Article:

Unger, Layla and Sloutsky, Vladimir M. (2025) Dynamics of Learning New Words From Context. Journal of Experimental Psychology: General. ISSN: 1939-2222

https://doi.org/10.1037/xge0001768

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2025, Vol. 154, No. 7, 1784–1799
https://doi.org/10.1037/xge0001768

Dynamics of Learning New Words From Context

Layla Unger[1, 2] and Vladimir M. Sloutsky[2]
[1] Department of Psychology, University of York
[2] Department of Psychology, The Ohio State University

Often the only source of information for learning a new word is its surrounding language context. For example, even if one has never seen a rambutan, it is possible to learn that "rambutan" is a kind of fruit just from hearing "I like sweet, juicy rambutans." What processes unfold at the moment upon encountering a new word in context that lead to successful word learning? We conducted three experiments to evaluate the role of working memory, which may be critical for linking a new word to the meaning implied by its surrounding language context. In each experiment, we assessed word learning from sentences in which new words occurred either before or after an informative context. In Experiment 1, we tracked gaze during reading to gain insight into the real-time processing of the surrounding language context and the new word. Results highlighted the importance of working memory resources for holding the language context in mind while processing the new word, regardless of which was encountered first. Experiment 2 replicated the importance of working memory resources for learning new words heard in fluent speech, and Experiment 3 replicated this finding while controlling for overall engagement measured from performance on an unrelated task. Together, these findings support the conclusion that successful word learning from context depends on maintaining the context in working memory while linking it to a new word.

Public Significance Statement

Humans are unique among animals in our ability to learn and use thousands of meaningful communicative signals—that is, words. Word learning is all the more remarkable because we somehow get a sense of what many words mean just from hearing or reading them in the context of everyday language. For example, from reading "Rambutans boast a juicy, sweet flavor," we can get a sense that a "rambutan" is a fruit even if we have never encountered one before. The experiments here were designed to shed light on how we accomplish this feat. The results point to a key role for the ability to hold the other words accompanying a new word (e.g., "juicy," "sweet," "flavor") in mind while linking the meaning they imply to the new word.

Keywords: word learning, working memory, eye tracking, reading, learning from context

Supplemental materials: https://doi.org/10.1037/xge0001768.supp

Words provide us with the building blocks to communicate about a limitless range of ideas, from science and philosophy to what we want for dinner. Much of what we know about how words are learned comes from early in development when young children often pick up their first words by mapping them to objects, events, and other observable referents in the world around them. Research into early word learning has revealed a suite of cognitive processes that help children map words to referents, from following a

This article was published Online First May 15, 2025.

Michele Diaz served as action editor.

Layla Unger https://orcid.org/0000-0002-8888-1265

Data, analysis scripts, and stimuli have been made available for blind peer review at https://osf.io/2prvy/ and will be made publicly available pending publication. A preliminary version of this work was presented at the 46th annual meeting of the Cognitive Science Society. All activities involved in this research were approved by the institutional review boards at The Ohio State University and the University of York.

The authors declare no conflicts of interest with respect to their authorship or the publication of this article. This work was supported by the National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development, under Grants R01HD078545 (awarded to Vladimir M. Sloutsky) and R01HD111458 (awarded to Layla Unger and Vladimir M. Sloutsky).

Layla Unger played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing. Vladimir M. Sloutsky played a lead role in funding acquisition and resources and a supporting role in methodology and writing–review and editing.

Correspondence concerning this article should be addressed to Layla Unger, Department of Psychology, University of York, Heslington, York YO10 5DD, United Kingdom. Email: layla.unger@york.ac.uk

speaker's gaze to tracking the objects that are consistently present when a word is heard (e.g., Markman & Wachtel, 1988; Smith & Yu, 2008; Tomasello & Farrar, 1986). Yet, although these processes are useful for building early vocabularies of common words that have real-world counterparts, they cannot support the acquisition of thousands of words that are typically learned across a lifetime. When one of these words is first encountered, often the only source of information about what it means is its surrounding language context. For example, from words like "juicy" and "sweet," the sentence "Rambutans boast a juicy, sweet flavor" conveys the sense that a "rambutan" is a fruit. In contrast with mapping, little is known about the cognitive processes that unfold in the moment upon encountering a new word in the language that lead to successful word learning.

In what follows, we outline two candidate cognitive processes that have been proposed either as computational or conceptual models. We then highlight key unknowns about these processes, with a particular emphasis on an underexplored role of working memory (WM). The focus on working memory is motivated by the fact that a new word and its surrounding context are encountered at different points in time but must be linked for successful word learning.

Finally, we present three studies designed to shed new light on the dynamics that unfold upon encountering a new word in a surrounding language context and their relationship with working memory.

Mechanistic Processes for Learning Words From Context

The candidate processes for learning words from the context that we synthesize here either explicitly or implicitly invoke the distributional hypothesis (Harris, 1954; Landauer & Dumais, 1997; Miller & Charles, 1991; Rubenstein & Goodenough, 1965). According to this hypothesis, words similar in meaning occur in similar language contexts. For example, different words for fruits often occur in the context of "juicy" and "sweet." There is extensive evidence that these distributional regularities are ubiquitous in language (e.g., Jones & Mewhort, 2007; Landauer & Dumais, 1997; Lund & Burgess, 1996; Melamud et al., 2016; Mikolov et al., 2013; Pennington et al., 2014). In turn, these regularities provide opportunities for learning a new word based on its occurrence in similar contexts to words similar in meaning, such as that "rambutan" is a fruit from its occurrence with "juicy" or "sweet" (Savic et al., 2022). Yet, whereas the distributional hypothesis suggests that such learning is possible in principle, it does not speak to the cognitive processes that link new words to known words from distributional regularities. Here, we draw from multiple prior literatures to evaluate the landscape of candidate processes (see Figure 1 for a schematic illustration).

One candidate process comes from the increasingly popular proposal that processing incoming language input involves prediction (DeLong et al., 2005; Kutas et al., 2019; Nieuwland et al., 2020; Van Petten & Luka, 2012; Willems et al., 2016). According to this proposal, as people process incoming language input, they are routinely predicting upcoming words. Importantly, the actually observed words (when compared to the predicted ones) serve as error feedback signals to improve predictions in the future. Building on this idea, some researchers have proposed that when a learner predicts an upcoming word and is incorrect, they link the erroneously predicted words to the observed word (Borovsky et al., 2012; Ervin, 1961). For example, a learner who hears, "I can't wait to eat some sweet, juicy" might predict one or more known words for fruits. When a novel word such as "rambutan" occurs instead, the words that the learner predicts and the word that they hear will be coactive. Therefore, it is possible to learn that "rambutan" is a fruit by linking it to erroneously predicted known words for fruits. Critically, this process can only unfold when an informative context is encountered before a new word. We therefore refer to this as the Forward account.

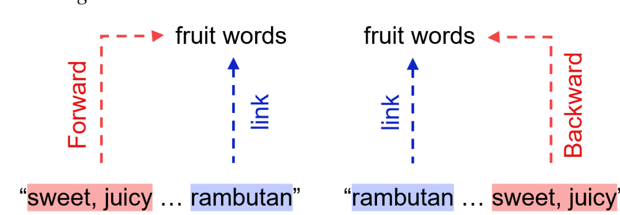
What about when a new word precedes an informative context, as in "Rambutans boast a sweet, juicy flavor"? The Forward account provides no explanation for learning in this situation. Instead, the new word must somehow be linked to known words that are likely to have preceded the informative context. We therefore refer to candidate explanations of this process as Backward accounts. One such Backward account is simply the inverse of the Forward account: that is, that new and known words can become coactive because processing language involves making backward retrodictions of words likely to have preceded those currently observed (Chaffin et al., 2001; Onnis et al., 2022). Another potential backward process comes from popular mechanistic models for learning how sequences—including words in language—tend to unfold over time, such as recurrent neural networks (e.g., Borovsky & Elman, 2006; Elman, 1990). Like the Forward account, these models also invoke the prediction of upcoming words, but use feedback from prediction error in a different way. Specifically, upon encountering a word that was not well predicted, error feeds back to update the representations of preceding words to increase the likelihood of making more accurate predictions in the future. Thus, when the word "rambutan" is unknown, it cannot be used to predict "sweet" or "juicy," as in "Rambutans boast a sweet, juicy flavor." Therefore, encountering "sweet" and "juicy" after "rambutan" then drives a feedback signal that updates a representation of "rambutan" to better predict these words. Because representations of known fruit words have already been formed from their occurrence with "sweet" and "juicy," the updated representation of "rambutan" becomes similar to the representations of known fruit words.

Together, the Forward and Backward accounts lay out a range of possible processes that might harness distributional regularities in language to foster word learning. However, these processes have not been contrasted or studied systematically.

First, it is unknown whether word learning is different or equivalent in the Forward and Backward directions. For example, it

Figure 1

Schematic Illustration of Forward and Backward Routes for Learning Words From Context



Note. See the online article for the color version of this figure.

is possible that one of the Forward or Backward accounts captures the primary way in which people learn words from context. If this is the case, then word learning should be more successful in either the Forward or Backward route. Alternatively, it is possible that each account only captures a complementary part of the story. If both Forward and Backward processes unfold similarly readily, then word learning may be equivalent across routes.

Evidence suggestive of differences comes from studies of another language comprehension challenge: resolving temporary ambiguities. Ambiguities are common during language processing, including when a word has multiple meanings or senses (e.g., "ball" can refer to a toy or formal dance) or is muffled in speech. As in determining the meaning of a novel word, the surrounding language context can help resolve these ambiguities (e.g., when "ball" occurs in a context relating to a game vs. a formal event). Accordingly, just as context can activate a familiar word that can be linked to a new word during word learning, context can activate a specific interpretation of a familiar word that can resolve an ambiguity. Because of these parallels, we can use the larger body of prior research on the use of context to resolve ambiguities to draw inferences about the use of context to learn words. However, we will also note key ways in which word learning from context may differ from ambiguity resolution from context, which highlights the importance of directly studying word learning from context.

Prior studies suggest that ambiguity resolution is more challenging in the equivalent of the Backward route when the ambiguity precedes an informative context (Frazier & Rayner, 1990; Rayner & Frazier, 1989; Samuel, 1991; see also Jesse & McQueen, 2011). Based on these findings, word learning may be less successful when it can only take place via the Backward route versus the Forward route. However, there are important differences between ambiguity resolution and word learning. To resolve an ambiguity, it is necessary to choose between multiple known interpretations of a word or speech sound, which may further involve needing to overcome a more frequent and dominant interpretation in favor of a weaker one (e.g., interpreting "ball" as a formal dance instead of the more frequent toy meaning). Therefore, the relative difficulty of the Backward versus Forward route in ambiguity resolution may come in part from the need to use the following context to sift between candidate interpretations and possibly reject a dominant one. In contrast, these demands are minimized in word learning because entirely novel words do not have multiple known interpretations. Therefore, the question of differences versus equivalence in the Forward and Backward routes to word learning remains open.

Another key unknown regarding the Forward and Backward routes comes from the fact that both directions involve linking a new word that is encountered at 1 point in time to a meaning implied by its context that is encountered at another. Therefore, working memory may be needed to maintain and manipulate information over time. Yet, the existing accounts have not explicitly incorporated a role for working memory, and its contribution to word learning from context has received little empirical attention.

Working Memory in Learning Words From Context

In principle, it is possible that working memory is important for linking a new word to the meaning implied by its context. This possibility is consistent with the influential account of the importance of working memory for language comprehension in general proposed by Just and Carpenter (1992). According to this proposal, working memory capacity is needed to store and interpret recently encountered language and to use these contents to anticipate and interpret upcoming language input. For example, working memory has been implicated in relating the meaning of a familiar word to the overall sense conveyed by its surrounding sentence context (Van Petten et al., 1997). To be successful, both the Forward and Backward routes outlined above may require such working memory involvement. However, although there is evidence implicating working memory in the storage of novel word forms (Baddeley et al., 1998; Gathercole & Baddeley, 1993), direct evidence regarding a role for working memory in learning the meanings of new words from context is relatively limited.

A handful of studies point to a role for working memory in word learning via the Forward route. In Daneman and Green's (1986) study, individuals with larger working memory spans were more successful at learning new words that occurred toward the end of an informative paragraph. Similarly, in a study with 9- to 11-year-old children, learning words that occurred after an informative sentence was correlated with a composite metric that included working memory span (Hill & Wagovich, 2020). Even less evidence speaks to a role for working memory in the Backward direction. To the best of our knowledge, the only relevant findings come from a small-scale study of learning words followed by informative contexts in 9- to 10-year-old children, which yielded mixed results regarding the role of working memory (Cain et al., 2004). Finally, across both routes, there is some evidence that adult second language learners with greater working memory capacity more successfully pick up new words that are preceded and followed by informative contexts in the second language (Perez, 2020).

Beyond this handful of studies, further evidence regarding the role of working memory in word learning is largely indirect. For example, (a) working memory is correlated with vocabulary size across school-age child development (Nilsen & Graham, 2009; Roman et al., 2014; Sesma et al., 2009), and (b) it has been estimated that hundreds to thousands of words enter vocabularies from encounters in language contexts (Nagy et al., 1987). Taken together, (a) and (b) suggest that working memory may be a factor affecting vocabulary size.

Suggestive indirect evidence also comes from studies outside the realm of word learning. Regarding the Forward route, as outlined above, one component of this route involves using a recently encountered context to predict upcoming words. Working memory has been implicated in this component of the Forward route in multiple studies. For example, upon encountering a word that is inconsistent with its preceding context, individuals with medium to high working memory capacity show electrophysiology-derived signatures of detecting an inconsistency. In contrast, these signatures are reduced or absent in individuals with low working memory (Van Petten et al., 1997; Yang et al., 2020; see also Nieuwland & Van Berkum, 2006).

In contrast with the suggestive evidence for working memory in the Forward route, the interpretations that can be drawn from nonword learning studies regarding the Backward route are less clear. Arguably, some research on resolving ambiguities in language point to a plausible role for working memory in the Backward route. Specifically, individuals with greater working memory spans also more readily use an informative context to resolve a preceding ambiguity (Miyake et al., 1994). However, as noted above, resolving ambiguities involves sifting between and possibly suppressing different known interpretations of familiar words. It is these processes that have been attributed to the role of working memory in ambiguity resolution. At the same time, these processes are not relevant to word learning because novel words do not have multiple known interpretations. Thus, inferences that can be drawn from ambiguity resolution to word learning are tentative at best.

Taken together, there is suggestive evidence for a role for working memory in word learning from context. However, the dynamics with which working memory is recruited when learning words via the Forward and Backward routes remain unclear. As with the question of whether word learning is different or equivalent in the Forward and Backward routes, it is unknown whether working memory is recruited differently or equivalently across routes. In the following section, we discuss the insight that may be gained into these unknowns using eye tracking.

Illuminating the Recruitment of Working Memory During Word Learning With Eye Tracking

Tracking gaze during reading has been used extensively to illuminate the cognitive dynamics that unfold during online language processing. Broadly, the amount of time spent looking at a given word or multiword section of text is indicative of the effort used to process it (Birch & Rayner, 1997; Rayner et al., 2006). Such effort can include the degree to which the reader is actively drawing upon the storage or manipulation aspects of working memory during language processing.

Several findings support this interpretation of gaze duration. For example, ambiguous words with multiple equally common meanings that a reader may need to sift between using working memory are inspected longer than words with a single dominant, common meaning (Duffy et al., 1988; Rayner & Duffy, 1986). Moreover, this effect is reduced when a preceding context strongly cues a single meaning for the ambiguous word (Binder & Morris, 1995; Duffy et al., 1988). Thus, gaze duration can capture demands to keep multiple interpretations active in working memory. Similarly, gaze durations increase when the meaning implied by the context of an ambiguous word changes across a discourse (Binder & Morris, 1995; Witzel & Forster, 2014). Thus, gaze duration can capture demands to manipulate and choose between word meaning interpretations in working memory. In the realm of word learning, one study also points to a relationship between gaze duration and working memory demands. In Chaffin et al.'s (2001) study, readers encountered either novel words or low- or high-frequency familiar words before an informative context. After reading the context, readers subsequently spent longer rereading the novel versus familiar words, suggesting that readers were drawing upon working memory resources to maintain contextual information in mind while updating their sense of the novel word. Thus, gaze durations can reflect the active storage and manipulation demands on the working memory processes putatively involved in both language comprehension in general and word learning in particular.

We next provide an overview of how gaze to novel words and informative contexts can illuminate the dynamics of these processes during word learning. The key value of gaze is to identify when working memory resources are recruited during word learning from context, and whether this recruitment varies across word learning via the Forward and Backward routes.

Gaze to Novel Words

Gaze to novel words likely reflects different processes depending on whether the reader is looking at the novel word before or after they have encountered a context that is informative about the novel word's meaning. When a novel word is read prior to reading an informative context (as in the Backward route), the reader does not yet have information about what the new word means. Thus, gaze duration to the novel word at this point only reflects processes elicited by encountering a new word, such as surprise at encountering a new word or effort involved in decoding and storing the word form (Baddeley et al., 1998; Gathercole & Baddeley, 1993). For this reason, gaze duration to the novel word in the Backward route can be thought of as a baseline for simply processing a novel word itself, when there is yet no opportunity to use its context to get a sense of its meaning.

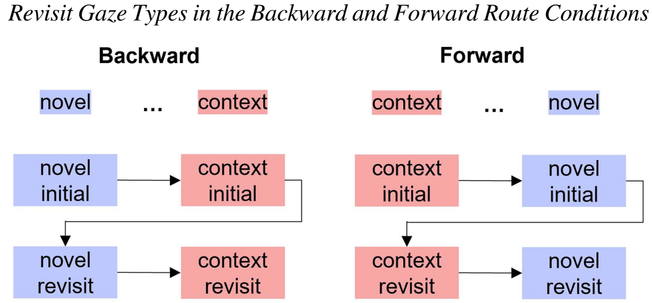
In contrast, encounters with a new word after an informative context has been read are likely more indicative of processes involved in learning the meaning of the word. Specifically, time spent reading a new word after an informative context has been read likely indicates the degree to which the reader is using contextual information that they are maintaining in working memory to update their sense of what the new word might mean. The point at which this type of encounter first takes place differs in the Forward and Backward routes. In the Forward route, the informative context precedes the new word. Therefore, encounters with the new word after the informative context take place starting with the very first time the new word is read. In the Backward route, the informative context occurs only after the new word. Therefore, the first time the new word is read, the informative context is not yet available. Instead, encounters with the new word after the informative context only take place if and when the reader looks back to revisit the new word after they have read the context (see Figure 2). Accordingly, any gaze duration differences during these encounters can capture differences between the Forward and Backward routes in the degree to which context is maintained in memory while linking it to the new word. Differences in maintaining the context in memory while linking it to the new word in individuals with high versus low working memory capacity can likewise be inferred from this measure.

Gaze to Context

The informative value of gaze durations to context is similar to the value of gaze durations to the novel word. When the context is read

Figure 2

Schematic Illustration of How Gaze Is Broken Down Into Initial and Revisit Gaze Types in the Backward and Forward Route Conditions



Note. See the online article for the color version of this figure.

prior to the novel word, gaze duration likely indicates the effort involved in storing this context in working memory, as well as perhaps predicting upcoming words. Time spent reading the context after the novel word has been read can instead capture the process of maintaining the novel word in working memory while linking it to the meaning implied by the context. These periods include the initial encounters with the context in the Backward route and revisit encounters in the Forward route. Following the same logic as above, gaze duration during these periods can capture differences in engagement in this process during the Forward and Backward routes and in readers with high versus low working memory capacities.

Present Experiments

The goal of the present experiments was to illuminate the online processes that unfold upon encountering an opportunity to learn a new word from its surrounding context. We investigated two routes to learning words from context: a Forward route in which the context precedes the novel word and a Backward route in which the context only occurs after the novel word. Because these processes inherently involve linking together information that is encountered at different points in time, the experiments were specifically designed to illuminate the processes of maintaining and manipulating information in working memory.

For this research, we examined word learning from sentences designed to be equivalent except for the occurrence of a novel word before versus after an informative context, such as "The monkey's favorite food is doffs" (Forward)/"Doffs are the monkey's favorite food" (Backward). As illustrated by this example, the informative context in a sentence pointed to a meaning for the novel word that is related to a familiar known word such as "bananas." Learning words from these sentences thus parallels the real-world scenario of learning a word from context that is similar in meaning to a known word, such as learning that "conflagration" is similar in meaning to "fire."

In Experiment 1, we used eye tracking to illuminate how the dynamics of processing these sentences varied with working memory capacity. To anticipate our results, we found that regardless of the relative ordering of novel words and contexts, processing involved maintaining the context in mind while linking it to the novel word. In contrast with individuals with low working memory capacities, individuals with higher working memory capacities both (a) engaged in this process to a greater extent and (b) were more successful at learning word meanings. Experiment 2 was designed to rule out the possibility that this pattern was idiosyncratic to reading, where individuals have the opportunity to look back and forth between different parts of a sentence. To accomplish this goal, Experiment 2 replicated the investigation into the role of working memory capacity in the Forward and Backward routes to word learning, with participants listening to spoken language input. Finally, Experiment 3 was designed to rule out the possibility that the apparent positive relationship between word learning and working memory was spurious—instead, individuals who are more engaged, attentive, or good at doing tasks may simply have performed better both on word learning and the assessment of working memory. Thus, Experiment 3 added a task that assesses an ostensibly unrelated cognitive construct of inhibition. If the apparent relationship between word learning and working memory was spurious, then working memory should disappear after controlling for performance on another task.

Transparency and Openness

All stimuli and data are available on the Open Science Framework (OSF) at https://osf.io/2prvy/. The OSF repository additionally includes all R code for analyses and figures with extensive comments for reproducibility. Prior to peer review, a talk was presented on preliminary findings from this study (Unger & Sloutsky, 2024).

Experiment 1

Method

All research activities received ethics approval from The Ohio State University Institutional Review Board (Protocol #2017B0149). This study was not preregistered.

Participants

The sample included young adults recruited from the undergraduate population at a large Midwestern university and adults recruited from the broader population in the surrounding city. Participants were only asked to affirm that they were 18 or over and were not asked to report their gender, sex, race, or ethnicity. Participation was compensated with either course credit or a $10 gift card. From an initial sample of 86 participants, we excluded eight due to poor eye-tracking data quality (see criteria in the Results section), yielding a final sample of 78 participants.

Sentence Materials

The primary materials consisted of sentences generated using the following criteria. Each sentence contained one novel word accompanied by context words that are: (a) informative about the novel word's meaning and (b) highly familiar to adults, as described below. The full set of sentences can be found in Supplemental Materials.

Informative Contexts. Context words are informative about a novel word's meaning because they reliably co-occur in language input with a target familiar word. For example, in the sentence, "At Jessie's birthday party they had delicious fimp," the familiar words "birthday," "party," and "delicious" reliably co-occur with the target familiar word "cake." Thus, the novel word "fimp" shares its context with "cake." All target familiar word meanings referred to a specific concrete noun, such as cake.

We generated a Forward and Backward version of each sentence. For example, the novel word "fimp" occurs before the informative context in the sentence "The fimp at Jessie's birthday party was delicious," and after the informative context in "At Jessie's birthday party they had delicious fimp." To ensure that the context portions of the Forward and Backward versions are equally informative about the meaning of the novel word, all sentences were normed with a separate sample ($N = 31$) using a sentence-completion or "cloze" task. In the task, participants saw either the Forward or Backward version of each sentence, with the novel word replaced with a blank space. Participants were prompted to complete the blank space with a familiar word. Accordingly, we selected only sentences where more than 85% of participants entered the target familiar word for both the Forward and Backward versions. From a larger set of candidate sentences, we selected 16 sets of Forward and Backward sentence versions that met the above criteria.

Familiar Contexts and Target Word Meanings. An important consideration for the generation of sentence materials was their high familiarity for typical adults. This is a consideration due to the possibility that apparent effects of individual differences in working memory on word learning from context might actually stem from prior language experience. As noted by Acheson and MacDonald (2009), greater experience with language (e.g., through extensive reading) may foster more robust knowledge of words and how they are combined into larger units such as phrases and sentences. Such robust knowledge may in turn lead to both the better storage of verbal input in working memory and better performance on language tasks such as comprehension and the use of context to resolve lexical ambiguities. However, this concern has been applied to interpreting apparent working memory effects in difficult language tasks, such as comprehending sentences with unusual and complex syntactic structures or using context to infer that an ambiguous word refers to its rare rather than common meaning. Extensive language experience is vital for success on such difficult tasks because it familiarizes people with more unusual elements of language. In comparison, extensive language experience may have a limited impact on processing words and word combinations that are already very common in everyday language.

Due to these considerations (as well as an aim to generate materials that can also be used in future research with children), the sentence contexts we generated contained high-frequency words and pointed to high-frequency target familiar word meanings. We measured word frequency from corpora of everyday conversational language input to both children (MacWhinney, 2000) and adults (Du Bois et al., 2000–2005; child corpora were included to generate stimuli that can also be used in future developmental studies). We then divided words into 10 equal-sized frequency bins, where the 10th bin contained the most frequent words. Word frequency was very high for both context words ($M = 9.94$, $SD = 0.362$) and target familiar words ($M = 9.93$, $SD = 0.263$).

In addition, target familiar words were all selected from words used in the MacArthur-Bates Communicative Index, a measure of word knowledge in children up to 30 months of age. Data from this instrument on typical ages of word learning indicate that all target familiar words are learned early in development. Additionally, 80% of context words had available data for calculating the ages at which they are typically learned (Brysbaert & Biemiller, 2017; Fenson et al., 2007), which indicated that the context words are typically learned in early childhood ($M = 3.72$ years, $SD = 2.75$).

Equating Forward and Backward Sentences. We additionally ensured that Forward and Backward versions were matched on two important characteristics. To quantify how well-matched sentences were, we used Bayesian *t*-tests because they estimate the strength of evidence for both differences (Bayes factors greater than 1) and equivalence (Bayes factors smaller than 1). First, versions were matched in length (number of words in Forward: $M = 10.9$, $SD = 2.77$; Backward: $M = 10.6$, $SD = 2.43$; t test Bayes factor = 0.197 providing moderately strong evidence for equivalence). Second, we ensured that both versions reflected the order in which words tend to co-occur in everyday language. For example, consider the Forward sentence version "At Jessie's birthday party they had delicious fimp" and the Backward sentence version "The fimp at Jessie's birthday party was delicious." To equate these sentences, it is useful to check that "cake" is just as likely to (a) follow the context words in everyday language that it follows in the Forward version as it is to (b) precede the context words in everyday language that it precedes in the Backward version. Therefore, we measured these ordered co-occurrence tendencies in corpora of language input to both children (MacWhinney, 2000) and adults (Du Bois et al., 2000–2005; child corpora were included to generate stimuli that can also be used in future developmental studies). We calculated co-occurrence tendencies using Positive Pointwise Mutual Information (PPMI), which measures the degree to which words co-occur more reliably with each other than with other words. Thus, for every ordered co-occurrence between context and target familiar words in a stimulus sentence, we calculated the PPMI of this same co-occurrence in everyday language. For example, for the sentence "At Jessie's birthday party they had delicious fimp," we calculated the PPMIs for the occurrence "cake" after "at," "birthday," "party," "they," "had," and "delicious." Likewise, for the sentence "The fimp at Jessie's birthday party was delicious," we calculated the PPMIs of "cake" before "at," "birthday," "party," "was" and "delicious." This approach confirmed that Forward and Backward sentence versions similarly captured the ordered co-occurrence tendencies of everyday language (PPMI of co-occurrences in Forward: $M = 1.25$, $SD = 1.77$; Backward: $M = 1.16$, $SD = 1.66$; t test Bayes factor = 0.142 providing moderately strong evidence for equivalence).

We then generated two lists of sentences, with each containing eight Forward and eight Backward sentence versions. Each sentence was randomly assigned to appear in its Forward or Backward version in a given list. In the experiment, one of the two lists was randomly selected for each participant.

Pictures

As described below, word learning was assessed by prompting participants to match novel words to corresponding pictures. We, therefore, used pictures that depicted the target meanings of novel words, such as a picture of cake for "fimp" in the sentences "At Jessie's birthday party they had delicious fimp"/"The fimp at Jessie's birthday party was delicious."

Novel Words

Novel words were one- to two-syllable words that followed the phonotactic regularities of English. Words were taken from the Novel Object and Unusual Name database of novel words used in studies with children (Horst & Hout, 2016).

Working Memory Span

We assessed working memory using the extensively used reading span task originally developed by Daneman and Carpenter (1980). We used a version of this task adapted for open-source use (Klaus & Schriefers, 2016). This assessment consists of two parts: reading sentences and judging them for their semantic coherence and storing words for later recall. It is worth noting that in some versions of this task, the word that participants store is the final word of the sentences judged for semantic coherence. In contrast, the version used in this study presented separate, single words for storage after each sentence. This version therefore mitigates the concern that the measure of working memory span is heavily influenced by participants' skill and experience with processing the way that words are combined in phrases and sentences (Acheson & MacDonald, 2009). Materials for this assessment consist of: (a) sentences that are semantically coherent (e.g., "The man went out to buy a new car") or incoherent (e.g., "The lemonade players kicked the ball") and (b) individual high-frequency words (e.g., "table").

Apparatus

Gaze data were collected using an EyeLink Portable Duo eye-tracking system that measures eye gaze by computing the pupil-corneal reflection at a sampling rate of 500 Hz. To minimize movement during eye tracking, participants made responses using a gamepad controller.

Procedure

Participants first completed a word learning from context task, then a working memory assessment. Eye-tracking data were collected only during the word learning from context task.

The word learning from context task consisted of two phases: exposure and test. In exposure phase trials, participants read sentences containing novel words, and in test trials, participants were tested on whether they learned the meanings of the novel words. The 16 sentences were divided into four blocks, so that within a block, participants read four sentences, then were tested on the four novel words in these sentences. Each block contained an equal number of Forward and Backward sentence versions, with order randomized within blocks for each participant. Sentences were assigned to blocks so that within a block, no two sentences pointed to target meanings for novel words that were related. For example, a sentence in which the target meaning of the novel word was "sink" occurred in a different block from a sentence in which the target meaning of the novel word was "bowl" (see Supplemental Materials for all block assignments).

During exposure phase trials, participants first pressed a button on a gamepad controller to see the sentence. The sentence was then depicted on the screen across 2–3 lines of text (letters were presented in 52pt font, such that the height of letters subtended ≈ 1.4 degrees of visual angle). Reading was self-paced until participants pressed the same button to indicate the completion of the sentence. During test trials, participants were presented with a single novel word from one of the sentences in the same block. Participants were also shown two pictures side by side: one depicting the target meaning of the novel word and one depicting the target meaning of another novel word from the same block. The same two pictures were always presented together, so that one picture was the correct choice on one test trial, and the other the correct choice on another test trial. At the beginning of each test phase in a block, participants were instructed that if they thought they chose the wrong picture on one trial, they could choose it again on a following trial. During each test trial, participants were prompted to use the left and right gamepad buttons to choose the picture corresponding to the meaning of the novel word.

Following completion of the word learning from context task, participants then completed the working memory assessment. This assessment consisted of Daneman and Carpenter's (1980) reading span task, as adapted by Klaus and Schriefers (2016) for open access use. This assessment consisted of two components: a processing component in which participants read sentences and judged them for the semantic coherence, and a storage component in which participants stored single words for subsequent recall. The assessment was organized so that each processing component sentence was followed by a storage component word. Participants used the up and down arrows on the gamepad to judge semantic coherence of sentences at their own pace, then were shown the storage word for 1,200 ms. After a block consisting of 2–6 sentence–word pairs, participants were then prompted to recall as many of the storage words as they could remember. Responses were typed by the experimenter for subsequent scoring and analysis. The task included one block for each span size ranging from 2 to 6, presented in a random order.

Results

All data and analysis code are available in the materials on the OSF at https://osf.io/2prvy/. In addition, tables of full analysis results are presented in Supplemental Materials.

For all analyses, we first excluded participants with poor eye-tracking data quality (see criteria below), yielding a sample of $N =$ 78. Following Conway et al. (2005), we scored performance by calculating the average proportion of words correctly recalled, without weighting scores according to difficulty (e.g., recalling three out of six and one out of two words were treated as equivalent).

We then analyzed the effects of Forward versus Backward routes and varying working memory capacity (henceforth WM capacity). WM capacity was treated as a continuous variable in analyses, but for ease of visual interpretation, we also used a median split to divide participants into high and low WM capacity groups in graphs. We analyzed these effects of Forward versus Backward routes and WM capacity separately for word learning performance and gaze dynamics.

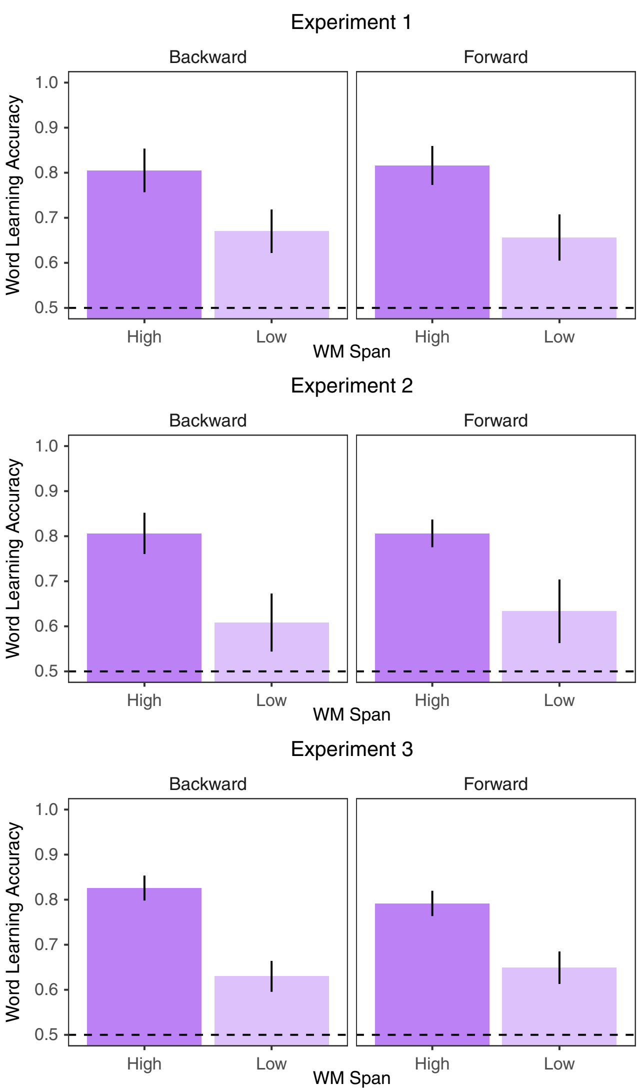
Because we were interested in both differences and equivalences (e.g., differences and equivalences in word learning via the Forward and Backward routes), all analyses were conducted as Bayesian mixed effects models in R using the brms package (Bürkner, 2017). The Bayesian approach yields a probability distribution for the magnitude of each fixed effect entered into a mixed effects model. From this probability distribution, it is possible to calculate: (a) a most probable point estimate of the magnitude of the effect, and (b) a "credible interval," which is a range that covers the most probable values. For example, for a 95% credible interval, the interpretation is that there is a 95% probability that the magnitude of the effect falls within the range covered by the interval. Therefore, a variable is commonly interpreted as having an effect when the 95% credible interval does not contain 0.

We selected the random effects following Bates et al. (2015) by starting with maximal models and eliminating random effects that did not contribute to the goodness of fit. Across analyses, the only random effect that consistently contributed to the goodness of fit was a random intercept for the participant. Therefore, we used this random effects structure across analyses. We additionally conducted frequentist versions of these analyses (included in code available on the OSF at https://osf.io/2prvy/) to ensure that results were equivalent.

Word Learning

To investigate word learning, we analyzed the effects of route (Forward vs. Backward) and WM capacity on word learning accuracy using Bayesian mixed effects logistic regression, with binary accuracy on each word learning trial as the outcome variable. Results are depicted in Figure 3. The analysis indicated that participants with higher WM capacity were more likely to learn words (point estimate = 3.70, 95% CI [0.38, 7.17]). In contrast, word learning did not vary between the Forward and Backward routes (point estimate = 0.27, 95% CI [−1.66, 1.13]). Moreover, there was no evidence for an interaction between route and WM capacity (point estimate = 0.39, 95% CI [−1.60, 2.38]). Thus, WM capacity contributed similarly to word learning success in both the Forward and Backward routes.

Figure 3
Word Learning in Experiments 1–3



Note. Graphs depict differences in word learning performance between participants with high and low WM capacity in the Forward and Backward route conditions. Error bars depict standard errors of the mean. WM = working memory. See the online article for the color version of this figure.

To follow up this analysis, we checked whether the apparent effect of WM capacity might instead have arisen from the possibility that some participants were overall inattentive during the experiment, leading to both poor word learning and poor WM assessment performance. To rule out this possibility, we harnessed the fact that the WM assessment included two tasks: a processing task in which participants judged the semantic coherence of sentences (used to place demands on WM) and a storage task in which participants stored individual words (used to assess WM capacity). Because the processing task is relatively simple, adults should easily perform well if they are attentive. Performance on this task was high overall, with an average accuracy of 95%. However, accuracy also ranged from 68% to 100%, indicating that some participants may have been inattentive. We therefore reran the primary analysis, only including participants with a judgment accuracy of ≥90% (excluded seven participants). The results of this analysis replicated the same pattern as the full sample, yielding evidence that the probability of word learning varied only with WM capacity (point estimate = 3.76, 95% CI [0.32, 7.45]).

Finally, we conducted a follow-up test of the apparent equivalence between the Forward and Backward conditions using a model comparison approach. The logic of this approach was to test whether a model in which WM capacity but not route condition predicted word learning accuracy provided a better fit to the data than models included route condition as a main effect or a factor that interacted with WM capacity. Specifically, we used leave-one-out cross-validation (Bürkner, 2017; Vehtari et al., 2020) to compare five models of word learning accuracy: a base model with no fixed effects, a model with route condition as a main effect, a model with WM capacity as a main effect, a model with both route condition and WM capacity as main effects, and a model with both main effects and their interaction. For data both with and without participants who performed poorly on the judgment task, this comparison revealed that the best fitting model was the model with only WM capacity as a main effect. This provides further evidence for equivalence between the Forward and Backward route conditions.

Illuminating the Role of Working Memory With Gaze Dynamics

The above analysis of word learning accuracy implicates WM in word learning from context. At the same time, this analysis does not shed light on what components of the learning process recruit WM. Gaze dynamics can illuminate this question because longer gaze durations to a given word or section of text capture greater WM demands. As described in the Introduction, gaze durations can thus reveal whether WM demands are implicated in processes that occur upon: (a) an initial encounter with the novel word, (b) an encounter with the novel word after reading the context, (c) an initial encounter with the informative context, or (d) an encounter with the context after reading the novel word.

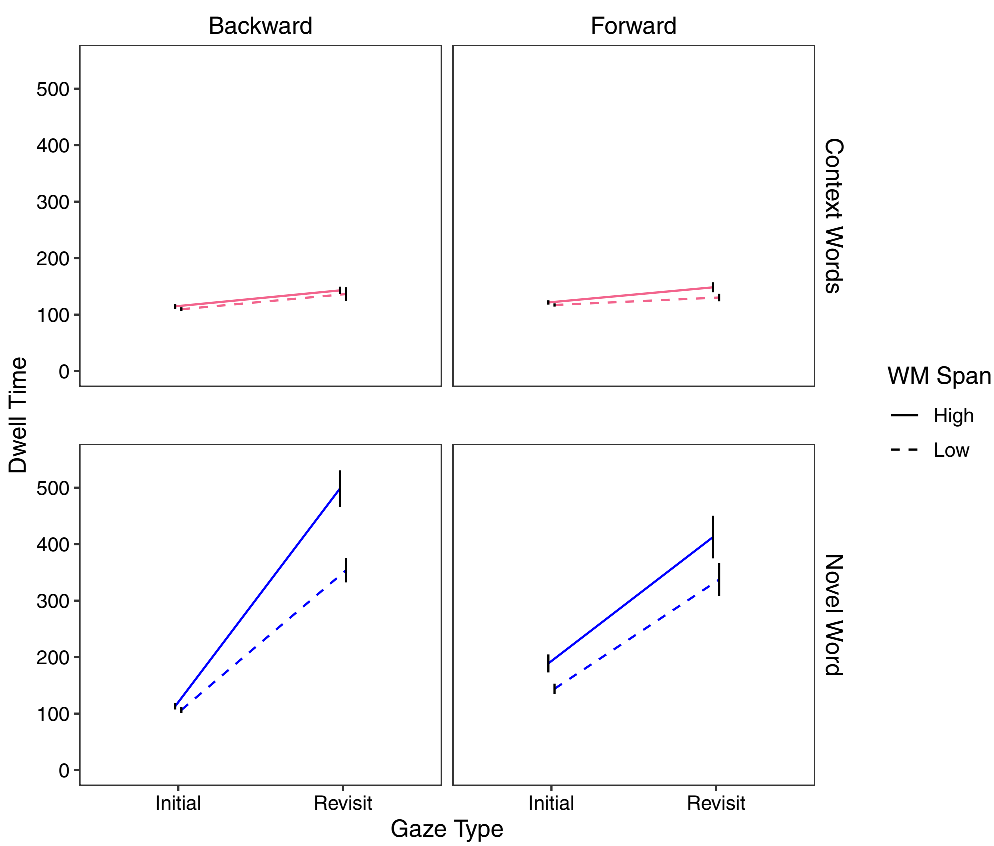
To investigate the use of WM resources during word learning from context, we first processed gaze data by computing trackloss on each trial—that is, the proportion of samples in the trial in which gaze data were missing. From the original sample of 86 participants, we removed data from participants of more than 40% (N = 8), leaving 78 participants. From the remaining participants, we removed trials with trackloss of more than 40%, resulting in the removal of 1.9% of trials.

We then defined an area of interest (AOI) for each word in each sentence using the eyekit Python package for analysis of reading behavior with eye tracking (Carr, 2023). We divided these AOIs into two categories: (a) the AOI for the novel word (henceforth, “Novel”) and AOIs for each word in the informative context (henceforth, “Context”). We then calculated two types of gaze durations for each AOI: (a) “Initial” gaze duration, starting from the first look to the AOI and ending with the first look away from it, and (b) “Revisit” gaze duration, including the total amount of time spent looking at the AOI after the Initial period (see Figure 2). Only gaze durations of >0 ms in an AOI were included in analyses. No other gaze data were removed prior to analysis. The reported results and figures therefore represent gaze durations for Novel and Context AOIs that participants spent at least some time looking at. All participants spent at least some time initially looking at and revisiting Novel and Context AOIs across sentences. Together, this process yielded Initial and Revisit gaze durations in the Forward and Backward conditions for Novel and Context words for participants with varying WM capacities, which are depicted in Figure 4. As described in the Introduction (see Gaze to Novel Words and Gaze to Context discussions), gaze to the Novel and Context words can provide different insights into the processes recruited during word learning. We therefore conducted the same analyses separately for gaze to Novel and gaze to Context words. Specifically, for each AOI type, we conducted an omnibus analysis of: (a) route condition (within subjects, Forward vs. Backward), (b) gaze type (within subjects, initial vs. revisit), and (c) working memory capacity (continuous individual differences variable). We then conducted any follow-up analyses needed to further illuminate the results of the omnibus analysis.

Gaze to Novel Word. The omnibus analysis revealed evidence for a three-way interaction between gaze type, route condition, and WM capacity (point estimate = 445.71, 95% CI [73.77, 813.47]). To tease apart these interactions, we conducted separate analyses for Initial and Revisit gaze durations, each testing the effects of route condition and WM capacity. To help the reader parse these effects, we first summarize the overall pattern before reporting the results. Overall, participants with greater WM capacity spent longer reading the novel word after they had already read the context and could thus use the context in WM to get a sense of what the novel word meant.

Figure 4

Gaze Dynamics in Participants High and Low WM Capacity in the Forward and Backward Route Conditions



Note. Error bars depict standard errors of the mean. WM = working memory. See the online article for the color version of this figure.

Within the Initial gaze, there was evidence of an interaction between route condition and WM capacity (point estimate = 150.01, 95% CI [28.44, 276.76]). The patterns of Initial gaze depicted in Figure 4 suggest that this interaction was due to a greater tendency for participants with higher WM capacity to spend longer initially reading the novel word in the Forward condition, when they had already encountered the context, than in the Backward condition, when they had not yet encountered the context. This pattern is further illustrated in Figure 5.

To further probe this interaction, we conducted follow-up analyses testing the effect of WM capacity on Initial gaze duration separately within the Forward condition and the Backward condition. In line with the patterns shown in Figures 4 and 5, Initial gaze to novel words was longer in participants with greater WM capacity only in the Forward condition, when the context had already been encountered and could thus be held in WM to get a sense of novel word meaning (point estimate = 178.09, 95% CI [16.07, 339.15]). In contrast, WM capacity was unrelated to initial gaze duration in the Backward condition, when readers had not yet encountered the context (point estimate = 45.35, 95% CI [−9.81, 102.43]).

Within Revisit gaze, there was only a main effect of WM capacity in which participants with greater WM capacity spent longer revisiting novel words (point estimate = 466.66, 95% CI [102.28, 830.23]). This result aligns with the analysis of Initial gaze: in both analyses, individuals with higher WM capacity spent longer reading the novel word after having read the context.

Gaze to Context. The omnibus analysis revealed no evidence for systematic effects of any variables on gaze durations to words in the context (all 95% CI contained 0).

Discussion

Overall, individuals with greater WM capacity were more successful at learning words from context. This role of WM capacity was implicated both in the Forward route when the informative context is encountered first and can thus be used to form predictions that support word learning, and in the Backward route when the informative context is only encountered afterward and must be used to retroactively update one's sense of what a new word means.

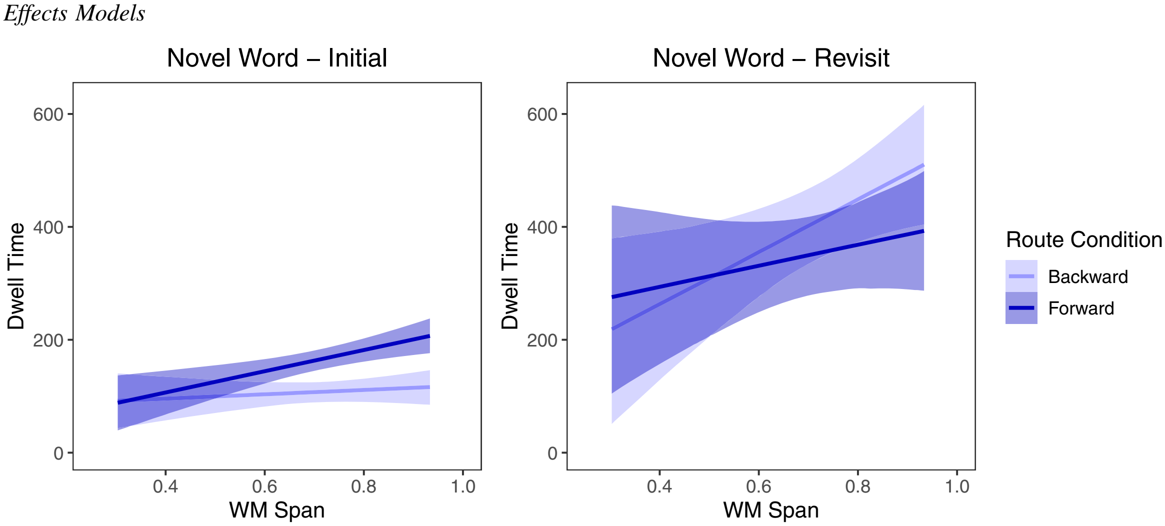
The gaze dynamics provide insight into the contribution of WM capacity. Individuals with greater WM capacity spent longer looking at the novel word after having encountered the context. In comparison, WM capacity was unrelated to time spent looking at the novel word before encountering the context. Given that higher WM capacity was associated with more successful word learning, this overall pattern suggests the important role of WM in storing, maintaining, and manipulating the context in mind while linking its implied meaning to a novel word, *regardless of whether the context or new word was encountered first*.

The goal of Experiment 2 was to test the possibility that the results of Experiment 1 might be idiosyncratic to reading. During reading, an individual has the opportunity to actively look back and forth between novel words and their surrounding contexts. This property of reading makes it possible to use eye tracking to gain insight into the online dynamics of language processing. Indeed, much of the insight offered by the eye-tracking results of Experiment 1 comes from gaze durations while revisiting novel words. At the same time, there is nothing to look back and forth between while processing spoken language. Therefore, the working memory dynamics revealed in Experiment 1 might not extend to spoken language. For example, the opportunity to look back and forth during reading might have contributed to the apparent equivalence of Forward and Backward routes in word learning.

Alternatively, the gaze dynamics that unfold during reading might capture processes that also occur mentally for spoken language. For example, looking back at a previously encountered word while reading might have an analog in processing spoken language in which the listener thinks back to a word they recently heard.

Figure 5

Interaction Between Route Condition and WM Capacity Within Initial and Revisit Gaze Durations From the Fitted Mixed Effects Models



Note. Error ribbons depict 95% CIs. WM = working memory. See the online article for the color version of this figure.

Therefore, Experiment 2 was a replication of Experiment 1, with the exception that participants listened to prerecorded versions of the sentences. Although this design cannot yield the same insights as eye tracking, it allowed us to test whether working memory played a similar role in successful word learning from spoken language contexts.

Experiment 2

Method

Participants

Participants were 51 adults recruited from the same population as Experiment 1 who had not participated in Experiment 1.

Materials and Procedure

Materials were similar to Experiment 1, with the exception that sentences were prerecorded rather than presented as text on a computer screen. Recordings were made by a female speaker fluent in English. To ensure that any differences between Forward and Backward route conditions could not be attributed to differences in the spoken articulation of the novel words, the same recording of each novel word (e.g., "fimp") was used in both versions of a given sentence.

The procedure was similar to Experiment 1, with the exception that no eye-tracking data were collected, and participants listened to sentences in the word learning from context task on headphones.

Results

Word Learning

As in Experiment 1, we analyzed the effects of route and WM capacity on word learning accuracy. The results mirrored those of Experiment 1. There was a main effect of WM capacity (point estimate = 3.63, 95% CI [2.08, 5.33]), in which word learning was more successful in participants with higher WM capacity. There was no evidence for an effect of route condition or an interaction between WM capacity and route condition (both 95% CIs contained 0). The results remained the same when analyses were restricted to participants who performed at an accuracy level of $\geq$90% in the semantic coherence judgment task (main effect of WM capacity point estimate = 5.51, 95% CI [3.03, 8.35]). Moreover, as in Experiment 1, model comparisons revealed that the data were best fit by a model that included only WM capacity (and not route condition) as a main effect.

Discussion

The results replicated the role of WM capacity in word learning from context in both the Forward and Backward routes. Accordingly, WM capacity is implicated in these processes across both reading and spoken modalities.

A remaining concern with the apparent positive relationship between WM capacity and word learning from context is that it may be spurious. One broad possibility is that this relationship might result from the possibility that individuals who were generally more motivated and alert would perform better at any set of tasks that require some form of effort than those who are not, leading to a misleading relationship between performance on the tasks. This concern is partially mitigated by findings from Experiments 1 and 2 that results are unchanged when analyses eliminate potentially inattentive participants based on semantic coherence judgment task performance. To further address this concern, in Experiment 3, we replicated Experiment 2 with an added ostensibly unrelated task. Specifically, participants in Experiment 3 additionally completed a standard Flanker inhibition task (Eriksen & Eriksen, 1974). This task is ostensibly unrelated because it does not involve language or the requirement to maintain information in working memory. Likewise, success in the word learning and working memory tasks does not depend on inhibition. At the same time, flanker task performance is influenced by overall motivation (Hübner & Schlösser, 2010; Ivanov et al., 2012; Yamaguchi & Nishimura, 2019) and attentiveness (Jugovac & Cavallero, 2012; Martella et al., 2011). The logic of this approach is that if the relationship between WM capacity and word learning is spurious, then WM capacity should no longer be related to word learning after controlling for performance on the additional task.

Experiment 3

Method

Participants

Seventy-five adults were recruited, including young adults recruited from the undergraduate population at a university in the United Kingdom (University of York Ethics Committee protocol #202433) and adults recruited from users of the Prolific platform located in the United States. Participation was compensated with either course credit or at a rate of $13 per hour.

Materials and Procedure

Materials and procedures for the word learning and working memory tasks were identical to Experiment 2. In addition, participants completed a control task consisting of a standard version of the commonly used Flanker inhibition task. In this task, participants judged the direction of a central arrow in the presence of flanking arrows that pointed in the same direction (congruent trials) or the opposite direction (incongruent trials). Participants completed 48 trials of this task, half congruent and half incongruent. The standard pattern of performance on this task is that participants are slower and less accurate on incongruent versus congruent trials due to the challenge involved in inhibiting the conflicting input from the flanking arrows.

Results

The goal of the analyses was to evaluate whether WM capacity is associated with word learning above and beyond performance on the Flanker control task. To accomplish this goal, we first calculated metrics of performance on the Flanker task. This task yields both accuracy and reaction time (RT) on congruent and incongruent trials. Performance on the Flanker task is commonly captured by RT on correct trials. To additionally capture accuracy, we also calculated a "Balanced Integration Score" (BIS), a metric that integrated accuracy and RT and is robust to speed–accuracy trade-offs (Liesefeld & Janczyk, 2019). We, therefore, calculated the correct

RT and BIS for the following: (a) all trials, (b) incongruent (i.e., more difficult) trials only, and (c) the difference between congruent versus incongruent trials. Distributions of these measures are reported in Supplemental Materials.

Using these metrics, we first analyzed whether word learning was associated with Flanker task performance. Specifically, we entered each Flanker performance metric as a fixed effect in a separate Bayesian mixed effects logistic regression model with word learning as the outcome. BIS metrics for incongruent and all trials were associated with word learning (all 95% CI did not contain 0). Neither the difference in BIS between congruent versus incongruent trials nor any of the correct RT metrics were associated with word learning (all 95% CI contained 0). Thus, performance on even an ostensibly unrelated task was associated with word learning, suggesting a role for overall attentiveness/engagement in word learning success.

Next, we analyzed whether word learning was associated with WM capacity above and beyond the BIS metrics of Flanker task performance that were associated with word learning. Specifically, we added WM capacity as a predictor to each BIS metric. WM capacity did indeed predict word learning above and beyond each BIS metric (controlling for BIS on incongruent trials: point estimate = 2.02, 95% CI [0.68, 3.36]; controlling for BIS on all trials: point estimate = 2.01, 95% CI [0.78, 3.33]).

Discussion

The goal of Experiment 3 was to test whether the apparent association between word learning and WM capacity was merely due to more general individual differences, in which individuals who are more motivated or attentive simply perform better on both the word learning and working memory tasks. Indeed, performance on an ostensibly unrelated task (Flanker inhibition) was associated with word learning. Critically, however, WM capacity was associated with word learning above and beyond performance on this ostensibly unrelated task. These results thus reinforce the importance of WM capacity in word learning from context. In the General Discussion section, we return to the specificity of the relationship between WM capacity and word learning from context.

General Discussion

We acquire much of our word knowledge by gleaning the meanings of words from their surrounding language contexts. For example, it is possible to learn that a “rambutan” is a fruit just from encountering it in the context of words associated with fruits such as “juicy” and “sweet.” The goal of the present study was to illuminate the online processes that unfold when learning new words from context. The results provided evidence that these processes draw upon working memory, as greater working memory capacity was associated with more successful word learning. Moreover, gaze dynamics revealed that working memory demands are particularly strong when processing a new word after having encountered an informative context. This result transpired regardless of whether the new word was initially encountered before or after an informative context. Together, these findings point to a key role for working memory in the maintenance and manipulation of the context while linking its implied meaning to a new word, regardless of whether the context precedes or follows the new word.

Implications for Cognitive Mechanisms

To date, there have been only a handful of proposals about the cognitive mechanisms involved in learning words from context (e.g., Borovsky et al., 2012; Chaffin et al., 2001). Moreover, each proposal has focused only on how word learning might unfold in one direction, either when an informative context occurs before (Forward) a new word, or when it occurs after (Backward). In contrast, the present results suggest that an equivalent recruitment of working memory to update the sense of a new word after encountering an informative context, regardless of which one occurs first. When the informative context occurred first and could thus be available in working memory starting from the initial encounter with the new word, participants spent longer initially inspecting the new word, then accumulated further time revisiting the new word after looking back at the context. When the informative context only occurred after the new word, participants spent an extended time revisiting the new word after they had read the context. These tendencies to spend longer looking at a novel word after encountering an informative context were all greater in participants with greater working memory capacities. Together, gaze patterns suggest that word learning from context is a dynamic, possibly iterative process of extracting an implied meaning from an informative context held in working memory and using it to update one’s sense of what a new word means.

These empirical insights suggest that prior Forward and Backward accounts are each insufficient to explain word learning from context because each only captures learning in one direction and overlooks a role for working memory that appears to equate learning across directions. These results could thus help enrich and expand mechanistic accounts of how words are learned from context to encompass the use of working memory in both Forward and Backward routes.

At the same time, it is worth evaluating the specificity of the observed relationship between word learning from context and working memory. Because this relationship was correlational, it is always important to consider whether other factors might account for it. The present findings rule out some obvious alternatives. One alternative is that word learning from context and working memory only appeared to be related because participants who are more motivated or alert are likely to perform better at any tasks that involve some effort. The likelihood of this alternative is undermined by: (a) consistently observing the relationship after removing participants who may have been inattentive based on their performance on a semantic judgment task included in the working memory test and (b) the results of Experiment 3, in which the relationship remained after controlling for performance on an additional unrelated task.

Another alternative is that participants with more language experience and fluency performed better at both the word learning from context and working memory tasks because both involved processing language input (Acheson & MacDonald, 2009). However, we endeavored to mitigate this possibility in our design of stimuli in the word learning from context task. Specifically, we created sentence contexts containing highly familiar, early-learned words that likewise implied that the meanings of novel words were similar to highly familiar, early-learned words. This approach reduced the likelihood that differences in language experience and fluency would represent a large source of variability in word learning from context performance. Nevertheless, because the present studies did not directly control for language knowledge and fluency, a key question for further research will be to examine whether such an alternative factor accounts for the observed relationship between word learning and working memory. Likewise, further research could examine whether the relationship with word learning is specific to working memory for verbal material (as measured here) or generalizes to working memory more broadly.

Contrast With Ambiguity Resolution

In comparison with the minimal prior investigation into the use of context for learning new words, the use of context to disambiguate familiar words has received extensive study. These processes might seem at least superficially similar. For example, just as "juicy" and "sweet" could be used to infer that a new word refers to a fruit, they could also disambiguate that an instance of the word "orange" refers to its fruit meaning rather than its color meaning. However, the present results highlight an important distinction between these apparently similar processes. Studies of disambiguation suggest that it is easier and places less demand on working memory resources when the informative context occurs before an ambiguous word (Duffy et al., 1988; Frazier & Rayner, 1990; Rayner & Frazier, 1989; Samuel, 1991; see also Jesse & McQueen, 2011). In contrast, the present results suggest a parity in word learning from context processes regardless of the order in which the context and familiar word are encountered. This discrepancy might be due to different working memory demands in word learning versus ambiguity resolution. In ambiguity resolution, the greater working memory demands incurred when the informative context occurs after the ambiguous word likely stem from the need to sift between multiple already-activated interpretations of the ambiguous word. In contrast, these demands are absent in word learning because new words (by virtue of being new) do not have multiple interpretations.

Implications for Word Learning During Development

Although word learning takes place throughout the lifespan, it is particularly pronounced during development, when vocabularies grow from zero to thousands of words. At the same time, development is also a period of substantial changes in working memory capacity (Fry & Hale, 2000). There is suggestive prior evidence that the development of working memory contributes to childhood vocabulary growth, such as correlations between children's working memory capacity and vocabulary size (Nilsen & Graham, 2009; Roman et al., 2014; Sesma et al., 2009). The present study reinforces this relationship and highlights future directions for investigating it further. For example, in the present study, even participants with low working memory capacities learned some words from context, suggesting that low working memory capacities may limit but not eliminate word learning in adults. However, given lower overall working memory resources in childhood, might a child with a low working memory capacity relative to other children struggle to pick up even some words from context?

A separate question raised by the present study concerns the observed parity between the Forward and Backward routes for word learning. In the present study, both overall word learning and the effect of working memory were comparable in the Forward and Backward routes. However, it is unclear whether this parity holds throughout development. According to preliminary data collected in our lab, younger (6- to 7-year-old) children may learn words encountered in spoken language contexts more successfully via the Forward versus the Backward route. This discrepancy may reflect developmental changes in the way working memory is recruited to maintain and extract word meanings from informative contexts. If borne out, such findings could have implications for pedagogical texts. For example, if children do indeed learn words more readily via the Forward route, the effectiveness of pedagogical texts could be improved by consistently placing target vocabulary words after informative contexts.

Scope of Word Learning From Context

To illuminate questions surrounding the role of working memory in the Forward and Backward routes to word learning from context, the present study focused on a relatively simple word learning challenge: learning that a new word is similar in meaning to a specific familiar word, such as learning that "fimp" is similar in meaning to "cake" from "The fimp at Jessie's birthday party was delicious." Like the new words in this study, many words that are incorporated into our vocabularies are similar in meaning to a particular word we already know, such as "conflagration" (similar to "fire"), "ornamentation" (similar to "decoration"), and "champion" (similar to "winner"). However, many other words that we learn may represent new members of a group of words similar in meaning. For example, in the "sweet, juicy rambutan" example used above, "rambutan" is similar in meaning to any known words for fruits, rather than one specific known word.

From the present study, we cannot illuminate whether the scope of both the Forward and Backward routes extends to incorporating new words as new members of groups of words similar in meaning. However, there is prior evidence to suggest that this may be the case in the Forward route. As described in the Introduction section, the Forward route is an extension of proposals that language processing involves predicting upcoming words. Studies on this topic indicate that online language processing involves the prediction of multiple probable upcoming words, such as the prediction of multiple fruits upon encountering "sweet" and "juicy," rather than an all-or-none prediction of a single probable word (Brothers et al., 2023; Federmeier & Kutas, 1999; Luke & Christianson, 2016). Thus, when a new word such as "rambutan" occurs after "sweet" and "juicy," it could be linked to the multiple fruit words that have been activated from prediction via the Forward route. In contrast, although mechanistic proposals regarding the Backward route would in principle also support the incorporation of a new word into a group of words similar in meaning, to the best of our knowledge there is no prior evidence that speaks to this possibility. Therefore, the present study can serve as a foundation for future investigations into the scope of the Forward and Backward routes for word learning from context. For example, future research could investigate the breadth of semantic priming produced by words learned via the Forward and Backward routes. Using this approach, future research could test whether new words learned via the Forward and Backward routes exhibit broad patterns of priming multiple known words similar in meaning or become narrowly linked to just one or a few such known words.

The scope of the present study also examined word learning from relatively local contexts, consisting of a single sentence in which a new word was embedded. In everyday life, broader contexts may be relevant to the meaning of a new word, such as the topic of a conversation or passage in which a new word appears. We speculate that working memory may play an even more important role in word learning from broader contexts because the amount of time and intervening language that transpires between an informative context and a new word can be larger, which may place stronger demands on maintaining the context in working memory (e.g., Gunter et al., 2003). Yet the present study cannot speak to this possibility, or to whether the contributions of working memory to word learning from broader contexts might differ in the Forward and Backward routes. These possibilities thus represent a key target for future research.

Finally, the present study evaluated the dynamics of word learning within a relatively small number of sentences (16, split between the Forward and Backward conditions). For generalizability and statistical sensitivity, future research that seeks to replicate or extend the present work would benefit from a larger number and variety of word-learning contexts.

Conclusion

We amass vocabularies containing thousands of words, even without spending hours studying a dictionary or the benefit of a teacher pointing out the meanings of words one by one. Instead, we pick up new words from the everyday language we read and hear. The goal of the present studies was to shed new light on the cognitive processes involved in this feat of learning words from their surrounding language contexts. The results revealed a key role for working memory in which working memory resources are used to maintain the context while linking its implied meaning to a new word, regardless of which is encountered first. These results have implications for cognitive models of word learning and highlight future research directions for better understanding word learning during development.

Constraints on Generality

The present experiments were conducted with adult participants recruited from Western populations, including undergraduate participants, participants recruited from the broader community, and Prolific users. These populations were targeted to recruit participants who could readily comprehend the English-language sentence materials. Thus, it is unknown whether the findings generalize to speakers of other languages. However, this possibility is supported by evidence that languages other than English are also rich in the kind of contextual support for word learning used in this research, in which new words occur in similar contexts to known words similar in meaning (e.g., Grave et al., 2018; Lample & Conneau, 2019). In addition, it is unknown whether the findings generalize to developing or aging populations. As noted above, this is a key target for future research.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50–68. https://doi.org/10.1037/a0014411

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv. https://doi.org/10.48550/arXiv.1506.04967

Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1186–1196. https://doi.org/10.1037/0278-7393.21.5.1186

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition*, *25*(5), 653–660. https://doi.org/10.3758/BF03211306

Borovsky, A., & Elman, J. (2006). Language input and semantic categories: A relation between cognition and early word learning. *Journal of Child Language*, *33*(4), 759–790. https://doi.org/10.1017/S0305000906007574

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*(3), 278–302. https://doi.org/10.1080/15475441.2011.614893

Brothers, T., Morgan, E., Yacovone, A., & Kuperberg, G. (2023). Multiple predictions during language comprehension: Friends, foes, or indifferent companions? *Cognition*, *241*, Article 105602. https://doi.org/10.1016/j.cognition.2023.105602

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*(4), 1520–1523. https://doi.org/10.3758/s13428-016-0811-4

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*(4), 671–681. https://doi.org/10.1037/0022-0663.96.4.671

Carr, J. W. (2023). *Package eyekit*. https://jwcarr.github.io/eyekit/index.html

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 225–235. https://doi.org/10.1037/0278-7393.27.1.225

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18. https://doi.org/10.1016/0749-596X(86)90018-5

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Du Bois, J. W., Chafe, W. L., Meyer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000–2005). *Santa Barbara corpus of spoken American English, parts 1–4*. Linguistic Data Consortium.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Ervin, S. M. (1961). Changes with age in the verbal determinants of word association. *The American Journal of Psychology*, *74*(3), 361–372. https://doi.org/10.2307/1419742

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *Macarthur-bates communicative development inventories*. Paul H. Brookes Publishing.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*(2), 181–200. https://doi.org/10.1016/0749-596X(90)90071-7

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*(1–3), 1–34. https://doi.org/10.1016/S0301-0511(00)00051-X

Gathercole, S. E., & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259–272. https://doi.org/10.1007/BF03174081

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). *Learning word vectors for 157 languages*. arXiv. https://doi.org/10.48550/arXiv.1802.06893

Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience*, *15*(5), 643–657. https://doi.org/10.1162/jocn.2003.15.5.643

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hill, M. S., & Wagovich, S. A. (2020). Word learning from context in school-age children: Relations with language ability and executive function. *Journal of Child Language*, *47*(5), 1006–1029. https://doi.org/10.1017/S0305000919000989

Horst, J. S., & Hout, M. C. (2016). The novel object and unusual name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Hübner, R., & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review*, *17*(6), 821–826. https://doi.org/10.3758/PBR.17.6.821

Ivanov, I., Liu, X., Clerkin, S., Schulz, K., Friston, K., Newcorn, J. H., & Fan, J. (2012). Effects of motivation on reward and attentional networks: An fMRI study. *Brain and Behavior*, *2*(6), 741–753. https://doi.org/10.1002/brb3.80

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, *18*(5), 943–950. https://doi.org/10.3758/s13423-011-0129-2

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. https://doi.org/10.1037/0033-295X.114.1.1

Jugovac, D., & Cavallero, C. (2012). Twenty-four hours of total sleep deprivation selectively impairs attentional networks. *Experimental Psychology*, *59*(3), 115–123. https://doi.org/10.1027/1618-3169/a000133

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122

Klaus, J., & Schriefers, H. (2016, December 7). *Measuring verbal working memory capacity: A reading span task for laboratory and web-based use*. https://doi.org/10.31219/osf.io/nj48x

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (2019). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–237). Psychology Press. https://doi.org/10.4324/9781315792385-11

Lample, G., & Conneau, A. (2019). *Cross-lingual language model pretraining*. arXiv. https://doi.org/10.48550/arXiv.1901.07291

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed–accuracy trade-offs(?). *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/s13428-018-1076-x

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/BF03204766

MacWhinney, B. (2000). *The CHILDES project: The database*. Psychology Press.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Martella, D., Casagrande, M., & Lupiáñez, J. (2011). Alerting, orienting and executive control: The effects of sleep deprivation on attentional networks. *Experimental Brain Research*, *210*(1), 81–89. https://doi.org/10.1007/s00221-011-2605-3

Melamud, O., Goldberger, J., & Dagan, I. (2016). context2vec: Learning generic context embedding with bidirectional LSTM. *Proceedings of the 20th SIGNLL conference on computational natural language learning* (pp. 51–61). Association for Computational Linguistics.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. arXiv. https://doi.org/10.48550/arXiv.1301.3781

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1–28. https://doi.org/10.1080/01690969108406936

Miyake, A., Just, M. A., & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, *33*(2), 175–202. https://doi.org/10.1006/jmla.1994.1009

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*(2), 237–270. https://doi.org/10.3102/00028312024002237

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B*, *375*(1791), Article 20180522. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., & Van Berkum, J. J. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155–167. https://doi.org/10.1016/j.brainres.2006.08.022

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–249. https://doi.org/10.1016/j.cogpsych.2008.07.002

Onnis, L., Lim, A., Cheung, S., & Huettig, F. (2022). Is the mind inherently predicting? Exploring forward and backward looking in language processing. *Cognitive Science*, *46*(10), Article e13201. https://doi.org/10.1111/cogs.13201

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics.

Perez, M. M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in*

Second Language Acquisition, *42*(4), 749–773. https://doi.org/10.1017/S0272263119000706

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255. https://doi.org/10.1207/s1532799xssr1003_3

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(5), 779–790. https://doi.org/10.1037/0278-7393.15.5.779

Roman, A. S., Pisoni, D. B., & Kronenberger, W. G. (2014). Assessment of working memory capacity in preschool children using the missing scan task. *Infant and Child Development*, *23*(6), 575–587. https://doi.org/10.1002/icd.1849

Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*, *8*(10), 627–633. https://doi.org/10.1145/365628.365657

Samuel, A. G. (1991). A further examination of attentional effects in the phonemic restoration illusion. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *43*(3), 679–699. https://doi.org/10.1080/14640749108400992

Savic, O., Unger, L., & Sloutsky, V. M. (2022). Exposure to co-occurrence regularities in language drives semantic integration of new words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(7), 1064–1081. https://doi.org/10.1037/xlm0001122

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology*, *15*(3), 232–246. https://doi.org/10.1080/09297040802220029

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Unger, L., & Sloutsky, V. M. (2024). *Dynamic processes of word learning from context* [Conference session]. Proceedings of the 46th Annual Meeting of the Cognitive Science Society, Rotterdam, Netherlands.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*(3), 238–242. https://doi.org/10.1111/j.1467-9280.1997.tb00418.x

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2020). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models* (R package Version 2.4.1) [Computer software]. https://mc-stan.org/loo/

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506–2516. https://doi.org/10.1093/cercor/bhv075

Witzel, J., & Forster, K. (2014). Lexical co-occurrence and ambiguity resolution. *Language, Cognition and Neuroscience*, *29*(2), 158–185. https://doi.org/10.1080/01690965.2012.748925

Yamaguchi, M., & Nishimura, A. (2019). Modulating proactive cognitive control by reward: Differential anticipatory effects of performance-contingent and non-contingent rewards. *Psychological Research*, *83*(2), 258–274. https://doi.org/10.1007/s00426-018-1027-2

Yang, X., Zhang, X., Zhang, Y., Zhang, Q., & Li, X. (2020). How working memory capacity modulates the time course of semantic integration at sentence and discourse level. *Neuropsychologia*, *140*, Article 107383. https://doi.org/10.1016/j.neuropsychologia.2020.107383

Received August 7, 2024
Revision received February 21, 2025
Accepted March 12, 2025 ■