2.2 Methods

2.2.1 Participants

47 younger adults and 41older adults participated in the study.[1] Three younger adult participants were excluded due to incomplete data, four younger adult participants were excluded for proficiency in languages other than English, and one participant was excluded for not taking the tasks seriously. Therefore, 80 participants—39 younger adults (age range = 18 – 23 years, *M* = 19.18 years, *SD* = 1.25, females = 30) and 41 older adults (age range = 58 – 79 years, *M* = 65.34 years, *SD* = 4.80, females = 19)—were analyzed in the present study. The younger adult participants were recruited from the Penn State Psychology Department Subject Pool and the older adult participants were recruited from https://www.prolific.co/. A detailed characterization of the participants is provided in Table 3. All participants were Native English speakers with normal or corrected-to-normal vision with no self-reported neurological or psychological conditions. 66 participants were right-handed, as indicated by the Edinburgh Handedness Inventory (Oldfield, 1971).[2] All participants provided informed consent and all experimental procedures were approved by the Institutional Review Board of The Pennsylvania State University.

---

[1] Only participants in which a complete dataset was obtained are reported. During online data collection, many participants experienced technical issues due to the resource demands of the self-paced task and terminated their participation and withdrew their consent to participate.

[2] While language studies typically recruit only right-handed individuals due to evidence suggesting differences in hemispheric lateralization in language processing in left- and right-handed individuals, because this was a behavioral study without any neural measures, both left- and right-handed individuals were recruited. Any future studies that use neural measures will only recruit right-handed individuals.

Table 3. Neuropsychological and Language Production Information for Study 1

Neuropsychological Information	**Younger**	**Older**
	Mean (SD)	**Mean (SD)**
Age (years)***	19.18 (1.25)	65.34 (4.80)
Education (years)***	13.46 (1.25)	16.35 (3.39)
Language Production		
Picture Naming Accuracy	0.93 (0.04)	0.92 (0.05)
Picture Naming RT (ms)***	1631.18 (649.89)	2499.22 (892.65)
Total Verbal Fluency	92.11 (22.72)	85.42 (27.32)
Categorical (Animals and Groceries)	51.19 (15.66)	46.80 (15.15)
Phonemic (Letters F and A)	40.92 (10.62)	38.63 (14.09)
Elicitation Task MLU (words)	17.33 (5.51)	15.73 (5.21)
Elicitation Task %Adjectives	4.68 (2.23)	4.59 (2.94)
Elicitation Task %Conjunctions**	2.20 (1.22)	1.45 (1.25)
Working Memory		
Forward Digit Span	13.33 (3.76)	14.90 (4.05)
Backward Digit Span	10.71 (3.89)	10.93 (4.27)
Reading Span*	0.66 (0.24)	0.78 (0.16)
Processing Speed		
Simple Speed (ms)	340.08 (59.12)	354.66 (63.40)
Complex Speed (ms)***	509.44 (87.28)	726.12 (315.55)
Reading Experience		
Author Recognition***	6.69 (3.58)	33.20 (15.61)
Vocabulary***	46.15 (6.55)	55.44 (10.34)

Note. *$p < .05$, ** $p < .01$, ***$p < .001$

2.2.2 Neuropsychological Measures

All participants completed a neuropsychological test battery to assess working memory, processing speed, and reading ability, as these factors are likely to moderate language prediction and production abilities. All working memory tasks were programmed in PsychoPy Version 2020.2.3 and the processing speed tasks were programmed in PsychoPy Version 3.2.4 (Peirce et al., 2019). The working memory assessment included measures of verbal working memory: forward and backward digit span tests and a reading span test that was a modified version of the Pitt Reading Span Test (Conway et al., 2005; Loboda, 2021). Processing speed was assessed with both a simple reaction time task, in which participants pressed the spacebar as soon as a stimulus appeared, as well as a choice reaction time task, in which participants saw an “X” in one of four positions on the screen and pressed the button corresponding to that position as quickly and accurately as possible. Participants’ reading experience was assessed using the author recognition task (Acheson, Wells, & MacDonald, 2008) and the WAIS-III Vocabulary Subtest (Wechsler, 1997). The reading experience tasks were created using Penn State Qualtrics ("Qualtrics," 2020).

Working Memory Tasks

Participants completed three verbal working memory tasks—forward and backward digit span, as well as a reading span task. During the digit span tasks, participants saw a sequence of numbers appear on the screen, one number at a time. After the number sequence was presented, participants were asked to recall all the numbers in the sequence in the same order (forward) or reverse order (backward) as they were initially presented. Each number in a given sequence was presented for one second. The number sequences varied in length from two to nine numbers for the forward digit span task and from two to eight numbers for the backward digit span task. Participants saw three of each number sequence (e.g., 3 times 2-number sequence, 3 times 3-

number sequence, etc.). All number sequences were presented sequentially so that after every 3 sequences, the difficulty increased by adding one additional number for participants to remember. During scoring, participants received one point for each correct response up until they received at least 2/3 of responses incorrect within a span. Correct responses after this point were not counted. This was done to replicate the scoring procedure in the lab environment, in which we stop administering the task once participants get two sequences incorrect within a span.

The reading span task was a modified version of the Pitt Reading Span Test (https://ubiq-x.gitlab.io/rspan/; Loboda, 2021). During this task, participants performed two interleaved tasks—sentence judgements and remembering letters. During the sentence judgements, participants were asked to judge whether a particular sentence made sense based on the content of that sentence, indicating via a button press whether they thought the sentence was correct (e.g., John was asked to sit on a chair.) or incorrect (e.g., John was asked to sit on a marmalade.). Participants were informed that they had 10 seconds to respond during the sentence judgement portion of the trial. In between making sentence judgements, participants saw one letter that they had to remember. Each individual letter was presented for one second. At the end of the trial, participants were asked to recall the letters in the same order they were presented. This was not timed. The number of sentences/letters participants saw in a trial varied from two to seven. There were three of each sequence and the order was randomized, totaling in 18 trials. For example, in a three sentences/three letters trial, the trial proceeded as follows: *sentence judgement → letter → sentence judgement → letter → sentence judgement → letter → recall all letters*. A key of all possible letters participants were asked to remember was available on the recall screen to mimic the Pitt Reading Span Test. During letter recall, if participants could not remember a particular letter for a particular position, they were instructed to enter the number zero. For example, if the letters to recall were “lqr” but a participant could not remember the letter in the second position, they would enter their response as “l0r”.

Scoring was done according to the partial-credit load scoring procedure described in Conway et al. (2005). As a brief explanation, when participants recalled the letters they were instructed to remember, each letter was worth one point. The number of points earned (i.e., letters correctly recalled) was divided by the total number of letters across trials, which was 81 letters in the version used in this study. For a letter to be considered correctly recalled, it had to be entered as the correct letter in the correct position within the sequence. Using the above example, if a participant recalled “lqr”, then three points (the maximum for this trial) would have been earned, because all letters were correctly recalled. If they entered l0r, then two points would have been earned because the two letters recalled were the correct letters in the correct positions. Lastly, if they had entered lr, then one point would have been earned because although there was an “r” in the sequence, it was incorrectly entered as the second letter in the sequence, when it was actually the third letter in the sequence.

Processing Speed Tasks

To measure processing speed, participants performed simple and choice reaction time tasks. During the simple reaction time task, participants saw the letter ‘X’ appear in a white box on the computer screen. They were instructed to press the spacebar as quickly and accurately as possible as soon as the ‘X’ appeared. There were 80 trials with 10, fully randomized interstimulus interval (isi) values ranging from 1.0000 seconds to 2.9998 seconds, with each isi value being used eight times. Response times for each trial were recorded. During the choice reaction time task, participants saw four white squares in a single row in the center of the computer screen. Participants were instructed to press the ‘c’, ‘v’, ‘b’, or ‘n’ key on their keyboard when the letter ‘X’ appeared in the corresponding white box. For example, if the ‘X’ appeared in the second box from the left, participants had to press the ‘v’ key as quickly and accurately as possible to respond correctly. There were 80 trials total and the same 10 isi values that were used in the simple reaction time task were used in the choice reaction time task. Each isi was paired with each of the four response keys twice, with the order being random. Although reaction time tasks do not typically have a practice portion to minimize practice effects, I included a brief practice portion for both the simple and choice reaction time tasks to ensure participants understood and performed the task correctly due to data collection occurring online without an experimenter present to answer questions.

For both the simple and choice reaction time tasks, reaction time values less than 200 ms, as well as reaction time values 2.5 standard deviations above a participant’s mean were removed as outliers. Additionally, only accurate trials were included in calculating the average reaction time in both tasks for each participant. Therefore, the reported reaction time values for the simple and choice reaction time tasks are for accurate trials after the removal of outliers.

Reading Experience Tasks

Participants completed two untimed tasks to assess their reading experience—the author recognition task (Acheson et al., 2008) and a modified WAIS-III Vocabulary Subtest (Wechsler, 1997). During the author recognition task, participants were given a list of names in which some names were real authors (e.g., Stephen King, J.R.R. Tolkien) and other names were foils. Participants were instructed to select all the names they knew were real authors and were informed there was a penalty for guessing. Total score was calculated as the number correct minus the number incorrect. During the WAIS-III Vocabulary Subtest (hereafter referred to as the WAIS vocabulary test), participants were given 21 words to define, with words becoming progressively more difficult to define (e.g., lower frequency words, abstract words). Participants were instructed to define the word to the best of their ability without using search engines for their answers. They were also instructed that they did not have to respond in complete sentences and that if a word had multiple definitions (e.g., the word 'ship' can refer to the act of shipping packages or a vessel that floats on water), any definition was considered correct. If participants did not know a definition, they could skip it and write "NA." The first 12 words from the WAIS vocabulary test were not given to save time as it was assumed participants knew the definitions to these words. Participants started with word 13 from the WAIS vocabulary test. Each definition was worth a maximum of two points (it was possible to receive 0, 1, or 2 points) and scoring was done according to the WAIS-III Vocabulary Scoring Guide. All participants received full credit (24 points) for the first 12 words that were excluded from the task and could receive a maximum of 66 total points.

2.2.3 Language Production Measures

To examine the relationship between language prediction and language production, participants completed several off-line production measures. All language production tasks were programmed in PsychoPy Version 2020.2.3 and were conducted online through Pavlovia (https://pavlovia.org/). Because all experimental tasks were conducted online, typing was used in place of overt speech production. Language production assessments included a typed verbal fluency task, typed picture naming task, and a typed elicitation task, in which participants responded freely to an open-ended prompt.

Verbal Fluency Task

The verbal fluency task included two categorical (animals and grocery items) and two phonemic (letters "F" and "A") measures. During the task, participants were instructed to type as many words as they could think of that belong to the specified category for the categorical condition and as many words that begin with the specified letter for the phonemic condition (Ardila, Ostrosky-Solís, & Bernal, 2006). During standard verbal fluency tasks, participants are given one minute to respond per measure; however, for online data collection, participants were given two minutes to respond per measure to account for variability in typing speed. Participants completed the two categorical measures before the two phonemic measures and were informed of the category or letter at the start of the two-minute response period. As such, participants were not given additional time to think of their responses outside of the two-minute response period. For the categorical task, participants were instructed to list all the items they could think of that belonged to a specified category. They were also instructed not to repeat the same item with different endings (e.g., elephant, elephants) and to avoid listing varieties of the same item (e.g., green pepper, red pepper). For the phonemic fluency tasks, participants were instructed not to repeat words, say the same word with different endings (e.g., faint, faints, fainted), nor say proper names of people (e.g., Francis) and places (e.g., France).

During scoring, repetitions were only counted once, as were the same words with different endings. Proper nouns were discarded and if superordinate categories were followed by subordinate items, the superordinate category was not counted, but the individual items were. For example, if a participant listed “bird” followed by “robin,” “crow,” and “owl,” then “bird” was discarded but the three specific responses were each counted. After coding the verbal fluency responses, items were separated into total verbal fluency (collapsed across all four fluency measures), total categorical (collapsed across the two categorical measures) and total phonemic (collapsed across the two phonemic measures).

Picture Naming Task

For the picture naming task, participants saw one of 80 color photographs presented on a white background per trial. All pictures were sized to a height of 396 pixels and aspect ratio was preserved. All pictures used in this task were taken from normed databases (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Brodeur, Guérard, & Bouras, 2014; McRae, Cree, Seidenberg, & McNorgan, 2005; Moreno-Martínez & Montoro, 2012) and the task itself is an abbreviated version of the picture naming task conducted in Diaz et al. (2020). A subset of pictures from the original task were used, and during picture selection, any object names that overlapped with the self-paced reading task, specifically the predictable verbs (e.g., the picture of a rake was removed because “rake” was used as a verb in the predictable condition) or target nouns, were removed. Additionally, to ensure there was enough variability in naming difficulty, a mix of low (range = 4.49 – 6.88; mean = 6.01, *SD* = 0.61; 28 total items), medium (range = 6.93 – 8.59; mean = 7.80, *SD* = 0.54; 27 total items), and high frequency (range = 8.93 – 11.01; mean = 9.89, *SD* = 0.68; 25 total items) picture names were selected based on the logged HAL frequency ranges described in Gertel et al. (2020). Except as a metric for picture selection, lexical frequency during the picture naming task was not further examined in this dissertation.

During the picture naming task, participants were instructed to type the name of the object depicted as quickly and as accurately as possible and press the enter key once they were finished typing their response. Participants were also able to use the backspace key to correct any typing errors; however, participants were instructed not to worry about exact spellings of words. This allowed greater flexibility in acceptable responses. For example, answers that would have been correct if overt articulation could be used but were misspelled because of typing disfluencies or because participants knew the item name but not the orthographic spelling of the item name (e.g., if a participant typed “cantalope” instead of “cantaloupe”) were considered correct. Accepted misspellings were consistent across participants; however, it is important to note that an accepted misspellings rubric was not created due to the high amount of variability in possible misspellings. Participants were also instructed to be as specific as possible when typing the name of the pictures, for example, if they saw a picture of a "robin" to type "robin" instead of "bird;" however, if participants were unsure of the picture name, they were instructed to make their best guess. Please refer to Appendix A for a rubric of acceptable responses. Due to the nature of online data collection, acceptable alternative responses were also coded as correct to account for any regionalisms (e.g., bucket vs. pail).

In addition to coding for accuracy, the time of the first letter-key pressed was recorded to calculate reaction time and was used as a proxy for the onset of articulation. Using the time of the first letter-key pressed helped ensure that the reaction times were not confounded by typing speed and word length and were more consistent with how response times are recorded during overt picture naming, which are based on the time that the first sound is produced. All response times reported are for accurate trials only, with outliers excluded. Outlier trials consisted of any responses less than 250 ms and that were more than 2.5 standard deviations from the mean for each participant.

Typed Elicitation Task

The speech elicitation task allowed for naturalistic language production and contrasts the verbal fluency and picture naming tasks, which are structured language production measures. During this task, participants were asked to type freely in response to the following prompt: "Please describe one of your favorite childhood memories." Using an autobiographical prompt serves to encourage older adults' production because these memories are more entrenched for older adults and will ensure the older adults are not biased against (Spreng et al., 2018; Spreng &

Turner, 2019). During the task, participants were informed that they would have five minutes to type freely in response to a prompt and that their response would be visible as they typed. Additionally, participants were informed that the backspace key would be disabled for the duration of the task. This served to prevent participants from going back and changing their response, simulating naturalistic speech, which is inherently disfluent. Participants were aware that the goal of this task was not to have a polished piece of writing without mistakes and that I was interested in their free-flowing thoughts. On the following screen, participants saw the prompt and were instructed to think about and plan their upcoming response. This screen progressed automatically after 28 seconds, allowing participants time to read the onscreen instructions, as well as time to think about their response. They were instructed to begin typing as soon as the next screen appeared. During the timed typing portion, participants had a countdown timer visible to them and at the end of the five minutes, the end of experiment screen automatically appeared, and their responses were recorded.

Typed responses were input into the CLAN software package to ascertain measures of syntactic complexity. Prior to analyzing the data using CLAN, minor typos were corrected. If the participant's intent was unclear, then the word was left as is. Additionally, in instances when participants did not use punctuation, it was added at the boundary of two distinct clauses where it was grammatically appropriate to ensure the software package processed the data correctly. Otherwise, grammar was not corrected. For the typed task, the mean length of utterance (MLU) was calculated to examine the volume of output, with utterances being defined according to participants' use of punctuation. I also examined the total number of coordinate phrases and the total number of complex nominals for each participant's typed response (Lu, 2011). Coordinate phrases were identified by their coordinating conjunctions (for, and, nor, but, or, yet, so), which are words that link clauses or sentences. Complex nominals are expressions in which the head noun in a noun phrase is preceded by a modifying noun or adjective (e.g., wind turbine; Newmeyer, 1979). To determine these two syntactic complexity measures from the speech elicitation task, each typed response was coded by hand to track the number of coordinate conjunctions used and the number of complex nominal phrases used. Additionally, the percentage of total conjunctions and adjectives per typed response were also ascertained using the CLAN EVAL function (operationalized as total number of conjunctions or adjectives/total number of words).[3]

2.2.4 Self-Paced Reading Materials

107 pairs of predictable and non-predictable sentences were created in which predictability of the sentence was manipulated by including a constraining or non-constraining verb (see Table 4 below for example sentences, Appendix B for all sentence pairs and Appendix C for verb lexical properties). The constraining verbs (mean length = 5.66 letters, range = 3 – 9 letters; mean logged HAL frequency = 7.59, range = 2.30 – 12.15) only apply to a limited number of potential target nouns, which encourages predictive processing, whereas the non-constraining verbs (mean length = 5.72 letters, range = 3 – 10 letters; mean logged HAL frequency = 9.87, range = 5.52 – 14.92) apply to numerous potential target nouns, which limits the ability to predict. Comparing the predictable and non-predictable verbs, there was no significant difference in the number of letters, $t(195.44) = -0.30$, 95% CI [-0.50, 0.37], $p = .77$; however, there was a significant difference in logged HAL frequencies $t(209.23) = -8.40$, 95% CI [-2.83, -1.75], $p < .001$; with the non-predictable verbs being more frequent. This is consistent

[3] Because the hand-coded coordinate conjunctions and complex nominals were similar to the percent of conjunctions and percent of adjectives calculated using the CLAN software, I decided to use the more objective measure calculated using CLAN for subsequent analyses. The values from CLAN are reported in all Neuropsych scoring tables.

with prior studies (Hintz et al., 2017) in which the non-predictable condition verbs were also more frequent. The non-predictable verbs are less constraining, meaning they are applicable to more nouns and are more commonly used (i.e., more frequent). The significant difference in word frequency was not a concern for interpretation of any results because having the higher frequency verbs in the non-predictable condition biases against observing results in predictability. Therefore, any results in which the reading times for predictable sentences are faster than the reading times for non-predictable sentences further supports that predictive processing was likely engaged since the non-predictable verbs should be read faster if results were due to frequency effects.

When selecting the predictable and non-predictable verb pairs, the verb classes were considered to ensure that the predictable verbs were more constraining in their use (e.g., verbs associated with a specific action, such as the verb "peel"). All verb classes were determined in accordance to their listing in the book, *English Verb Classes and Alternations* and can be found in Appendix D (Levin, 1993). Additionally, to ensure that the verb–noun pairings are used together in naturalistic language, the Online OXFORD Collocation Dictionary of English (https://www.freecollocation.com) was used. The collocation dictionary is based on the 100 million word British National Corpus (Leech, 1992) and includes over 150,000 collocations for about 9,000 headwords. Sentence pairs were included so long as the predictable verb and target noun were listed together as a common word combination and the non-predictable verb and target noun were not listed as a common word combination. For cases in which the noun was not listed in the dictionary, the norming data conducted by the lab was used to ensure the predictable verb and target noun are commonly associated. In all sentences, the verb that precedes the target noun was the only difference between the sentence pairs and all predictable verbs were used only once. Additionally, adjectives were added between the verb and target noun to allow enough time for prediction to occur. All sentences continue after the target noun to ensure prediction was calculated based on the critical region (target noun + 1 word), which accounts for potential spillover effects of predictive processing (Hintz, Meyer, & Huettig, 2015). Sentence subjects and overall sentence structure varied across sentences to ensure effects were not a result of a specific sentence structure. Seventy-five filler sentences were also used, with the goal of preventing participants from adapting to the study and becoming aware of the predictable–non-predictable sentence manipulation. After 20% of trials, participants answered yes/no comprehension questions to ensure they were attending to the task and not simply pressing the spacebar without reading the sentences.

Table 4. Example Predictable and Non-Predictable Sentences

Predictable Sentences	**Non-Predictable Sentences**
The boy peeled the large banana while he was sitting in the cafeteria.	The boy drew the large banana while he was sitting in the cafeteria.
Kate earns plenty of money from babysitting the neighborhood children.	Kate has plenty of money from babysitting the neighborhood children.
Todd and Sara rake brightly colored leaves into a pile as they fall in autumn.	Todd and Sara bring brightly colored leaves into a pile as they fall in autumn.

Note. The critical verb is presented in green text and the target noun is presented in purple text.

Sentence Norming

Two separate sentence lists were created for sentence norming in which participants saw one version of each sentence pair—either the predictable or non-predictable sentence. Sentence pairs were split to prevent participants from seeing the same sentence construction twice, in which the only difference was the verb. Eighty Penn State undergraduate students from the Psychology Department subject pool completed sentence norming in person while 44 participants completed sentence norming online and were recruited using Prolific. All in-person sentence norming was done using E-Prime and all online sentence norming was done using Qualtrics. During the norming task, participants were provided with the beginning of the sentence and were asked to fill in the missing target noun to complete the sentence. Participants were told there were no right or wrong answers and to limit responses to one or two words. This instruction was given because no target nouns were more than two words in length, and this also prevented participants from completing the sentences with longer phrases (e.g., responding "the cute, brown puppy," when "puppy" or "dog" sufficed). Sentence completion results were used to calculate cloze probability to determine if the sentence pairs were both predictable and not predictable. Predictable sentences were retained if the cloze probability was at least 0.5 (See Appendix E for cloze probabilities and accompanying statistics for both predictable and non-predictable items). A range of cloze probabilities (0.5 – 1.0) was used to avoid ceiling effects (Brothers et al., 2015; Hintz et al., 2017). Brothers et al. (2015) found significant prediction effects with predictable sentences having a minimum cloze probability of 0.5, suggesting that this manipulation should elicit the desired effects. Additionally, in a visual world paradigm study, Hintz, Meyer, and Huettig (2017) found significant effects while including predictable sentences with cloze probabilities below 0.5 (range: 0.06 – 0.8, average: 0.39). In the current study, when calculating cloze probabilities, an answer was considered "correct" if the participant’s response could reasonably complete the sentence with the constraining verb that was used. This was done because contextual sentences were not provided before the sentences to be completed, which would narrow the possible responses. For example, in the sentence “The boy peeled the large ____”, answers such as “banana”, “apple”, “orange”, or “potato” can all reasonably complete the sentence and were considered “correct” for calculating cloze probabilities.

2.2.5 Self-Paced Reading Task Procedure

The self-paced reading task was programmed in PsychoPy Version 3.2.4 and was run online using Pavlovia.org. Participants completed one of two counterbalance lists (details discussed below) to ensure participants saw only one version of each sentence pair. During the self-paced task, participants were instructed that they would be reading sentences on a screen, one word at a time (see Figure 2). Each word was presented in the center of the screen in white-colored text in Arial Font on a gray background. There were 182 total trials (107 predictable/non-predictable sentences + 75 filler sentences). Participants were also instructed that they would be reading these sentences at their own pace and must press the spacebar on their keyboard to reveal the next word in the sentence. Response times were recorded for each time the spacebar was pressed and were used during data analysis to determine reading times. Participants were informed that they would be answering yes/no comprehension questions after some sentences, but not all sentences. Importantly, participants were not explicitly told to try to predict the upcoming words in the sentence and were not told any explicit strategies to use during the task.

At the start of each trial, participants saw a white fixation cross in the center of the screen for 1.5 seconds, then the first word of the sentence would automatically appear. After the automatic presentation of the first word, all subsequent words were revealed using the spacebar. Comprehension questions appeared after 20% of trials, resulting in 38 total questions, and were evenly split amongst predictable, non-predictable, and filler sentences. Additionally, the comprehension questions were constructed in a manner so that they could be the same across lists and were applicable to both the predictable and non-predictable version of a sentence. During the comprehension questions, participants saw the entire question on the screen at one time (See Figure 3). They were instructed at the start of the study that if the answer to the question was yes, to press the letter “z” key on their keyboard and if the answer to the question was no, to press the letter “m” key. The words “YES” and “NO” appeared in the bottom left and bottom right corners of the screen, respectively, to remind participants which key to press. Lastly, there were an equal number of yes and no responses to all questions and these were further split to ensure there were an equal number of yes and no responses to questions in the predictable, non-predictable, and filler sentences.

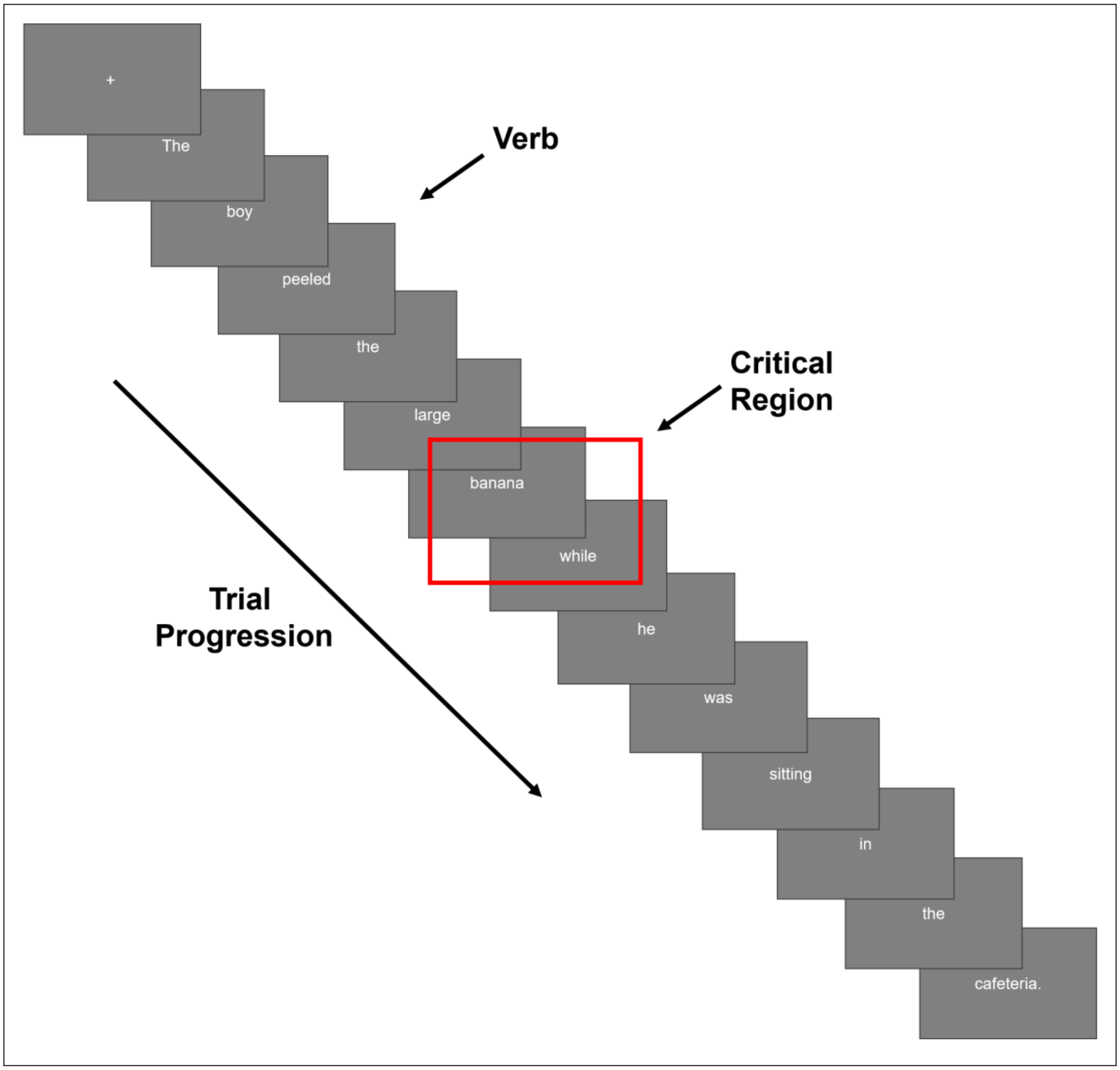

Figure 2. Study One Example Trial. Participants began each trial with a fixation cross followed by the first word of the sentence, which was presented automatically. To see each subsequent word in the sentence, the participant pressed the spacebar and reaction times for each spacebar press were recorded. This example shows the predictable version of the sentence. The non-predictable version of the sentence, “The boy drew the large banana while he was sitting in the cafeteria.” was presented in the same manner. The critical region is outlined in red.

Figure 3. Example Comprehension Question. After 20% of sentences, a comprehension question would follow. These 38 questions were consistent across conditions and served to ensure participants were attending to the task.

There were two different experimental lists so that participants only saw the predictable or non-predictable version of each sentence pair. Filler sentences were constant across lists. Additionally, the frequency and length of the predictable and non-predictable verbs was controlled across lists. Because questions followed only some sentences and were specific to a particular sentence, the order of the sentences within each list was pseudo-randomized to ensure that the questions appeared when appropriate; therefore, all participants saw the sentences in the same order. To achieve pseudo-randomization, the RAND() function in excel was used to assign each sentence a random number between 0 and 1. The excel sheet was then sorted by these random numbers from smallest to largest so that trials could then be presented sequentially within PsychoPy, which ensured the questions were correctly matched to the corresponding sentences. These same random numbers were used across lists to ensure trial order was consistent across both lists.

2.2.6 Study One Experimental Procedure

Study one was conducted on two separate days to prevent fatigue effects and maximize data quality. The entire experiment took approximately 2 – 2.5 hours and each experiment session took approximately one hour. On the first day, participants were provided with a study link to Penn State Qualtrics.[4] After giving informed consent, participants answered a series of survey forms asking questions about their demographics including their age, gender, years of education completed, first language, any other known languages, and ages those languages were acquired, race and ethnic background. Participants then completed the author recognition form and WAIS-III Vocabulary Subtest, followed by Ishihara plates to assess for potential red-green colorblindness. After the completion of the Qualtrics survey, participants were redirected to the self-paced reading and processing speed reaction time tasks that were hosted on Pavlovia.org. Participants first completed the self-paced reading task, then completed the simple reaction time task followed by the choice reaction time task. After the choice reaction time task, participants who were recruited from Prolific were provided with a completion code to receive payment and participants recruited through SONA Systems received course credit. After completion of the first session, participants were provided with an invitation to access session two. During session two, participants were directed to Pavlovia.org to complete the remaining tasks. Before proceeding with the tasks, all participants provided informed consent and were instructed to exit

[4] The procedure for day one was altered partway through data collection so that the self-paced task was completed before the Qualtrics survey. There was an issue redirecting from Qualtrics to Pavlovia, resulting in participants' screens freezing or the task not loading. While a potential confound, this was done in an effort to prevent data loss and preserve funds. The ordering of the tasks within Qualtrics and Pavlovia was not altered.

the tasks if they did not consent to participating. Task order proceeded as follows: forward digit span, backward digit span, typed elicitation task, typed picture naming task, reading span, and typed verbal fluency. Participants were able to take breaks between each task and were informed after the typed elicitation task that they reached a half-way point. Upon completion of session two, participants were again provided with a completion code to receive payment on Prolific or receive course credit through SONA Systems. Please note that this same study procedure was followed for studies two through four, with the only difference being the different manipulations in the self-paced reading paradigm.[5]

[5] There was also a change to the number of trials in the two reaction time tasks for studies 3 and 4. This will be explained in more detail in the method section of study 3.