sequence frequent in their input (e.g., *Brush your teeth*) or only to the event associations reflected in the verb-object relationships (e.g., *brush-teeth*, *brush-hair*), a contrast that was not directly tested in previous research. We asked whether children would be faster to anticipate sentence-final objects in higher-frequency frames, compared to less frequent ones (e.g., *Brush your teeth* vs. *Brush your hair*), and whether gaze patterns to high-frequency frames reflect event knowledge, or real-time summation of phrasal frequency using modified frames (e.g., *Brush her teeth* vs. *Brush her hair*). We found that children generated a stronger prediction for the object in higher frequency sequences (*Brush your teeth* vs. *Brush your hair*) even when both outcomes were highly plausible (*hair* and *teeth* are both “brush-able”). However, when the event remained constant while the frame was modified by a single word, rendering those modified sequences less frequent in the input (*Brush your teeth* vs. *Brush her teeth*), children were equally likely to predict the higher frequency ending of the frame regardless of whether it was modified or not. These findings suggest that children relied primarily on the frequency of the verb-object associations for prediction, and do not provide evidence for children’s use of multi-word information during online processing.

There are several possible explanations for our findings. First, it is possible that our results are driven by children’s knowledge of abstract frames with varying levels of productivity in the different slots rather than non-adjacent verb-object associations. Specifically, the intervening

element (pronoun, article, or preposition) in each of the frames we used is much less productive as an abstract slot compared to that of the noun at the end of the frame (i.e. there are a lot more options that can be used in the noun slot as opposed to the middle slot); slots that are more productive end up being organized around the most frequent items (*teeth* may act as a prototype for the noun slot in a frame that begins with *brush*; Bybee, 2006). Such difference in the productivity of the slots may have led children to represent multi-word information that relies on abstract representations rather than multi-word associations (*brush* may be followed by a lot of nouns, and there are a few options for what may intervene between them; *teeth* is the prototypical option and is preferred over other options). Unfortunately, the current design does not allow us to examine the impact of slot productivity as our stimuli did not systematically vary on this dimension. Only future work can help address this issue².

Second, it is possible that the brief duration of the intermediate element between the verb and the object did not provide enough time for children to modify their expectations for the sentence-final object. The spoken pronoun/article/preposition had an average duration of 379 ms, only a little longer than the 250-350 ms range during which children in this age group (Trueswell, Sekerina, Hill, & Logrip, 1999) and adults (Matin, Shao, & Boff, 1993) show a divergence in their eye-movements in response to linguistic stimuli. Many visual world paradigm designs incorporate intervening items, such as adjectives that allow time for eye-movements to unfold for this very reason. Additionally, these intermediate elements carried only minimal (if any) semantic content that the listener could leverage to generate an anticipatory fixation. It is therefore possible that the particular structures that we selected for this experimental manipulation were not ideal to observe

² We thank an anonymous reviewer for suggesting this possible interpretation of our findings.

the phenomenon of interest, and that the predictive processes in question unfold over a slower timescale than the duration of a mono-syllabic article, pronoun, or preposition.

Third, given that we relied on naturalistic variation in frame frequency, it is possible that the frequency differences between the high- and low-frequency frames across the conditions (i.e. the difference between the frequency of FF-High and MF-High, and between FF-Low and MF-Low) were not large enough to modify fixation behavior and influence language processing, and hence result in an interaction effect between Frame Frequency and Frame Type. However, the number of naturally occurring trigrams that we could feasibly incorporate into this design were limited, and therefore, it is possible that our items may not have provided a strong enough manipulation to drive differences in eye-movements. This possibility has been recently attested in a study that examined children's fixations to a target upon hearing word bigrams that varied in their frequency (Jones, Cabiddu, & Avila-Varela, 2020). The results showed that children's fixations appeared to be influenced by the categorical presence or absence of the bigram in child-directed speech and by the graded changes of frequency, such that bigrams that appeared an average of 19 times per million words, and those that appeared an average of 206 times were processed in a similar manner by 2-year-olds. Hence, it is likely that the range of frequencies of our Frequent Frames (see Appendix A) was not large enough to capture differences in processing. Finally, it is theoretically possible that the impact of multi-word frequency on processing that has been previously attested in adults' comprehension (e.g., Arnon & Snider, 2010; Hernández, Costa, & Arnon, 2016; Reali & Christiansen, 2007; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay, Derwing, Libben, & Westbury, 2011), and children's productions (Arnon & Clark, 2011; Bannard & Matthews, 2008) does not impact children's comprehension of spoken language,

and that during online language processing, children do not rely on the frequency of multi-word sequences.

Further information that can aid in interpreting these results may be gleaned from the next experiment. Experiment 2 was designed to better distinguish between the impact of pairwise associations and multiword frequency and to examine whether children can learn novel multiword frequencies from a short exposure and use them in predictive processing. We utilized a training design which allowed us to control the frequency of multi-word associations by presenting learners with novel events described using multiword sequences. The second study also included longer intervening durations between the informative items in the frame and the sentence-final outcome. This design allows us to shed additional light on children's reliance on multiword frequency during online processing.

3. Experiment 2

3.1. Methods

3.1.1. Participants

The same children who participated in Experiment 1 were also invited to take part in Experiment 2 during the same laboratory visit. Participants always completed Experiment 1 before being asked if they would like to continue to Experiment 2. All children agreed to participate in the second task.

3.1.2. Stimuli

3.1.2.1. Story Materials

Participants heard four sets of stories (8 stories), such that each one conveyed novel events comprising relationships among two agents, two actions, and four thematic objects (2 per action). Stories were presented on a computer, with each "page" in the story corresponding to a colorful

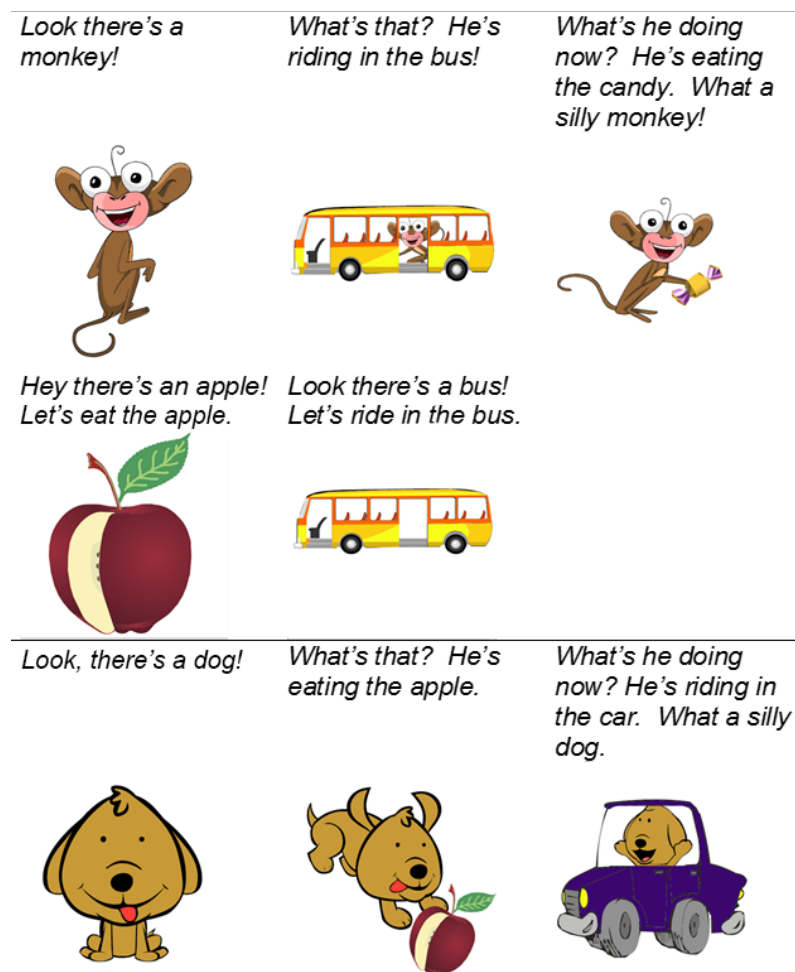
cartoon image (400 x 400 pixels in size). These images were created using the comic strip creation website, Toondoo (<http://www.toondoo.com>). Each image in the story was accompanied by spoken narration that reinforced the visually depicted events in the story. The narration was pre-recorded by a native English-speaking female at 44.1 kHz on a mono channel and normalized to mean intensity of 70 dB.

The goal of each narration was to develop combinatorial (agent-action-object) and associative (action-object) relations that were equivalent in frequency. Therefore, the stories followed a narrative structure similar to that used in Borovsky et al. (2014) with several key modifications. Like in prior work, each story described actions carried out by two different agents (e.g. a monkey and a dog), who each performed two different actions (e.g. riding and eating) on different objects (see Appendix B for a list of all agent-action-object combinations used in the experiment). These narrative associations served to establish four main combinatorial relations among the two agents, two actions, and four objects in the story set. Crucially, in the current task, listeners also viewed two additional action-object relations in sentence frames that did not mention either of the agents (e.g. ride-bus and eat-apple, see Fig. 5). This additional action-object association served to level the summed association frequency for two of the four combinatorial relations in the story. More concretely, consider the example in Figure 4. Here, the child's prediction following a sentence like "*The monkey rides in the...*" has two potential outcomes depending on how the child uses the frequency information in the story. One possibility is that the child may have attended to all pairwise frequencies across the story. This option would suggest that the child should not generate a specific expectation for BUS or CAR when hearing the sentence "*The monkey rides the...*" since ride-bus and ride-car appeared with equal frequency in the story (twice each). However, if the child computes frequencies at the level of the trigram

(*monkey-ride-bus*; which appears only once), in addition to pairwise relations, then, the statistics of the story would lead the child to specifically expect BUS after hearing the sentence.

Figure 5

Example of Images and Accompanying Spoken Narration from Experiment 2



As illustrated in Fig. 5, the events in each story were presented in identical order across each of the eight stories in the experiment. First, the narrator introduced one of the agents (e.g. MONKEY, *Look! there's a monkey*), and two specific event associations are established for that particular agent (e.g. riding in a bus, eating a candy). Next, the listener heard two events described

that are not associated with a novel agent, one object that was associated with the first introduced agent (*Let's ride in the bus*) and a second that corresponds to the next-to-be introduced agent (*Let's eat the apple*). Finally, the second agent appears (e.g. DOG, *Look! there's a dog*), along with two novel event associations (e.g. *riding in a car*, *eating an apple*). The event relations were counterbalanced across versions of the experiment such that all potential combinations of the agent-action-object appeared across versions. For example, in one version, a dog may have eaten an apple and have ridden in a car, while in another version the dog may have eaten a candy and have ridden in a bus. Even though the goal of constructing these cartoon narratives was to develop very low-frequency events that participants were unlikely to have previously encountered in the world, this counterbalancing served as extra insurance against the possibility that participants may have had pre-existing biases to prefer certain constructed event combinations over others.

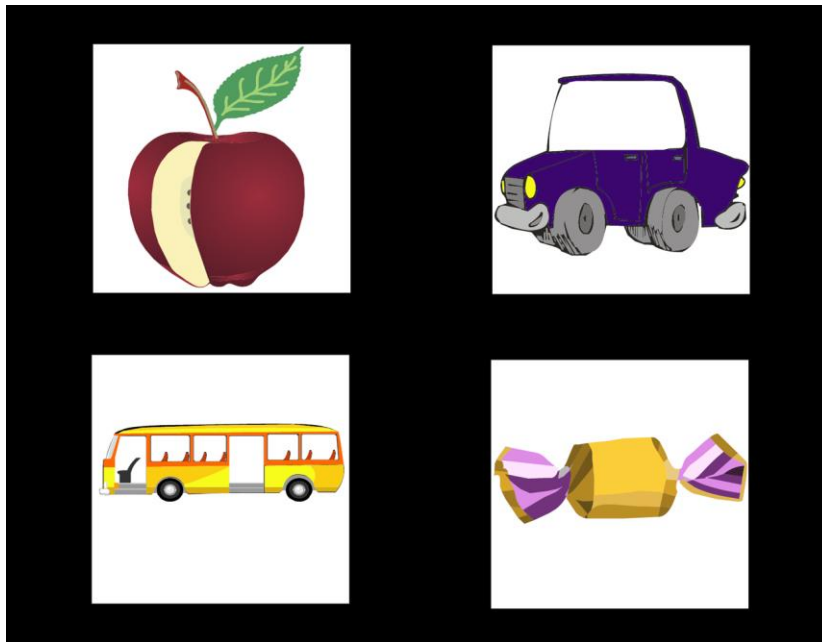
It should additionally be mentioned that, as in prior work with similar stimuli, the verbal description of the event relations never explicitly mentioned all of the elements of the event in the same sentence. For example, in the dog-rides-car example, the narration indicated that the dog was present in one scene in the story, and then later the narration conveys that *He's riding in the car*, but the narration never stated verbatim that *The dog is riding in the car*. The goal was to prevent the learner from hearing the same acoustic stimulus during the story phase that was presented in the sentence recognition phase of the experimental task. It should also be noted that because the combination of the event relations was controlled across versions, this arrangement further ensured that the same acoustic signal during the narration conveyed the same event relation for all agents across versions (i.e. The same auditory file was used to describe *He's riding in the car*, for both the dog, and the monkey across the separate experimental versions where each agent performed this action).

3.1.2.2. Sentence Recognition Stimuli

Each story was followed by a visual-world sentence recognition task that was designed to measure the real-time recognition of the two critical combinatorial relations from the story. Participants initially viewed four objects on the monitor. These four images consisted of the thematic objects that initially appeared in the prior story (e.g. BUS, CAR, CANDY, APPLE; see Fig. 5). Images initially appeared on the screen in silence for 2000 ms, before the onset of one of the two critical experimental sentences (e.g. *The monkey rides in the bus*). Each sentence consisted of the same form: Article – Agent – Action – Article – Theme. As in experiment 1, the duration and onset of each word within the sentences was aligned using Praat, such that the onset of each word in each sentence occurred at the same time across all sentences in the study. The alignment of word durations in each sentence allowed for direct comparison of looking behavior with respect to the timing of the auditory stimulus across all experimental sentences. For each sentence (e.g. *The monkey rides the bus*), each image was classified in one of four conditions: Target (BUS), Agent-Related (having been associated with the agent in the prior story, in the example in Fig. 6, this was the CANDY), Action-Related (i.e. the other ride-able item, CAR), and Unrelated (i.e. was not associated with the agent or the action of the sentence from the information presented in the novel story, in this case APPLE). Although we developed the novel relations in this study to be infrequently encountered by young children, we also counterbalanced the stimuli to prevent the possibility that participants may have had some pre-existing biases for some relationships over others. This counterbalancing scheme ensured that each image served in all Target and Distractor conditions across all versions of the study, yielding a balanced within-subjects design.

Figure 6

Illustration of Images that Appeared in the Visual World Sentence Recognition Task



3.1.3. Procedure

3.1.3.1. Experimental Task

As in Experiment 1, participants were seated in the same comfortable chair in the same experimental room equipped with an SR Eyelink 1000+ eye-tracker. The tracker was re-calibrated before starting Experiment 2. Participants were instructed that the task would be similar to the first, except that they would now hear stories about some creatures and that they should do their best to understand the story. Each child was told that after each story, they would hear additional sentences accompanied by four pictures and that they should point to the picture that “goes with the sentence.” Participants completed a single sentence comprehension practice trial before starting the experiment.

Each block in the study began with one of the four possible short story sets. Each story was narrated in a child-directed voice and each “page” in the story was depicted on the computer screen as a colorful cartoon image that matched the spoken description. After each story, participants heard two sentences outlining one of the agent-action-object relationships from the previous narrative, while they viewed items that varied in relationship to the sentence final target. Hence, participants completed eight trials in this sentence recognition task, and the entire task (including all story and sentence recognition phases) took less than 10 minutes to complete.

3.1.3.2. Eye-movement Recording

Eye movements were recorded with the same equipment and parameters as in experiment 1.

3.1.3.3. Analytical Approach

As in Experiment 1, our analyses sought to establish whether and when participants fixated towards the target and competitor objects in relation to the spoken sentence. Our primary question concerned whether or not participants would make anticipatory eye-movements towards the unspoken target of the sentence by utilizing the combinatory relations or lexically-associated relations from the prior story. Therefore, our analyses focused on the anticipatory time-window comprising the onset of the spoken verb to the onset of the sentence-final noun. As in Experiment 1, analyses were carried out over broad and fine-grained time- windows.

As in Experiment 1, we wished to verify that children were accurate and attentive in the analyzed trials. Therefore, we first removed trials in which children selected the incorrect target picture. This criterion led to the removal of 11 out of 272 total trials (4.0%), leaving 261 trials. As in Experiment 1, we next removed trials in which children did not view the target or distractor picture for more than 25% of the analyzed time-window. This criterion led us to remove a further 23 trials (8.8% of remaining trials) from the broad time-window analysis which was limited to the

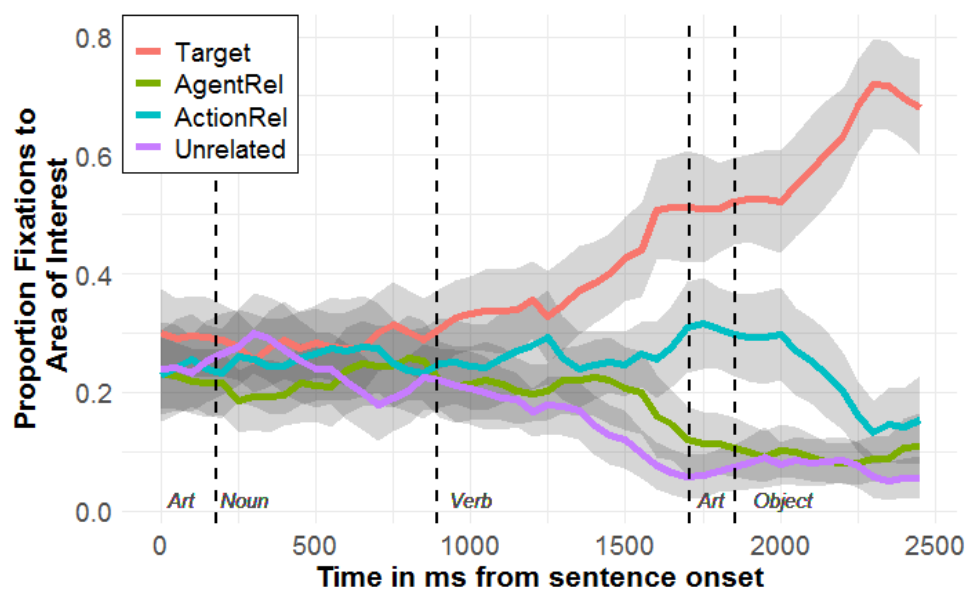
anticipatory time-window, leaving a total of 238 total trials in this analysis. For the fine-grained cluster analysis involving the entire sentence time-window, this criterion led to the removal of 8.0% of remaining trials (21 out of 261, with 240 remaining).

3.2. Results

Fig. 7 highlights the time-course of fixations towards the Target and three competitor images (Agent-Related, Action-Related, and Unrelated) as the spoken sentence unfolds. As expected, fixations towards the Target image increase dramatically after it is named in the spoken sentence. It is also visible that children fixated on the Target item with relatively greater magnitude than the other objects before the onset of the spoken object (i.e. predictively). These patterns are explored in the broad and fine-grained time-window analyses below.

Figure 7

Time-Course of Fixations Towards the Target and Competitor Items From Sentence Onset to Sentence Offset



Note. Error cloud represents CIs. Dashed lines mark word onsets.

3.2.1. Experimental Effects on Target Fixations During a Pre-determined Anticipatory Time-Window

To determine whether there was evidence for prediction of the Target object compared to other distractors, we first calculated the proportion fixations towards Target and competitor objects during the anticipatory time-window. Following prior work using similar paradigms, if looks towards the Target exceed those to the other competitors during this anticipatory period (e.g. Borovsky et al., 2012; Kamide et al., 2003), then this is taken as strong evidence for anticipation of the target due to the multi-word information gleaned from the story (e.g. monkey-rides-bus). If, however, looks towards the Target and Action-Related object (e.g. car) do not diverge before the sentence-final target is spoken, then this pattern is taken as support for the simple (pairwise) lexical-association account (reflecting the association between *ride* and ride-able objects, i.e. bus and car). We therefore entered area of interest (AOI: Target, Agent-Related, Action-Related and Unrelated) into a linear mixed-effects model along with Age (in months, centered) and Trial Index (centered) with the lme4 library (Bates, Maechler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2016). The initial model included random effects for participants and items; specifically, we included by-item and by-subject random intercepts, however, this model experienced convergence issues due to little variance explained by the random effects. This model was compared to a model with no random effects, and was not found to have any improvement in fit ($LR = 5.6e-07$, $p > .999$), so the final model reported here includes fixed effects only ($Accuracy \sim AOI + Age + Trial Order$); see Table 4; the model summary for the (non-converging) model with random effects is reported is available in the supplementary materials, section 7.4).

These analyses revealed no main effect of Age or Trial Index (see Table 4 for beta coefficients for AOI levels and Age and Trial Index). Follow-up pairwise tests of fixations towards

the Target compared with other competitor objects indicated that mean Target fixation exceeded those to each of the other competitors during the anticipatory time-window [Target vs. Action-Related: $t(920) = 5.25, p < .0001$; Target vs. Agent-Related, $t(920) = 8.04, p < .0001$; Target vs. Unrelated: $t(920) = 9.69, p < .0001$]³. That is, children relied on multi-word information to generate predictions about the final object.

Table 4

Regression Model for Experiment 2

Effect/Slope	Estimate	CI	t-value	p-value
(Intercept)	0.250	0.223 - 0.268	26.591	<.001
Area of Interest:				
Target	0.152	0.120 - 0.18	9.35	<.001
Action Related	0.013	-0.029-0.045	0.801	.423
Agent Related	-0.061	-0.093- -0.029	-3.729	<.001
Age	-5.62e-18	-0.018 – 0.018	0	>.999
Trial Index	-5.57e-18	-0.018 – 0.018	0	>.999

Note. A version of this table for model outputs with random effects is available in the supplementary materials, section 7.4.

3.2.2. Finer-grained Time-Course Analysis

As in Experiment 1, we used a cluster-based permutation approach to gain insight into the timing of the experimental effects. Again, we initially calculated a dependent measure of Target looking relative to that of each Distractor item (Agent-Related, Action-Related and Unrelated) separately as a log-gaze proportion ratio in each 50 ms bin that defines the bias of looking towards the Target

³ These statistical patterns were identical when carrying out the same analyses using an empirical-logit transformation of the data.

or Distractor on a scale from negative to positive infinity, centered around 0. The results of these cluster analyses are outlined in Table 5; they clearly support the early broad window analysis that fixations towards the Target diverged from those to each of the other competitors well in advance of the spoken Target onset (which occurs at 1854 ms after sentence onset).

Table 5

Significant Differences Between Target and Distractor Objects Identified by Cluster-Based Permutation Analyses

Comparison	Time-window
Target vs. Action Related	1350-2500 ms (cluster $t = 108.54$, $p < .0001$)
Target vs. Agent-Related	950-2500 ms (cluster $t = 200.07$, $p < .0001$)
Target vs. Unrelated	950-2500 ms (cluster $t = 235.51$, $p < .0001$)

Note. Timing is measured from the onset of the spoken sentence, which was 2455 ms in duration. The time-window where looking at the Target significantly exceeded that to each of the distractors is reported in ms, followed by the Cluster t statistic and Monte Carlo p value for that window, in parentheses. The onset of the sentence final target was at 1854 ms post sentence onset. **The reported p-values are Bonferroni adjusted.**

3.3. Discussion

Using a training paradigm, in which children were exposed to illustrated stories about animal characters engaged in unfamiliar actions, we sought to examine whether children form novel multi-word associations, and whether they use them in predictive processing. Importantly the frequency of multi-word combinations and individual lexical associations of the various agent-object and action-object pairwise combinations were fully controlled. If children simply generated predictions based on pairwise associations, rather than attending to higher-order multi-word information, then they would not have generated a specific prediction for one particular final object

and instead would have shown equivalent fixations toward the Target and Action-Related items (for example, participants would have looked towards the BUS and CAR equally as they heard the verb “rides” in *The dog rides the bus*). Only if listeners rapidly attended to and interpreted information in the multi-word frame would they generate a specific prediction for a single sentence-final outcome that was more likely based on the story.

In line with our predictions, the results of this manipulation showed that children generated specific predictions for a single sentence-final outcome based on the multi-word contingencies in the story, rather than the pair-wise lexical associations. Specifically, children’s fixations to the Target exceeded those to all other distractors during the anticipatory time-window. These results show that children were able to aggregate information about agents, actions, and objects from the stories. Specifically, even though they never heard the sentence “The monkey rides in the bus” in this exact syntactic structure in the story, children were able to create the multi-word association between *monkey*, *ride*, and *bus* after hearing “Look! there’s a monkey! What’s that? He’s riding the bus! What’s he doing now? He’s eating the candy. What a silly monkey!”. Creating such a multi-word association allowed them to predict BUS (and not CAR) long before it was produced, and shortly after hearing RIDE. These findings show that carefully controlling the frequency and content of multi-word information provides evidence for children’s use of multi-word information for predictive processing.

As we noted in our results’ section, Age and Trial did not contribute to this effect, suggesting that even the youngest children in our study (5 years old) were able to integrate multi-word information in real-time comprehension, and their ability to do so did not change (or adapt) over the course of the experiment. The lack of developmental change is consistent with a prior study that found that 5- to 10-year-olds could successfully use novel event knowledge to generate

linguistic predictions (Borovsky et al., 2014), and with findings from the language production literature showing that even preschoolers may be sensitive to these multi-word effects (Arnon & Clark, 2011).

4. General Discussion

In the real world, children and adults integrate multiple words to successfully process sentences. Yet, despite the growing evidence that multi-word sequences serve as building blocks in language learning, and that they impact language processing (Arnon & Christiansen, 2017; Arnon & Cohen Priva, 2013; Arnon & Snider, 2010), the impact of multi-word sequences on generating predictions during real time processing has been understudied. Specifically, while there is abundant evidence that both adults and young children rapidly generate real-time predictions about the form and meaning of upcoming speech, less work had explored whether and how children use multi-word information to generate predictions during online speech processing. Recent work on language production and comprehension suggests that adults and children are sensitive to multi-word contingencies above and beyond what can be simply explained by single word frequency effects (e.g. Arnon & Clark, 2011; Arnon & Snider, 2010; Bannard & Matthews, 2008), but it is not clear whether these findings would extend to real-time language comprehension. Therefore, the current project sought to explore whether multi-word frequencies impact real-time phrasal processing in young children. Our two experiments, using naturally occurring stimuli (Experiment 1) and newly constructed events (Experiment 2) provide the first thorough investigation of the effect of multi-word frequency on anticipatory and predictive processes. Importantly, our definition of multi-word sequences here refers to the associations between multiple words and their frequency; we do not refer to lexicalized wholes or abstract

frames. Further, we do not dive into the question of whether children treat multi-word sequences as holistic wholes, and our work does not address the issue of when those processes occur in development (see recent evidence indicating that 12-month-old infants distinguish between high- and low-frequency trigrams, suggesting that they extract multi-word sequences from early on; Skarabela et al., 2021).

In Experiment 1, children did not seem to use the frequency of known multi-word sequences to generate real-time predictions. Here, five to eight-year-olds generated predictions towards target objects based on verb-object (*brush-hair*) associations rather than multi-word information (*Brush your hair*). Yet, there were two main limitations to this particular design - the reliance on naturalistic variation in frequency, and the limited time to generate a predictive eye movement after the pronoun - that may have masked our ability to measure whether listeners were able to generate predictive eye-movements towards the upcoming stimuli. The design of Experiment 2 allowed us to further unpack children's ability to form predictions from multi-word sequences, by examining their learning of novel multi-word sequences and better controlling the frequency of pairwise associations vs. multiword frequency. We used a learning paradigm that did not rely on naturalistic variation in multi-word frequency, and that utilized a sentence structure that provided sufficient time for listeners to generate real-time predictions following a sentential verb. The findings in Experiment 2 indicated that multiword frequency did impact language processing and that children used such information to form real-time predictions during language comprehension. Our results support an account of language processing where such multi-word information is learned and computed rapidly – at least when it has semantic content. Importantly, our results represent the first evidence that, at least in *some* circumstances, multi-word frequency

can over-ride individual lexical associations and support real-time linguistic comprehension and prediction in children.

This finding raises intriguing questions regarding the relative impact of single-word and multi-word associations during online comprehension. Generating predictions from larger multi-word sequences may not always be optimal in cases where one word may provide greater informativity about the upcoming event than another (as in experiment 1), or in cases where the linguistic message is highly uncertain. In particular, children may weigh multi-word and word information differently depending on many sources of information, including the strength of the pairwise and multi-word associations, event familiarity, speaker reliability, and semantic informativity. Such effects have been found in adult production, where, for example, the impact of trigram frequency on phonetic duration is stronger than that of word frequency for highly frequent sequences (Arnon & Cohen Priva, 2013). This contextual sensitivity may help explain why our first experiment, using brief utterances, and more naturalistic stimuli found only moderate support for the impact of multi-word information on language comprehension. In the case of the stimuli in Experiment 1, listeners seemed to primarily rely on the most semantically informative unit, the verb, to guide predictions rather than on the less salient and informative function word (preposition/article/pronoun). However, in our constructed sentences in Experiment 2, these very same participants showed a different pattern of behavior: when each word in the multi-word sequence contributed clear semantic information (e.g. *monkey, ride, bus*), this manipulation led to a greater reliance on the multi-word sequence (relative to the verb-object associations) to form predictions. Our results also support the idea that proficient language users, including young children, can dynamically shift their strategies for linguistic prediction based on the information available in the linguistic task in that moment. This pattern is most consistent with recent proposals

that highlight adaptive processes in language comprehension as listeners flexibly shift their strategies to make optimal use of available information, such as that put forward in the “utility account” proposed by Kuperberg and Jaeger (2016).

Next, we note some limitations of the current work. First, as mentioned earlier, our experimental participants were school-age children, between the ages of 5-to-8 years of age. While important developmental changes in school-aged children occur in a number of linguistic processing abilities ranging from phonemic category distinction (Hazan & Barrett, 2000), to integration of referential cues to meaning in sentences (Snedeker & Trueswell, 2004; see Creel & Quam, 2015 for a review), this age group did not demonstrate strong evidence of developmental change in processing in either experiment. It should be noted that our testing of 5-to-8 year-old children was directly motivated by prior studies that found that school-age children (5- to 10-year-olds), but not preschoolers (3- to 4-year-olds) could successfully use a novel event to generate real-time predictions in a similar task (Borovsky et al., 2014). Future work is therefore needed to determine whether these patterns would also extend to younger children, and whether younger children would modify their ability to integrate multi-word information in sentence comprehension. However, the complexity of the current task may require significant modifications before it can be implemented with younger children.

Second, given that children’s linguistic experiences reflect their real-world knowledge and lived experiences of events, it is necessary to consider that novel events in Experiment 2 capitalized on the listeners’ preexisting verb/action knowledge, where novel agents interacted with various objects using actions that were semantically licensed by particular verbs. Thus, not all elements of these “novel” events were truly novel, and our results, like many from similar experimental tasks are not free from “contamination” from real-world knowledge. One concern about relying on these

pre-existing verb-object associations is that children may have used these pair-wise associations to drive behavior in Experiment 2. While this is certainly a possibility, our counterbalancing scheme ensured that all verbs were paired with all objects across all versions of the study. Therefore, if some verb-object associations were stronger than others (i.e., *ride-car* may be more frequent/strongly associated than *ride-bus*), these associations would have been balanced out in our design. While an even stronger test of our hypotheses would have been to generate a design where all relations were completely novel (with novel agents, actions and objects), it is likely that this design would have been challenging for children. Indeed, an experiment where adults learned novel combinatorial relations in an artificial language among novel agents, actions, and objects, required extensive repetition and training (Amato & MacDonald, 2010). To gain a fuller picture of how children learn about complex relations among words and events, much more work is needed, potentially using naturalistic (albeit complex) paradigms where learning is measured over multiple contexts over several days.

Third, in interpreting our results, we must acknowledge the possibility that children may have relied on visual priming from the pictures for predictive processing. The design of both our experiments does not allow us to ascertain the extent to which the visual context and linguistic cues have individually facilitated processing, or whether they worked in concert; hence, we do not think we can make any inferences about children's reliance, or lack thereof, on referential context. This issue has been directly investigated in several sentence processing studies with unclear results as to the extent to which children rely on referential context. For instance, some studies showed that referential processing may be harder for younger children (e.g. Kidd & Bavin, 2005; Rabagliati & Robertson, 2017; Snedeker & Trueswell, 2004), while others provide evidence that children can use referential contexts for generating accurate predictions (e.g. Huang & Snedeker,

2008; Reuter, Dalawella, & Lew-Williams, 2021). Only more work carefully controlling for the effects of pair-wise associations, multi-word sequences, as well as referential context can help address this issue in more detail.

Despite these limitations, our research provides initial support for the idea that children are sensitive to multi-word information in predictive language processing, at least in cases where the sequences contain multiple semantically informative units, as in Experiment 2. A clear and necessary path forward for future work would be to parametrically manipulate the potential utility of various multi-word sequences in predictive processing paradigms, perhaps by using a training paradigm that manipulates the local frequency of experimental test sentences that conform to prior multi-word associations across the course of a study. Further, it will be interesting to explore how different frequency and probability measures may impact and contribute to predictive processing. Specifically, instead of defining multi-word associations by their frequency in the input, it is possible to define them in terms of the conditional probabilities between the different elements (e.g. McCauley & Christiansen, 2019). As such, it will be intriguing to explore whether defining multi-word sequences in this manner provides a more sensitive measure of their impact on predictive processing. Moreover, individual differences are also likely to contribute to performance on these multi-word processing tasks, just as they do in other sentence processing tasks (Borovsky & Creel, 2014; Borovsky, Elman, & Fernald, 2012; Kukona et al., 2016; Mani & Huettig, 2012). Some potential candidates for skills that could contribute to multi-word sequence comprehension include working memory (Huettig & Janse, 2016), statistical learning skills (Misyak, Christiansen, & Tomblin, 2010), and cognitive flexibility/inhibitory control (Woodard, Pozzan, & Trueswell, 2016). It is likely that these same kinds of skills contribute to learning about multi-word sequences in natural language processing. Future work is necessary to explore

precisely how and when in development various cognitive skills may interact with processing multi-word chunks in sentences.

In sum, this project has demonstrated that young children can rapidly generate predictions during language processing based on higher-order multi-word chunks in sentences, although, this process is (at least) limited to conditions where there are multiple semantically informative units in the linguistic stream. Our findings lend support to recent accounts of flexible linguistic processing, and suggest that even very young, school-age children are highly sophisticated in their ability to rapidly compute and generate predictions for higher-order relations among multi-word sequences during linguistic comprehension.

5. References

- Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*(1), 143–148.
- Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250.
- Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*, *9*(3), 621–636.
- Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth - children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129.
- Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371.
- Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248.
- Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, *143*(1), 404–413.

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Boersma, P., & Weenink, D. (2012). *Praat: doing phonetics by computer* (5.3.22). 5.3.22.
- Borovsky, A, Ellis, E. M., Evans, J. L., & Elman, J. L. (n.d.). Lexical leverage: Category knowledge boosts real-time novel word recognition in two-year-olds. *Developmental Science*.
- Borovsky, Arielle, & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology*, 50(5), 1600–1613.
- Borovsky, Arielle, Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, 19(6), 918–932.
- Borovsky, Arielle, Elman, J. L., & Fernald, A. (2012). Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112(4), 417–436.
- Borovsky, Arielle, Sweeney, K., Elman, J. L., & Fernald, A. (2014). Real-time interpretation of novel events across childhood. *Journal of Memory and Language*, 73(1), 1–14.
- Bybee, J. (2006). From usage to grammar: The mind’s response to repetition. *Language*, 82(4), 711–733.
- Bybee, Joan. (1998). The emergent lexicon. *Chicago Linguistic Society*, 34, 421–435.
- Creel, S. C., & Quam, C. (2015). Apples and Oranges: Developmental Discontinuities in Spoken-Language Processing? *Trends in Cognitive Sciences*, Vol. 19, pp. 713–716.
- de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, 19(2), 235–250.

- de Carvalho, A., Dautriche, I., Lin, I., & Christophe, A. (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, *163*, 67–79.
- Ellis, E. M., Gonzalez, M. R., & Deák, G. O. (2014). Visual prediction in infancy: What is the association with later vocabulary? *Language Learning and Development*, *10*(1), 36–50.
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, Second Edition*. Baltimore, MD: Brookes Publishing.
- Fernald, A, Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental Psycholinguistics: On-line methods in children's language processing*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Fernald, Anne, Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective-noun phrases. *Cognitive Psychology*, *60*(3), 190–217.
- Gambi, C., Gorrie, F., Pickering, M. J., & Rabagliati, H. (2018). The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds. *Journal of Experimental Child Psychology*, *173*, 351–370.
- Gambi, C., Jindal, P., Sharpe, S., Pickering, M., & Rabagliati, H. (2020). The relation between preschoolers' vocabulary growth and their ability to predict and recognize words. *Child Development*, *00*(0), 1–19.
- Gambi, C., Pickering, M. J., & Rabagliati, H. (2016). Beyond associations: Sensitivity to

- structure in pre-schoolers' linguistic predictions. *Cognition*, 157, 340–351.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology*, 48(12), 1726–1737.
- Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, 28(4), 377–396.
- Hernández, M., Costa, A., & Arnon, I. (2016). More than words: Multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, 31(6), 785–800.
- Huang, Y. T., & Snedeker, J. (2008). Use of referential context in children's language processing. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, (May), 1380–1385.
- Huang, Y. T., & Snedeker, J. (2013). The use of lexical and referential cues in children's online interpretation of adjectives. *Developmental Psychology*, 49(6), 1090–1102.
- Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, 69(4), 589–606.
- Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118–135.
- Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, 31(1), 80–93.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). The processing of multiword expressions

- in children and adults: An eye-tracking study of Chinese. *Applied Psycholinguistics*, 41(4), 901–931.
- Jones, G., Cabiddu, F., & Avila-Varela, D. S. (2020). Two-year-old children's processing of two-word sequences occurring 19 or more times per million and their influence on subsequent word learning. *Journal of Experimental Child Psychology*, 199, 104922.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156.
- Kidd, E., & Bavin, E. L. (2005). Lexical and referential cues to sentence interpretation: An investigation of children's interpretations of ambiguous sentences. *Journal of Child Language*, 32(4), 855–876.
- Kukona, A., Braze, D., Johns, C. L., Mencl, W. E., Van Dyke, J. A., Magnuson, J. S., ... Tabor, W. (2016). The real-time prediction and inhibition of linguistic outcomes: Effects of language and literacy skill. *Acta Psychologica*, 171, 72–84.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Lew-Williams, C., & Fernald, A. (2007). Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447–464.
- Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two- and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, 146, 349–370.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (third edition):

- Volume I: Transcription format and programs, Volume II: The database. *Computational Linguistics*, Vol. 26, pp. 657–657.
- Mani, N., Daum, M. M., & Huettig, F. (2016). “Proactive” in many ways: Developmental evidence for a dynamic pluralistic approach to prediction. *The Quarterly Journal of Experimental Psychology*, 69(11), 2189–2201.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake - But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843–847.
- Mani, N., & Huettig, F. (2013). Towards a complete multiple-mechanism account of predictive language processing [Commentary on Pickering & Garrod]. *Behavioral and Brain Sciences*, 36, 365–366.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, 53(4), 372–380.
- McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, 126(1), 1–51.
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14(8), 348–356.
- Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science*, 2(1), 138–153.
- Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and

- review. *Psychological Bulletin*, *144*(10), 1002–1044.
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347.
- Rabagliati, H., & Robertson, A. (2017). How do children learn to avoid referential ambiguity? Insights from eye-tracking. *Journal of Memory and Language*, *94*, 15–27.
- Real, F., & Christiansen, M. H. (2007). Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology*, *60*(2), 161–170.
- Reuter, T., Dalawella, K., & Lew-Williams, C. (2021). Adults and children predict in complex and variable referential contexts. *Language, Cognition and Neuroscience*, *36*(4), 474–490.
- Reuter, T., Emberson, L., Romberg, A., & Lew-Williams, C. (2018). Individual differences in nonverbal prediction and vocabulary size in infancy. *Cognition*, *176*, 215–219.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase “Time and Again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 776–784.
- Skarabela, B., Ota, M., O’Connor, R., & Aron, I. (2021). ‘Clap your hands’ or ‘take your hands’? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, *211*(January), 104612.
- Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299.
- Swingle, D. (2011). The Looking-While-Listening Procedure. In *Research Methods in Child*

Language (pp. 29–42). Wiley-Blackwell.

- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613.
- Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134.
- Van Heugten, M., & Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Developmental Science*, *12*(3), 419–425.
- Von Holzen, K., & Mani, N. (2012). Language nonselective lexical access in bilingual toddlers. *J Exp Child Psychol*, *113*(4), 569–586.
- Woodard, K., Pozzan, L., & Trueswell, J. C. (2016). Taking your own path: Individual differences in executive function and language processing skills in child learners. *Journal of Experimental Child Psychology*, *141*, 187–209.
- Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*, *133*(1), 262–276.

6. Appendices

6.1. Appendix A

List of all Sentences and Their Frequency Information for all Conditions in Experiment 1

Condition	Frame Frequency	Object Frequency	Verb-Object Frequency
FF-High			
Brush your teeth	18.50	236.83	36.50
Close your eyes	19.00	293.50	26.17
Drink of water	10.83	950.33	21.50
Open the door	42.17	429.83	50.83
Throw the ball	21.50	809.00	24.33
Want your bottle	6.17	226.33	12.83
Wash your hands	41.33	390.17	66.83
FF-Low			
Brush your hair	5.50	491.67	24.33
Close your mouth	2.17	461.67	2.83
Drink of milk	4.50	581.67	19.67
Open the window	1.17	164.50	4.33
Throw the book	0.50	1285.67	1.50
Want your book	2.17	1285.67	20.17
Wash your hair	5.50	491.67	11.50
MF-High			
Brush her teeth	0.50	236.83	36.50
Close his eyes	0.17	293.50	26.17
Drink some water	2.50	950.33	21.50
Open a door	0.83	429.83	50.83
Throw your ball	0.17	809.00	24.33
Want a bottle	3.00	226.33	12.83
Wash her hands	2.00	390.17	66.83
MF-Low			
Brush her hair	0.50	491.67	24.33
Close his mouth	0.50	461.67	2.83
Drink some milk	3.00	581.67	19.67
Open a window	0.83	164.50	4.33
Throw your book	0.17	1285.67	1.50
Want a book	3.33	1285.67	20.17
Wash her hair	2.00	491.67	11.50

Note: These numbers represent the frequency of each item per 1 million words calculated from a corpus of ~6 million words.

6.2. Appendix B

Outline of the Agent, Action, and Thematic Object Sets Used in Experiment 2

Set	Agents	Action1	Object 1	Action2	Object 2
Set 1	Monkey, Dog	Eats	Candy, Apple	Rides in	Car, Bus
Set 2	Mouse, Cat	Tastes	Ice cream, Cake	Wears	Hat, Glasses
Set 3	Turtle, Frog	Turns on	TV, Computer	Cuts	Paper, Bread
Set 4	Lion, Duck	Flies	Kite, Airplane	Sits on	Rock, Fence

Note. In each story set, each agent performed both actions, and different agents were associated with different objects (e.g. if the monkey eats the candy, then the dog eats the apple). In each set, each verb was additionally paired with one of the potential two objects in a frame that did not mention either of the agents (e.g. Let's eat the apple). All possible combinations of Agents, Actions and Objects occurred across lists in the study.

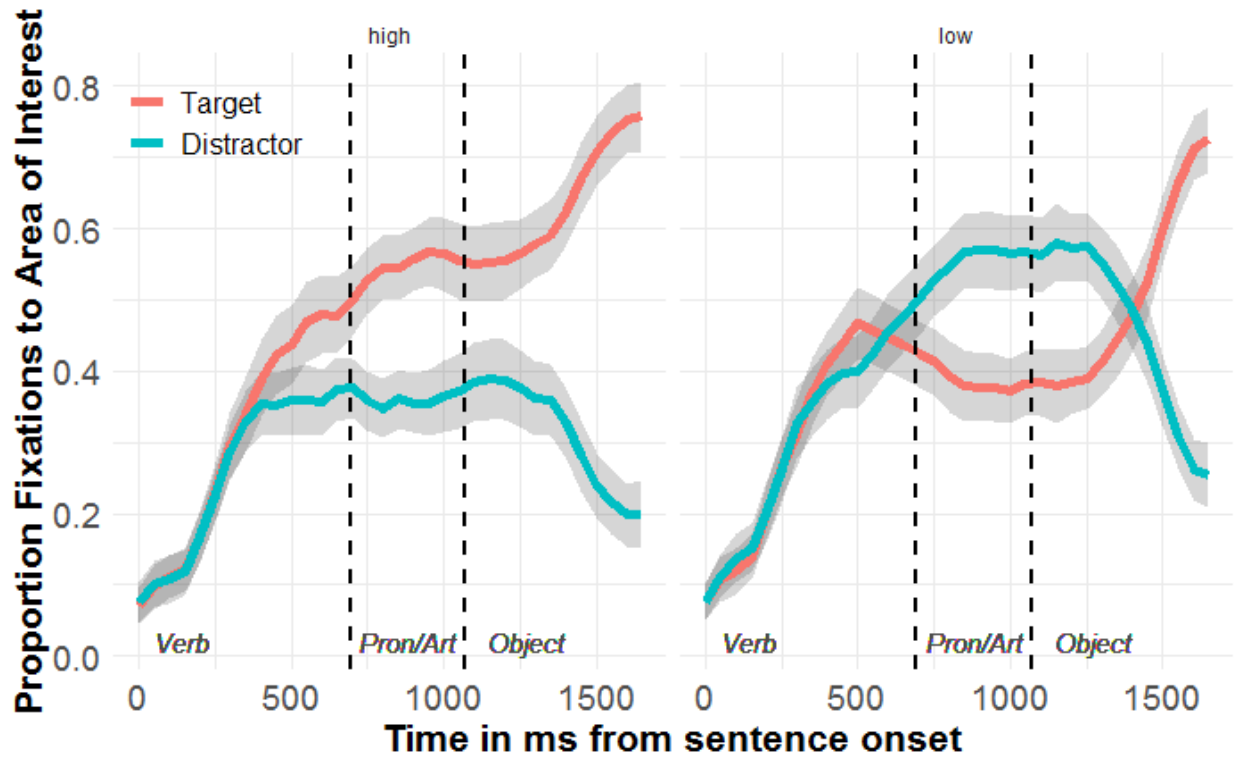
7. Supplementary Materials

7.1. Linear mixed effects regression model for Experiment 1 including random effects of item.

Effect/Slope	Estimate	CI	t-value	p-value
(Intercept)	0.496	0.461 – 0.531	27.672	<.001
Frame Type	-0.006	-0.077 – 0.064	-0.179	.861
Frame Frequency	0.122	0.072 – 0.172	4.778	<.001
Type x Frequency	-0.056	-0.156 – 0.043	-1.110	.268
Age	0.004	-0.020 – 0.029	0.342	.733
Trial Order	-0.031	-0.059 – -0.003	-2.202	.029
Random Effects				
σ^2	0.13			
τ_{00} Items	0.002			
ICC	0.02			
N_{Items}	14			
Observations	829			
Marginal R^2 / Conditional R^2	0.031 / 0.048			

7.2. Figure 1

Time-Course of Proportion Fixations to Target and Distractor Images for High- (Frequent and Modified) and Low-Frequency (Frequent and Modified) Frames



Note. Error cloud represents CIs. Dashed lines mark word onsets.

7.3. Table 1*Linear Mixed Effects Regression Model for Experiment 1 after removing all “drink” trials*

Effect/Slope	Estimate	CI	t-value	p-value
(Intercept)	0.511	0.478 – 0.544	30.082	<0.001
Frame Type	-0.006	-0.072 – 0.061	-0.167	0.871
Frame Frequency	0.142	0.090 – 0.195	5.293	<0.001
Type x Frequency	-0.069	-0.176 – 0.038	-1.271	0.204
Age	0.001	-0.026– 0.027	0.056	0.955
Trial Order	-0.024	-0.053 – 0.005	-1.645	0.102

Random Effects

σ^2	0.13
$\tau_{00 \text{ Item}}$	0.001
N_{Item}	12
Observations	718
Marginal R^2 /	0.042 /
Conditional R^2	0.052

7.4. Expanded version of Table 4

Linear Mixed Effects Regression Model for Experiment 2

Effect/Slope	Estimate	CI	t-value	p-value
(Intercept)	0.25	0.23 - 0.27	26.7	<.0001
Area of Interest:				
Target	0.152	0.12 - 0.18	9.38	<.0001
Action Related	0.013	-0.02-0.05	0.804	.42
Agent Related	-0.061	-0.09- -0.03	-3.74	.0002
Age	-1.44e-17	-0.001 – 0.001	0	>.9999
Trial Index	-1.24e-18	-0.008 – 0.008	0	>.9999
Random Effects				
σ^2	0.08			
τ_{00} Item	0.00			
τ_{00} Subject	0.00			
N Item	32			
N Subject	32			
Observations	920			
Marginal R ²	0.105			