Dynamic Processes of Learning Words from Context

Layla Unger (layla.unger@york.ac.uk)
Department of Psychology, University of York &
Department of Psychology, Ohio State University

Vladimir M Sloutsky (sloutsky@psy.ohio-state.edu)
Department of Psychology, Ohio State University

Abstract

Often the only source of information for learning a word is its surrounding language context. For example, even without seeing a rambutan, one can learn that it is a fruit just from hearing "I like sweet, juicy rambutans". What processes foster learning words from context? We investigated candidate processes that can unfold when the context precedes a new word and can foster learning via prediction, versus when the context occurs after and can only be used retroactively. We particularly sought to illuminate a role for working memory in linking a new word to the meaning implied by its context. Experiment 1 probed word learning during reading with eye tracking, and Experiment 2 probed word learning from speech. We found convergent evidence that regardless of whether the context precedes or follows a new word, word learning depends on maintaining the context in working memory while linking it to a new word.

Keywords: word learning; working memory; reading; eye tracking

Introduction

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 development, when young children often pick up their first words by mapping them to objects and other observable referents. Children harness a suite of cognitive processes to map words to referents, from following a speaker's gaze to tracking the objects that are consistently present when a word is heard (e.g., Smith & Yu, 2008; Tomasello & Farrar, 1986; Markman & Wachtel, 1988). Yet, for the thousands of words that are typically learned across a lifetime, often the only initial 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 word learning processes that unfold upon encountering a new word in language.

Here we first outline candidate cognitive processes that have been proposed either as computational or conceptual models. We then highlight key unknowns about these processes, with an emphasis on an underexplored role of working memory. 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, yet they must be linked for successful word learning. Finally, we present two studies designed to shed new light on the dynamics that unfold upon encountering a new word in language and their relationship with working memory.

Mechanistic Processes for Learning Words from Context

The candidate processes for learning words from context either explicitly or implicitly invoke the distributional hypothesis (Harris, 1954; Landauer & Dumais, 1997; Miller & Charles, 1991; Rubenstein & Goodenough, 1965), which posits that words similar in meaning occur in similar language contexts. For example, different words for fruits often occur in the context of "juicy" and "sweet" (e.g., Landauer & Dumais, 1997; Mikolov, Chen, Corrado, & Dean, 2013; Jones & Mewhort, 2007; Lund & Burgess, 1996). 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". However, the distributional hypothesis does not speak to the cognitive processes that link new words to known words from distributional regularities. Here, we evaluate the landscape of candidate processes.

One candidate process comes from the popular proposal that processing language input involves prediction, including the prediction of upcoming words (DeLong, Urbach, & Kutas, 2005; Kutas, Lindamood, & Hillyard, 2019; Van Petten & Luka, 2012; Willems, Frank, Nijhof, Hagoort, & Van den Bosch, 2016; Nieuwland et al., 2020). Importantly, the actually observed words serve as error feedback signals to improve predictions in the future. Building on this idea, some researchers have proposed that when a predicted word turns out to be incorrect, it becomes linked to the observed word (Borovsky, Elman, & Kutas, 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, so that fruit words are active when a novel word such as "rambutan" occurs instead. This co-activation links the familiar and novel words. 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"? In contrast with the Forward route, the new word must somehow be linked to known words likely to have *preceded* the informative context. We refer to candidate explanations of this process as Backward accounts. One Backward account is simply the inverse of the Forward account: that new and known words become co-active because processing language

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

involves making backward retrodictions of words likely to have preceded those currently observed (Onnis, Lim, Cheung, & Huettig, 2022; Chaffin, Morris, & Seely, 2001). Another account 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., Elman, 1990; Borovsky & Elman, 2006). Like the Forward account, these models also invoke the prediction of upcoming words, but use error feedback in a different way: Upon encountering a word that was not well predicted, error feeds back to update the representations of preceding words to improve future predictions. Thus, encountering "sweet" and "juicy" after "rambutan" updates a representation of "rambutan" to better predict these words, which will become similar to the representations of known fruit words that have already been formed to predict "sweet" and "juicy".

Together, the Forward and Backward accounts outline processes that might lead to word learning from distributional regularities. Yet, 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. Evidence suggestive of differences comes from studies of another language comprehension challenge: resolving ambiguities. Ambiguities are common in language, such as when a word has multiple meanings or senses (e.g., "ball" can refer to a toy or formal dance). As in determining the meaning of a novel word, the surrounding language context can inform the interpretation of an ambiguous familiar word (e.g., when "ball" occurs in a context relating to a game versus a formal event). 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; Gilbert, Davis, Gaskell, & Rodd, 2021; Samuel, 1991; Jesse & McQueen, 2011). However, the difficulties posed by Backward ambiguity resolution likely come from choosing between multiple activated interpretations, and possibly overcoming a dominant interpretation in favor of a weaker one (e.g., interpreting "ball" as a formal dance). Given that novel words do not have any known interpretations, the relative difficulties of Forward and Backward routes in word learning may not parallel those in ambiguity resolution.

Another key unknown regarding the Forward and Backward routes comes from the fact that both involve linking a new word that is encountered at one 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, this possibility is not explicitly incorporated across accounts and is little studied.

Working Memory in Learning Words from Context

The possibility that working memory is important for word learning from context is consistent with Just and Carpenter's (1992) influential proposal that working memory capacity is needed to store and interpret recently encountered language, and to use these contents to anticipate and interpret upcoming language input. Both the Forward and Backward routes may require such working memory involvement. However, direct evidence regarding a role for working memory consists mainly of a handful of correlations between working memory span and the success of word learning via the equivalent of the Forward route (Daneman & Green, 1986; Hill & Wagovich, 2020), as well as correlations between working memory span and vocabulary size (Roman, Pisoni, & Kronenberger, 2014; Sesma, Mahone, Levine, Eason, & Cutting, 2009; Nilsen & Graham, 2009). Moreover, the dynamics with which working memory is recruited when learning words via the Forward and Background routes remain unclear. 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 multi-word section of text is indicative of the degree to which the reader is actively drawing upon the storage or manipulation aspects of working memory during language processing. For example, ambiguous words with multiple equally common meanings are inspected longer than words with a single dominant, common meaning (Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986). Moreover, this effect is reduced when a preceding context strongly cues a single meaning for the ambiguous word (Duffy et al., 1988; Binder & Morris, 1995). Thus, gaze duration reflects demands to maintain multiple interpretations in working memory. In the realm of word leaning, Chaffin et al. (2001) found that readers spent a longer time revisiting novel versus familiar words after encountering a context that related to the word's meaning, suggesting that readers were using working memory resources to maintain contextual information in mind while updating their sense of the novel word. Thus, gaze durations reflect the storage and manipulation demands on the working memory processes putatively involved in language comprehension in general and word learning in particular. We next outline how gaze to novel words and informative contexts can illuminate the dynamics of these processes during word learning.

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 an a context that is informative about the novel word's meaning. When a novel word is read prior to reading an informative context, the reader does not yet have information about what the new word means, so gaze duration likely indicates just the effort involved in storing its word form in working memory.

Encounters with a new word *after* an informative context has been read are likely more indicative of the degree to which the reader is using contextual information in working memory to update their sense of what the new word might mean. The point at which this encounter first takes place differs in the Forward and Backward routes. In the Forward route, these encounters start with the first time the new word is read. In the Backward route, the context is not yet available at the first encounter with the new word. Instead, encounters with the new word after the informative context only take place if and when the reader *revisits* the new word after they have read the context. Accordingly, any gaze duration differences during these encounters between Forward and Backward routes can capture 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. When the context is read 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 Study

The goal of the present study was to illuminate the online processes that unfold upon encountering an opportunity to learn a new word from its surrounding context. We contrasted 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 study was designed to illuminate the processes of maintaining and manipulating information in working memory.

For this study, 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). In Experiment 1, we used eye tracking to illuminate the dynamics of processing these sentences in individuals with high versus low working memory capacities. 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. 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. Experiment 2 thus 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.

Experiment 1

Methods

Participants

The sample consisted of 78 adults (following exclusion of 8 participants due to $\geq$ 40% eye tracking trackloss) recruited from the undergraduate population at a large US university and the surrounding city. Participation was compensated with course credit or a $10 gift card.

Materials

Sentences. The primary materials consisted of sentences generated using the following criteria. Each sentence contained one novel word accompanied by context words that are familiar to adults and informative about the novel word's meaning. 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.

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". All sentences were normed with a separate sample (N = 31) using a sentence-completion or "cloze" task in which participants saw either the Forward or Backward version of each sentence, with the novel word replaced with a blank space, and were prompted to complete the blank space with a familiar word. From a larger set, we selected only sentences where more than 85% of participants entered the target familiar word for both the Forward and Backward versions. In addition, 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, moderately strong evidence for equivalence). Finally, we calculated a measure (Positive Pointwise Mutual Information, PPMI) of the regularities with which the words that co-occurred in sentences also tend to co-occurr in corpora of everyday language input, and ensured that these regularities were similar in Forward and Backward versions (PPMI of co-occurrences in in Forward: M = 1.25, SD = 1.77; Backward: M = 1.16, SD = 1.66) (t-test Bayes Factor = 0.142, moderately strong evidence for equivalence). Finally, we generated two lists of sentences (randomly assigned to participants), each with 8 Forward and 8 Backward sentence versions.

Pictures. The word learning task used pictures that depicted the target meanings of novel words.

Working Memory Span. We assessed working memory using the extensively used reading span task originally developed by Daneman and Carpenter (1980) and adapted for open-source use by Klaus and Schriefers (2016). This assessment consists of judging sentences for semantic coherence, and storing words for later recall. Materials for this assessment consist of: (1) 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 (2) individual high-frequency words (e.g., "table").

Apparatus. Gaze data were collected using an EyeLink Portable Duo eye tracking system with a sampling rate of 500Hz. Responses were made using a gamepad controller.

Procedure

Participants first completed a word learning from context task with eye tracking which included Exposure and Test phases. 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. Within a block, participants read four sentences (half Forward, half Backward in a random order), then were tested on the four novel words in these sentences. Reading was self-paced. During Test trials, participants were presented with a single novel word and matched it to one of two pictures: one showing the target meaning of the novel word, and one showing the target meaning of another novel word from the same block. The same two pictures were always presented together.

Participants then completed the working memory assessment, which consisted of Daneman and Carpenter's 1980 reading span task, as adapted by Klaus and Schriefers (2016) for open-access use. In this task, participants alternate between (1) judging sentences for semantic coherence (self-paced) and (2) storing an individual word for later recall (shown for 1200ms). After a block consisting of 2 to 6 sentence-word pairs, participants recalled as many of the storage words as they could remember (typed by an experimenter). The task included one block for each span size ranging from 2 to 6 in a random order.

Results

We first scored performance on the working memory assessment (the average proportion of words correctly recalled following Conway et al., 2005) and used a median split to divide participants into High and Low working memory (henceforth WM) span groups. Note that this split simplifies the interpretation of patterns (particularly the patterns observed in gaze dynamics), but we found equivalent results when treating WM as a continuous variable.

We then contrasted Forward and Backward routes in participants with High versus Low WM spans separately for word learning and gaze dynamics. Analyses were conducted as mixed effects models in R using the lme4 and car packages (Bates, Maechler, Bolker, & Walker, 2015; Fox & Weisberg, 2011). Following Bates, Kliegl, Vasishth, and Baayen (2015), we selected mixed effects models by starting with maximal models and eliminating random effects that did not contribute to the goodness of fit. Replicating analyses with Bayesian versions of mixed effects models yielded equivalent results (not reported here).

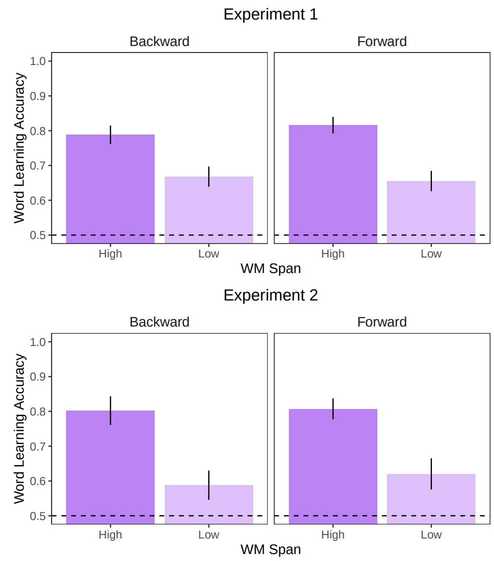

Figure 1: Word learning in Experiments 1 and 2. Graphs depict differences in word learning performance between participants with High and Low WM span in the Forward and Backward Route conditions. Error bars depict standard errors of the mean.

Word Learning

We analyzed the effects of Route (within-subjects, Forward versus Backward) and WM span (between-subjects, High versus Low) on word learning accuracy (see Figure 1). This revealed a main effect of WM span ($\chi^2(1) = 5.115$, $p = .024$), in which word learning was more successful in participants with High versus Low WM span. Neither Route nor its interaction with WM span were significant (*p*s $> .5$). Thus, WM span contributed to word learning success in both the Forward and Backward routes. We checked whether this result might instead have arisen because some participants were overall inattentive during the experiment, leading to poor word learning and WM assessment performance. We assessed attentiveness from the processing component of the WM task, on which performance was overall very high (95%), but ranged from 68 – 100%, indicating inattentiveness in some participants. Results were unchanged by the removal of participants with $<$ 90% accuracy (N = 7).

Illuminating the Role of Working Memory with Gaze Dynamics

Although the analysis of word learning accuracy implicates WM in word learning from context, it 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

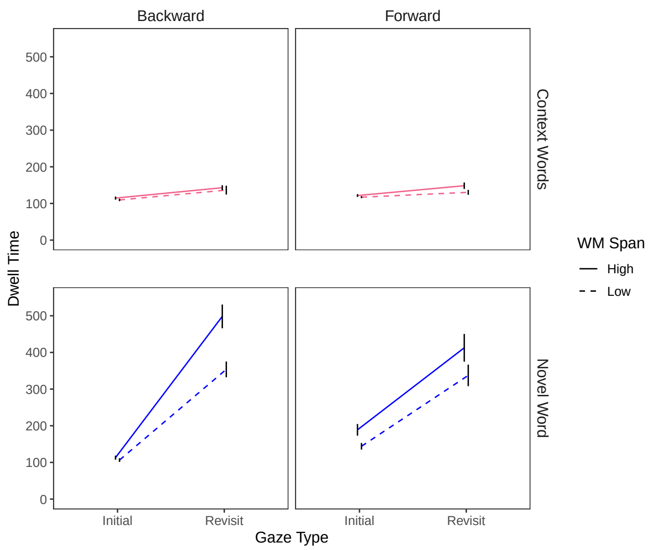

Figure 2: Gaze dynamics in participants High and Low WM span in the Forward and Backward Route conditions. Error bars depict standard errors of the mean.

WM demands. As described in the Introduction, gaze durations can thus reveal whether WM demands are implicated during: (1) an initial encounter with the novel word or informative context, or (2) an encounter with the novel word or informative context after the other has been read.

To investigate the use of WM during word learning from context, we first removed trials with trackloss of more than 40% (1.9% of trials), then defined AOIs for the words in each sentence and divided them into: (1) the AOI for the novel word (henceforth, "Novel"), and (2) AOIs for each informative context word (henceforth, "Context"). For each AOI, we calculated: (1) "Initial" gaze duration, starting from the first look to the AOI and ending with the first look away from it, and (2) "Revisit" gaze duration, including the total amount of time spent looking at the AOI after the Initial period. This process yielded Initial and Revisit gaze durations in the Forward and Backward conditions for Novel and Context words for participants with High and Low working memory span, shown in Figure 2. As described in the Introduction, gaze to the Novel and Context words can provide different insights into the processes recruited during word learning. We therefore conducted the same omnibus analysis separately for gaze to Novel and Context words with factors of: (1) Route condition (within subjects, Forward versus Backward), (2) Gaze type (within subjects, Initial versus Revisit), and (3) WM span (between subjects, High versus Low).

Gaze to Novel Word. The omnibus analysis revealed multiple main effects (of Gaze Type and WM span), two-way interactions (between Gaze Type and Route condition and between Gaze Type and WM span), and a nearly-significant three-way interaction between Gaze Type, Route condition and WM span (p = .05). We teased apart these interactions using separate analyses for Initial and Revisit gaze durations. To help the reader, we first summarize the overall pattern: participants showed a tendency to spend longer reading the novel word after they had read the context that was greater in participants with High WM.

Within Initial gaze, there was a main effect of Route condition ($\chi^2(1) = 35.012$, $p < .0001$), in which participants looked longer at the novel word in the Forward condition – i.e., when it was encountered *after* the context – than in the Backward condition – i.e., when it was encountered *before* the context. This main effect was qualified by an interaction between Route condition and WM span ($\chi^2(1) = 4.154$, $p = .042$) in which the tendency to spend longer looking at the novel word in the Forward condition was stronger in individuals with High WM. Thus, participants spent more time initially reading the novel word when they encountered it after an informative context. This tendency was greater in individuals with higher WM span.

Within Revisit gaze, there were main effects of Route condition ($\chi^2(1) = 4.205$, $p = .040$) and WM span ($\chi^2(1) = 3.876$, $p = .049$). The main effect of Route condition indicated that participants spent longer revisiting the novel word in the Backward than in the Forward condition. This result complements the results for Initial gaze. Within Initial gaze, participants spent longer initially looking at the novel word when they had already encountered the context (Forward condition) than when they had not (Backward condition). Within Revisit gaze, participants appeared to compensate for the fact that they had not encountered the context prior to initially reading the novel word in the Backward condition by spending longer revisiting it. Thus, participants emphasized looking at the novel word after having read the context, regardless of which they encountered first. The main effect of WM span indicated that participants with High WM spent longer revisiting the novel word after having read the context in both conditions. There was no interaction between Route condition and WM span (ps $>$ .23). Thus, as in the analysis of Initial gaze, individuals with higher WM span spent longer reading the novel word after having read the context.

Gaze to Context. The omnibus analysis only that participants spent longer revisiting versus initially reading context words ($\chi^2(1) = 24.153$, $p < .0001$).

Discussion

Overall, individuals with greater WM span were more successful at learning words from context, both in the Forward route (when the informative context can be used to form predictions that support word learning), and in the Backward route (when the informative context can only be used retroactively). Gaze dynamics indicate that WM was most strongly recruited for processing the novel word after having read the context, a tendency that was greater in participants with higher WM span. Together, this pattern suggests the important role of WM in maintaining and manipulating the context while linking its implied meaning to a novel word.

Experiment 2 tested whether these patterns might be idiosyncratic to reading, when it is possible to look back and forth between novel words and their surrounding contexts.

For example, the opportunity to look back and forth during reading might have contributed to the equivalence of Forward and Backward routes in word learning. Alternatively, these gaze dynamics might have an analogue in processing spoken language in which the listener thinks back to a word they recently heard. Therefore, Experiment 2 was a replication of Experiment 1, except that participants listened to prerecorded versions of the sentences.

Experiment 2

Methods

Participants were 51 adults recruited from the same population as Experiment 1. Materials and Procedure were similar to Experiment 1, with the exception that participants listened to prerecorded sentences (with the same recording of the novel word used in Forward and Backward versions of a sentence to avoid differences in articulation) and gaze was not tracked.

Results and Discussion

The same analysis of word learning as conducted for Experiment 1 revealed analogous results, in which word learning was more successful in participants with High vs. Low WM span ($\chi^2(1) = 16.037$, $p < .0001$; replicated with removal of participants who were $< 90\%$ accurate on the semantic coherence task). The replication of Experiment 1 implicates WM span similarly in reading and spoken modalities.

General Discussion

The goal of the present study was to illuminate the online processes that unfold when learning new words from context, from which much of our word knowledge is acquired. The results provided evidence that these processes draw upon working memory, as greater working memory span 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 processed an informative context, regardless of which was initially encountered first. 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.

The present findings have key implications for mechanistic accounts of word learning from context. The existing handful of accounts (e.g., Borovsky et al., 2012; Chaffin et al., 2001) each focus only on how word learning might unfold in either the Forward or Backward route. In contrast, the present results suggest that learning words from context involves using 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 was thus already available in working memory from the initial encounter with the new word, participants spent longer initially inspecting the new word, then spent 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 patterns suggest that word learning from context is a dynamic, possibly iterative process of extracting an implied meaning from an informative context and using it to update one's sense of what a new word means. These insights could help enrich and expand mechanistic accounts of word learning from context to encompass the use of working memory in both Forward and Backward routes.

Contrast with Ambiguity Resolution

In comparison with the minimal prior investigation into the use of context in word learning, the use of context to disambiguate familiar words has received extensive study. These processes appear similar: for example, just as "juicy" and "sweet" could be used to learn that a new word refers to a fruit, they could also disambiguate that an instance of the word "orange" refers to its fruit rather than its color meaning. However, the present results highlight an important distinction. Studies of disambiguation suggest that it is easier and places less demand on working memory when the informative context precedes an ambiguous word (Frazier & Rayner, 1990; Rayner & Frazier, 1989; Gilbert et al., 2021; Samuel, 1991; Jesse & McQueen, 2011; Duffy et al., 1988). In contrast, the present results suggest a parity in word learning from context processes regardless of the order in which the context and new 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 follows the ambiguous word likely stem from the need to reweight or reject already-activated interpretations. In contrast, these demands are absent in word learning because new words have no existing interpretations.

Role of Working Memory

In the present research, we have focused on interpreting a role for working memory that involves maintaining an informative context in mind while processing a new word. This interpretation rests on both: (1) gaze durations, in which longer gaze durations are indicative of greater use of working memory resources (Duffy et al., 1988; Rayner & Duffy, 1986), and (2) different patterns of gaze and word learning associated with individual differences in working memory span.

However, it is important to consider whether the effects associated with individual differences in working memory span are actually due to some other source of individual differences. In the present studies, we addressed the possibility that these effects were due to overall attentiveness / lack of motivation within the study session. Specifically, the effects remained even after removing participants whose performance on one component of the working memory task indicated inattentiveness. Another possibility is that effects were due to individual differences in linguistic knowledge (Acheson & MacDonald, 2009). For example, the better word learning in participants with higher working memory spans might instead come from these participants' better knowledge of the words in the informative contexts. However, we consider this alternative unlikely because the sentence contexts were designed to be used for both the present studies and future developmental studies with children, and therefore contain common, early-learned words likely to be highly familiar to all adult participants in the present studies. Nonetheless, the soundness of the conclusions reached in the present research could be tested by investigating whether other individual differences might account for the effects observed here.

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 a period of substantial changes in working memory span (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 span and vocabulary size (Roman et al., 2014; Sesma et al., 2009; Nilsen & Graham, 2009). Our findings reinforce this relationship and highlight 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?

Conclusion

We pick up new words from the everyday language we read and hear. The present studies aimed was to shed new light on the cognitive processes involved in this feat 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.

Acknowledgements

This work was supported by National Institutes of Health Grant R01HD078545 awarded to Vladimir M. Sloutsky.

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.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

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.

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.

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.

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.

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*, 769–786.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of memory and language*, *25*(1), 1–18.

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.

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.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *The American Journal of Psychology*, *74*(3), 361-372.

Fox, J., & Weisberg, S. (2011). *An r companion to applied regression* (Second Edition ed.). Thousand Oaks, CA: Sage.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of memory and language*, *29*(2), 181–200.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology*, *54*(1-3), 1-34.

Gilbert, R. A., Davis, M. H., Gaskell, M. G., & Rodd, J. M. (2021). The relationship between sentence comprehension and lexical-semantic retuning. *Journal of Memory and Language*, *116*, 104188.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146-162.

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.

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic bulletin & Review*, *18*, 943–950.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1-37.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological review*, *99*(1), 122-148.

Klaus, J., & Schriefers, H. (2016). Measuring verbal working memory capacity: A reading span task for laboratory and web-based use.

Kutas, M., Lindamood, T. E., & Hillyard, S. A. (2019). Word expectancy and event-related brain potentials during sentence processing. In *Preparatory states and processes* (pp. 217–237). Psychology Press.

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.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203-208.

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.

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

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1-28.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., ... others (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), 20180522.

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology*, *58*(2), 220–249.

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), e13201.

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.

Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *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.

Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*, *8*(10), 627-633.

Samuel, A. G. (1991). A further examination of attentional effects in the phonemic restoration illusion. *The Quarterly Journal of Experimental Psychology Section A*, *43*(3), 679–699.

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.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 1454–1463.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and erp components. *International Journal of Psychophysiology*, *83*(2), 176–190.

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.