

Supplemental Material S1. Additional methodological details.

Baselines

Receptive Probe Foils

An image of each word in the receptive probe was accompanied by three foils, including an unknown semantic foil, a known phonological foil, and a known unrelated foil. Known versus unknown status was determined using the toddler's most recent MCDI, so foils were drawn from the MCDI as much as possible. Our criteria for foil selection were broad to accommodate toddlers' diverse vocabularies. Unknown semantic foils either (a) shared a semantic feature with the baseline word (e.g., "rip" and "cut" both involve splitting something); (b) shared a superordinate category (e.g., "moose" and "giraffe" are both animals); (c) were antonyms (e.g., "old" and "new"); or (d) were from the same word class (e.g., "slow" and "soft" are both adjectives). For phonological foils, we selected words that shared as many phonemes as possible with the baseline word. All shared at least one phoneme, whether a vowel or consonant (e.g., the /i/ in "she" and "bee," the /l/ in "long" and "loud"), and many shared two or more phonemes (e.g., the /u/ and /s/ in "moose" and "juice"). Whenever possible, we selected phonological foils with the same number of syllables as the baseline word. Although we aimed for phonological foils to be known words, we occasionally used words marked as unknown on the MCDI when none of a toddler's known words matched our criteria. The third set of foils, known unrelated foils, had no apparent relation to the baseline word (e.g., "dish" and "cat") but were marked as "understood" on the child's most recent MCDI.

Receptive Probe Word Elimination

If a toddler pointed to an incorrect image or did not respond on their first attempt, the word was marked as incorrect. If a toddler pointed to a correct image on their first attempt, we presented it a second time later in the baseline session. For the second attempt, the slide consisted of the same three foil images as in the first round but a new image of the word itself, and image arrangement was re-randomized. (In our first set of toddlers, one subject failed to recognize three words at Baselines 1 and 2. When they correctly identified the words at Baseline 3, we suspected this was due to familiarity with the specific images, rather than new comprehension of the words. Indeed, when we used novel images for the second set of receptive probes at Baseline 4, the toddler did not correctly identify the words. Accordingly, we changed our procedures for subsequent subjects.) We completed this second round to ensure that children had not selected the picture correctly on the first round simply due to chance, visual appeal, or location on the screen. If the toddler pointed to an incorrect image or did not respond on the second attempt, the word was marked as incorrect. However, if on the second attempt the child again correctly pointed to the word, we marked it correct and eliminated it from the potential words to use in treatment.

Target Word Treatment Order

Prior to enrolling participants, we generated multiple lists that randomly ordered items from 1 to 10. We used a new list each time to determine the order of treatment of the pairs. There were five instances where we manipulated this order. In two cases, we moved a word pair to the end of the list to avoid it occurring in the same session as a semantically similar word in another pair (e.g., "jeans" and "tights"). In three cases, we moved a word pair that would have been first

to later in the order because we needed to obtain an additional baseline point for one or both words in the pair. This was for efficiency: we were able to start treatment using three finalized pairs and simply collect the last data point before the first session, without the possibility of eliminating a word intended as a target or control for that session.

Treatment Personnel

Clinicians were students (undergraduate and graduate) who completed general training on the VAULT protocol using the worksheets and training videos used in Mettler et al. (2023; see Supplemental Materials S2.), as well as training specific to the object and action variability treatment conditions. Clinicians completed at minimum one practice treatment session before working with a real participant. Training was overseen by senior members of the research team, many of whom are licensed, certified speech-language pathologists. Due to COVID-19 protocols at the start of the study and a desire for consistent treatment conditions for all subjects, clinicians wore face masks throughout the entire study. The scorekeeper tallied clinicians' doses live and provided updates on dose number. The reliability person was a senior member of the research team and often a licensed, certified speech-language pathologist. They provided support to the clinician as needed (e.g., supported material management, provided feedback to clinician on their dose rate and other VAULT principles), recorded clinicians' fidelity to treatment condition (see Fidelity to Treatment Condition below), and recorded the toddler's verbal output (see Reliability of Toddler's Productions below). In addition, they videorecorded each treatment session.

Determining Fidelity to Treatment Condition

For the action condition, the reliability person recorded whether there were 3 exemplars used in 2 different activities for each target word. For the object condition, they recorded whether there were 3 exemplars used in one activity and a different set of 3 exemplars used in a second activity for each target word. Fidelity was calculated out of 24 points per session: one point per exemplar per target word and one point per activity. Points were deducted if insufficient or excess exemplars were used or if exemplars were used in the wrong number of activities, depending on condition.

Exemplars by Word Class

For nouns, physical exemplars were deemed unique if they differed from each other in size, shape, posture (e.g., a sitting vs. a standing figurine), material (e.g., wood, plastic), texture (e.g., smooth, bumpy, hard, soft), or a combination thereof. Differing in color alone was insufficient. Multiple identical exemplars were counted as just one exemplar (e.g., multiple small brown roosters were counted as just one "rooster" exemplar). Clinicians were encouraged to choose exemplars that were as physically distinct as possible (e.g., a small plastic plane with propellers vs. a medium wooden plane without propellers). Additionally, they were instructed to integrate exemplars equally throughout a given activity to allow the child an opportunity to observe or interact with each of them. Below are descriptions of sample exemplars for different types of target words. See Figure 1 for images of sample exemplars for the word "airplane."

For verbs, exemplars were entities demonstrating the action or having the action done to them. For example, three exemplars of "pour" could be three items that can pour: a red bowl, a short plastic drinking cup, and a measuring cup. For "fix," three exemplars could be three different animate objects (e.g., figurines of a child, a young adult, and an older adult) fixing an item. Verbs were presented in multiple morphological forms (e.g., poured, pouring), but the intention was to teach children the root word, not verb morphology.

For prepositions, exemplars demonstrated the preposition relative to some other object. Three exemplars of "above" could be three items positioned above other items during an activity (e.g., a bird, airplane, and superhero flying above a building).

For adjectives and adverbs, exemplars demonstrated the characteristic or quality of the target. For example, three exemplars of "long" could be a long hair ribbon, a brush with an unusually long handle, and a doll's long hair in an activity playing with the doll's hair.