**Supplemental Material S1.** Supplementary Experiments 1 and 2.

## Supplementary Experiment 1

### Method

### Linguistic parameters

***Length***. We analyzed the length of a sentence by counting the number of words and number of syllables.

***Syntactic complexity***. Syntactic complexity was defined by number of clauses and Yngve depth (Yngve, 1960). Linguistic trees were generated using the Stanford Core Natural Language Processing website (https://corenlp.run/; Manning et al., 2014). The number of clauses was calculated in two ways: 1) declarative clauses which was defined as the number of S nodes in the linguistic tree and 2) phrasal nodes (XPs) in the linguistic tree were also analyzed.

Another way to measure syntactic complexity is by using Yngve depth to evaluate the tree depth (reflecting the average number of embedded structures in a sentence). We computed the max and total Yngve depth of each sentence. Scores are assigned by giving a score of 0 to the rightmost branch under a given node and then increasing the score of each branch by 1 going from right to left. The total Yngve depth of each word is the sum of all the branches that connect that word to the root node. The max Yngve depth is the word in the sentence with the most depth (i.e., highest score) and the total Yngve depth is the sum of depth over all words. For example, the sentence, "In addition to touching the yellow circle, touch the black circle" has a max depth of 3 and total depth of 18 (Figure S1). Total Yngve depth was not reported in the main text because it is similar to our sentence length measure, but is provided in this online supplement.

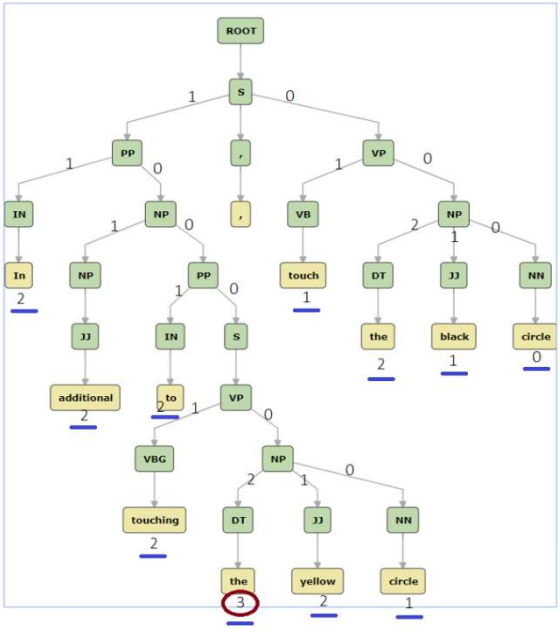


**Figure S1.** A linguistic tree illustrating the calculation of Yngve max depth (score circled in red = 3) and total depth (sum of all underlined scores = 18).

***Readability***. We used the "readability statistics" tool that is available in Microsoft Word to estimate the reading level for each part of the Modified Shortened Token Test. Note that participants are not reading sentences during the Token Test, instead, readability statistics were generated by computing a score for the reading level of the sentences typed into a document in Microsoft Word. Readability was measured by Flesch Reading Ease, the higher the score, the easier it is to understand, and Flesch-Kincaid Grade Level, which determines the minimum level of education required for the reader to understand the text.

## Results and Discussion

Descriptive statistics for the linguistic parameters calculated for each part of the Modified Shortened Token Test are presented in Table S1.

**Table S1.** Descriptive statistics for linguistic measures for each part of the Modified Shortened Token Test

| | **Part 1** | **Part 2** | **Part 3** |
|---|---|---|---|
| *Length:* | | | |
| Word length | 3.71 (0.49) | 4 (0) | 5 (0) |
| Syllable length | 4 (0.57) | 4.75 (0.5) | 5.75 (0.5) |
| *Syntax:* | | | |
| Declarative clause | 1 (0) | 1 (0) | 1 (0) |
| Phrasal nodes | 2 (0) | 2 (0) | 2 (0) |
| Max depth | 1.71 (0.49) | 2 (0) | 3 (0) |
| Total depth | 3.43 (0.98) | 4 (0) | 7 (0) |
| *Readability statistics:* | | | |
| Ease | 100 | 100 | 100 |
| Grade level | 0.0 | 0.0 | 0.0 |
| S*emantics:* | | | |
| Age of acquisition | 3.8 (0.95) | 3.89 (0.95) | 4 (0.96) |
| Concreteness | 3.86 (0.82) | 3.76 (0.96) | 3.67 (0.87) |
| Word frequency | 113 158 (93 811) | 134 270 (366 460) | 108 264 (327 806) |
| | **Part 4** | **Part 5** | **Part 6** |
| L*ength:* | | | |
| Word length | 8 (0) | 10 (0) | 9.31 (1.32) |
| Syllable length | 9.31 (0.63) | 11.23 (0.83) | 11.46 (2.70) |
| *Syntax:* | | | |
| Declarative clause | 1 (0) | 1 (0) | 1.31 (0.48) |
| Phrasal nodes | 4 (0) | 4 (0) | 5.23 (1.17) |
| Max depth | 4 (0) | 5 (0) | 3.23 (0.60) |
| Total depth | 14 (0) | 22 (0) | 14.31 (3.40) |

| | | | |
|---|---|---|---|
| *Readability statistics:* | | | |
| Ease | 100 | 100 | 94.4 |
| Grade level | 1.1 | 1.5 | 2.0 |
| *Semantics:* | | | |
| Age of acquisition | 3.86 (0.69) | 3.96 (0.76) | 4.75 (1.86) |
| Concreteness | 3.58 (1.13) | 3.55 (1.03) | 2.85 (1.04) |
| Word frequency | 222 521 (497 442) | 186 142 (457 906) | 121 963 (318 769) |

Based on the three factors identified in Experiment 1, a composite score was formed based on proportion items correct for relevant sections and then correlations were formed. There was no significant difference between the working memory and linguistic composites with respect to total Yngve depth, $t(41) = 1.38$, $p = .18$ (Table S2). Total Yngve depth is similar to sentence length, which also did not differ between the two composites, as reported in the main text.

**Table S2.** Composite score of total depth for factors identified

| | Syntactic complexity |
|---|---|
| Composite | Total depth |
| Basic attention | 3.64 (0.81) |
| Working memory | 17.4 (5.03) |
| Linguistic | 14.31 (3.40) |

*Note.* Basic attention = Parts 1 and 2; Working memory = Parts 3-5; Linguistic = Part 6

## Supplementary Experiment 2

### Method

***Language***. Each child completed the board-designed *DDSB narrative language* measure. The child listened to a story with corresponding images and then was asked to retell the story, answer comprehension and vocabulary questions, and share a personal narrative about a similar event. All other available test measures are described fully in the main text.

### Results and Discussion

A composite score was formed for each factor identified in Experiment 1 and then correlated with the board-designed narrative measure. The mean raw score for this narrative measure was 33.29 (SD = 15.82) with scores ranging from 0-58. Exploratory partial correlational analyses between the board-designed DDSB narrative measure with the working memory composite (controlling for language) and with the linguistic composite (controlling for working memory) were not significant, partial $r = .003$, $p = 1$, BF = 0.26, and partial $r = .11$, $p = .65$, BF = 0.29, respectively. Similarly, the Test of Narrative Language (TNL; Gillam & Pearson 2017) was not associated with any composites either, as reported in the main text.

## Supplementary References

Gillam, R., & Pearson, N. (2017). *Test of Narrative Language* (2nd ed.). Pro-Ed.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System demonstrations* (pp. 55–60).

Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society, 104,* 444–466.