False Memory of Grammatical Constructions: Evidence for Structured Construction-Level Generalizations

Abigail Fergus (abby.fergus@princeton.edu)[1] & Adele E Goldberg[1]
[1]Department of Psychology, Princeton University

Abstract

Three experiments and a control task combine to indicate that grammatical constructions are implicitly represented in memory, with prototypicality emerging from distributional experience, alongside significant verbatim memory. Participants are exposed to instances of a construction and, following a delay filled by an unrelated language task, falsely endorse novel instances of the same construction more often than paraphrases in a recognition task. Critically, false memories are particularly common for lures containing an unwitnessed word that is distributionally prototypical of the construction, compared to frequency-matched controls. A control study excludes the possibility that the latter effect is simply a word-level effect. Finally, we simultaneously find memory for specific exemplars. Overall performance is above chance yet implicit: participants underestimated their accuracy, which was unrelated to age or education.

Keywords: generalization; false memory; grammatical constructions

Introduction

A central question in the study of language is how people generalize beyond the specific utterances they encounter. Human learners routinely extract abstract grammatical knowledge from limited and noisy input, enabling them to comprehend and produce novel sentences they have never heard before. Usage-based and constructionist approaches propose that such generalizations are induced from experience with form–meaning pairings, or grammatical constructions (Bybee, 2010; Goldberg, 1995, 2006; Tomasello, 2003), yet direct evidence that people spontaneously and implicitly generalize to the level of constructions—without explicit instruction or awareness—remains limited.

Evidence exists that instances of a construction prime or activate other instances of the same construction (Chang et al., 2003; Mahowald et al., 2016; Ziegler, et al., 2018, 2019); and constructions filled with nonsense open-class words have been found to prime words semantically related to the constructions' meanings (Johnson & Goldberg, 2013). But these methods don't lend themselves to testing whether instances of constructions leave memory traces that cluster in long-term memory, perhaps involving a centroid type of instance emerging from distributional evidence.

Here we capitalize on the illusory memory paradigm, which demonstrates that people often falsely remember items that were never presented if those items are strongly supported by associations with encountered items (the DRM paradigm: Deese, 1959; Roediger & McDermott, 1995). In the classic DRM design, participants are presented with a list of words that are closely semantically related (e.g. *nap*, *pillow*, *night*), followed by a delay. When then queried about an unwitnessed associate, (e.g. *sleep*), they tend to falsely believe it had been witnessed. These illusory memories are recognized to reflect normal processes of 'gist' extraction rather than memory failure (Brainerd & Reyna, 2002; Gallo, 2010).

Here we test whether grammatical constructions, which are abstract pairings of form and function, give rise to illusory memories, analogous to the false memories of gist lures in the DRM paradigm. If constructions are mentally represented as clusters of pairings of form and function, learned from distribution (Goldberg, 1995, 2006) then exposure to a relatively small number of instances of a construction should result in the activation of a cluster representation in long-term memory. If so, we expect participants to have more difficulty distinguishing witnessed instances from unwitnessed instances of the same construction compared to paraphrases that employ an alternative construction, when open-class words and propositional content are controlled for.

The key ("head") words in argument structure constructions tend to be skewed in distribution, such that one head word accounts for more tokens than any other, with productive constructions including a long tail of low-frequency head words (Boas, 2003; Goldberg et al., 2004; Yi et al., 2019). We refer to the head word with the highest token frequency in each construction as "construction (cx-) prototypical." If constructions include statistical distributional information such that imperfect or "lossy" memories of a construction containing the cx-prototypical head word overlap more with memories of other instances of the construction, then a lure with a cx-prototypical head word should be more likely to elicit illusory memories than a lure containing a head word that is less prototypical for the construction, even when head words are matched for general frequency in the language.

We also hypothesize that structured representations at the level of constructions are implicit. We test this by asking people to estimate their accuracy on the task. We also test whether memory for constructions varies with age or education. Stimuli, data, and models for all experiments and preregistration for Exp. 2-3 can be found here. Preregistration for Exp. 1 can be found here.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

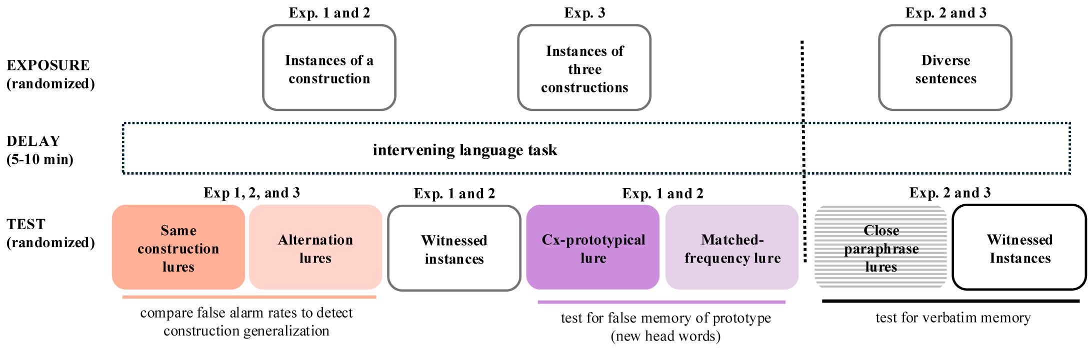

Figure 1: Design of experiments 1, 2 and 3.

Table 1: Example stimuli used in Exps. 1, 2, and control study

Target constructions (*example*)	Unwitnessed prototypical keyword	Frequency-matched lure words	Key words witnessed in sentences during exposure
Double-object construction e.g., *She rented him an apartment.*	*give*	*make, take*	*bake, build, buy, offer, read, rent, serve, show, steal, teach, tell, write*
Gossip construction e.g., *It's silly of them to stay here.*	*nice*	*big, good*	*beautiful, bold, cruel, foolish, gutsy, irresponsible, rich, silly, smart, sweet, thoughtful, wise*
Way construction e.g., *She clawed her way to the front.*	*make*	*find, work*	*claw, climb, creep, drag, elbow, fight, inch, pull, push, shuffle, squeeze, storm*

Table 2: Target (*example*) and Alternative (*example lure*) constructions. First six constructions counterbalanced in Exp. 3

Double-object construction *e.g., She rented him an apartment.*	Prepositional dative construction *e.g., She rented an apartment to him.*
Gossip construction *e.g., It was thoughtful of them to send a card.*	Adjective predication construction *e.g., It was thoughtful that they sent a card.*
Xer, the Yer construction *e.g., The greener your thumb, the more your plants will thrive.*	Conditional construction *e.g., Your plants will thrive when you have a greener thumb.*
Way construction *e.g., They inched their way across the room.*	Intransitive motion construction *e.g., They inched across the room.*

Exp. 1: Constructions

In experiment 1, we adapt the DRM paradigm to determine whether instances of a construction are associated with other instances of the same construction in long-term memory. We additionally test whether prototypical instances of a construction are particularly strong associates. For present purposes, we refer to any instance of a construction that contains the most frequently occurring head word in the construction, a prototypical instance of the construction (Ellis & O'Donnell, 2014; Goldberg et al., 2004). For example, 40% of all instances of the double-object construction include the verb *give* (e.g., *She gave him a book*). We therefore refer to instances of the double-object construction that include *give* as prototypical. See Table 1.

Method

We recruited 159 participants from Prolific (93 F, 63 M, 2 non-binary, 1 preferred not to say), who were paid for their time. *M* age = 35.6.

Each participant witnessed 12 instances of a randomly assigned construction, presented in random order. To make the study more naturalistic, participants were not warned that their memory would be tested. Participants then took part in a 5-10-minute unrelated language task, before engaging in the key recognition task: participants had to decide whether they had witnessed a randomly ordered combination of familiar sentences and four types of lure sentences from Figure 1:

1) "Same construction" lures: recombined witnessed words to form new instances of the witnessed construction.
2) "Alternation" lures: paraphrases of each same-construction lure, using same open-class words in an alternative (unwitnessed) construction (see Table 2).
3) "Cx-prototypical" lures: instances of witnessed construction with prototypical key word (unwitnessed); other open-class words had been witnessed.
4) "Matched frequency" lures: witnessed construction with an unwitnessed verb, which was as frequent overall as the prototypical key word.

Same construction and alternation lures were counterbalanced across participants, so no one saw both an instance of a construction and its paraphrase. Prototype and matched frequency lures were also counterbalanced, and no one saw both a cx-prototypical instance and a frequency-matched lure. This allowed us to compare the two, which only differed in the head word.

As shown in Table 1, three constructions were included: the double-object construction (Bresnan et al., 2007), the gossip construction (Goldberg & Herbst, 2021) and the *way*-construction (Israel, 1996; Jackendoff, 1990). Prototypical head words and frequency matched head words are specified in the 2nd and 3rd columns of Table 1, as are all other head words included during exposure (4th column). For example, a double object sentence lure with *give* is compared to double-object sentences containing the main verbs, *make* or *take*. Both *make* and *take* are more frequent than *give* overall, but they are each less frequent in the double-object construction than *give.* If participants show greater false memory for a new instance of the double-object construction with *give* than a new instance with *make* or *take,* it suggests that participants spontaneously generalize the witnessed instances of the double-object construction to include a prototypical instance.

To summarize, we hypothesize that participants will display a greater tendency to falsely remember certain types of unwitnessed sentences (lures) than others. In particular:

(a) Same-construction lures > alternation lures (familiar open class words in both)

(b) Cx-prototypical lure > frequency-matched lure (new key words in both)

Results

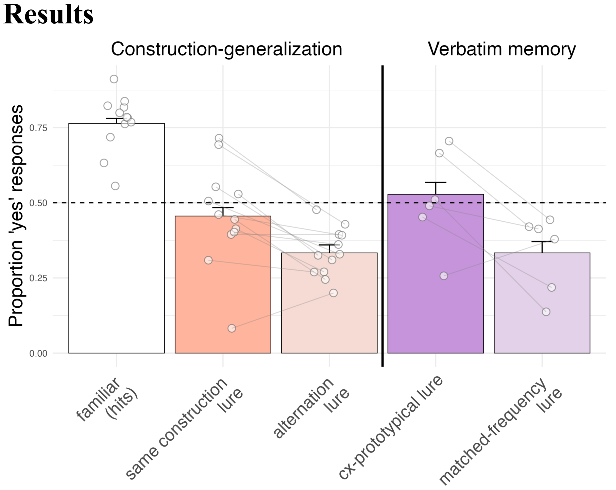

Figure 2: Exp. 1: Proportion of 'yes' (familiar) responses to witnessed sentences (white: hits/(hits + misses)) and each of the four lure types: same-construction vs. alternation lures (orange); cx-prototypical vs. freq. matched lures (purple). Points are items; segments indicate comparisons b/w items.

Overall Accuracy Using signal detection theory, we calculated *d'* for each participant: i.e., the normalized probability of saying yes to familiar items minus the normalized probability of saying yes to new items. Participants showed above chance sensitivity, with mean d' = 0.35, *CI* [0.30, 0.40]. A nonparametric Wilcoxon signed-rank test confirmed that d' values were significantly greater than zero (V = 10,276, $p < .0001$) despite the delay and lack of warning that a memory test was to be given.

False Memory for Same-Construction Lures Figure 2 (comparison of orange bars) also displays that participants had higher false alarm rates–lower accuracy–for same-construction lures than alternation lures despite sharing the same propositional content. This was confirmed with a generalized linear model that predicted response by condition, with random intercepts for participants, items, and constructions (β = -0.58, z = -2.45, $p < 0.02$).

False Memory for Cx-Prototypical Lures Comparing lures that included the prototypical head word (deep purple) with a frequency matched head word (light purple), the same model (with same random effects structure) confirms more illusory memories—lower accuracy—for the prototype lures (β = -0.81, z = -2.42, $p < 0.02$).

Self-rated Accuracy Participants underestimated their own accuracy ($M_{estimated\_acc}$ = .44, M_{acc} = .66; β = -.22, t = -11.06, $p < .0001$). There was no correlation between self-estimated accuracy and actual accuracy (r = .10, *CI* = [-.06, .25]). Nor was accuracy predicted by participants' age [19, 62] (r = -.08, *CI* = [-0.23, 0.08]) or education level (r = -.07, *CI* = [-0.22, 0.09]).

Discussion

Experiment 1 provides support for implicit, emergent generalizations at the level of grammatical constructions. Evidence that memories for constructions are structured around a centroid based on distributional evidence comes from the increase in false memories for sentences containing a new head word, prototypical for the construction in comparison to sentences containing other head words, matched for overall frequency. Participants were also more likely to falsely believe they had witnessed new instances of the witnessed constructions than close paraphrases that included the same open-class words.

Despite the illusory memories, overall recognition memory was well above chance and moderately high, according to signal detection theory. Accuracy did not correlate with age or education level, and participants believed their accuracy was significantly worse than it was.

Exp. 1 did not test verbatim memory for language, given that lures differed in meaning from exposure items. The only indication that people remembered language verbatim in Exp. 1 comes from the greater likelihood of false alarms to new sentences in the witnessed construction compared to close paraphrases expressed by an alternative construction. Yet this result is attributable to the formation of a generalization at the level of construction. We therefore test for verbatim memory in experiments 2 and 3.

But first we briefly address a potential confound in experiment 1. Could the illusory memory for prototypical instances of constructions in experiment 1 be attributed simply to lexical priming of a "gist" word rather than by exposure to the grammatical constructions? We address this in the following control study.

Control Study: Word Lists

Method

A total of 159 new participants from Prolific completed the task (79 F; 75 M; 4 non-binary; 1 preferred not to say) and were paid for their time.

The control study followed nearly the identical design as Exp. 1, except participants were first exposed to one of 3 lists of words, corresponding to the key words used in Exp. 1 (Table 1: far right column) instead of full sentences. It was therefore not possible to compare new instances of the construction with instances of a paraphrase, so the critical lures only included the prototypical word in comparison to the frequency-matched word. We additionally included 8 unwitnessed lure words at test. After the same 5–10-minute delay involving the same unrelated language task, participants performed a recognition task. Specific lists again varied by construction and by participant. If the greater number of false memories for prototypical instances of constructions compared to non-prototypical in Exp. 1 was due to lexical 'gist' memory generated by the key words rather than by exposure to the constructions, then we should find a comparably high level of false memory for the prototypical key words.

Results

Results of the word list control task do not suggest that lists of the key words used during exposure evoke higher levels of false memory for the "prototypical words" compared to frequency matched controls. To compare false alarms to prototypical lures across experiments, we used a linear model with experiment, condition (matched vs. prototype), and their interaction as predictors. False alarms were significantly lower overall in this control experiment (words) than Exp. 1 (constructions) (β = -0.17, t = -3.50, p < .001). Crucially, the better accuracy in the word control was driven by significantly more false alarms to the cx-prototypical items in Exp. 1 than here (M_{exp1} = .53 vs. M_{control} = .20) as confirmed by an experiment × condition interaction (β = -0.17, t = 2.48, p = .01). False alarms to prototypical words in the control were not significantly different than false memory to frequency-matched controls ($M_{word\text{-}freq\text{-}matched}$ = .17).

Discussion

The control study demonstrates that the false memory for instances containing construction-prototypical head words was not simply due to witnessing head words during exposure, since no evidence for elevated false memories was found for the same words after exposure to a list of head words. The comparison with word lists highlights another difference from word-level DRM tasks. While people have reliably encountered the 'gist' words that are falsely remembered (e.g., *sleep* after witnessing *nap, pillow, night*, etc.), the sentences used as probes in the current experiment may well have been witnessed for the very first time during the recognition task. It is possible that false memories for new whole sentences emerges because the specifics of sentences are never retained in the first place.

That is, the findings of Exp. 1 may be construed as supporting the idea that people only retain fuzzy memory traces of witnessed language (e.g., Brainerd & Reyna, 2002; Roediger & McDermott, 2000; Schacter & Dodson, 2001). As mentioned, above chance accuracy in Exp. 1 could have been due to memory for meaning, as the lure sentences all differed from the exposure sentences in content, together with a fuzzy memory of the constructions.

Exp. 2 addresses the issue of verbatim memory directly and replicates Exp. 1, while diluting the initial exposure to the target construction. In Exp. 2, an additional set of "isolate" sentences are randomly interspersed during exposure. These sentences, which differ from target constructions and from one another in form and meaning are used to test for verbatim memory.

Exp. 2: Replication & Verbatim Memory

Learning a language requires a combination of item-specific knowledge and generalizations. There is broad agreement that humans recognize high-frequency phrases such as idioms and collocations (Arnon & Snider, 2010; Dabrowska, 2014; Jacobs et al., 2017, Tremblay & Baayen, 2010).

However, traditionally, verbatim memory for sentences after a single exposure has been thought rare or nonexistent. Yet we are faced with a conundrum if we wish to posit verbatim memory for high frequency sentences without also positing some verbatim memory for some amount of time after an initial encounter. If no memory trace existed, then hearing a phrase 1000x would be the same as hearing it the very first time. In fact, the early work that is sometimes cited as providing evidence against any verbatim memory had actually observed simply that memory for meaning is stronger than memory for specific forms (Anderson, 1974; Divjak, 2019; Gernsbacher, 1985; Lombardi & Potter, 1992; Potter & Lombardi, 1990).

In fact, verbatim memory for language is known to exist under certain specific conditions, such as when individuals are warned their memory will be tested (Johnson-Laird & Stevenson, 1970), when utterances are emotionally charged (Murphy & Shapiro, 1994) or repeated aloud (Zormpa et al. 2019; Reder, 2014). To probe verbatim memory in typical, naturalistic settings of language-use, Gurevich, Johnson, and Goldberg (2010) reported a series of experiments that found significant verbatim memory for sentences encountered only once, without advanced notice of a memory test in interconnected texts (see also Gernsbacher, 1985; Kintsch & Bates, 1977). However, these studies had all included as participants relatively small samples of undergraduate students at R1 institutions. This population may display verbatim memory due to skills required to be academically adept or due to their relative youth. This is addressed here by testing effects of age or education.

Method

148 participants from Prolific completed the task (74 F; 73 M; 1 non-binary) and were paid for their time.

In addition to the 12 instances of three constructions used as exposure stimuli in Exp. 1, four sets of 8 unique "isolate" sentences were created that were unique from one another in form and meaning (Table 3). Each isolate was paired with a close paraphrase to be used as a lure, isolate sentences and lures were counterbalanced across participants. Word2vec confirmed the cosine semantic similarity between isolate sentences and their lures was high ($M = .92$).

Table 3: Example isolate sentences and corresponding lures

Example isolate	Example close paraphrase
Kids find ice skating really fun.	Ice skating is really fun for kids.
There's a car parked in the street.	A car is parked in the street.

Results

Verbatim Memory Comparing responses to isolate sentences and nearly synonymous lures (Fig. 3 white and gray bars on right), signal detection theory confirms d' is greater than 0, ($M\ d' = .21$, CI [.15, .27]), as confirmed by a nonparametric Wilcoxon signed-rank test ($V = 6050$, $p < .0001$).

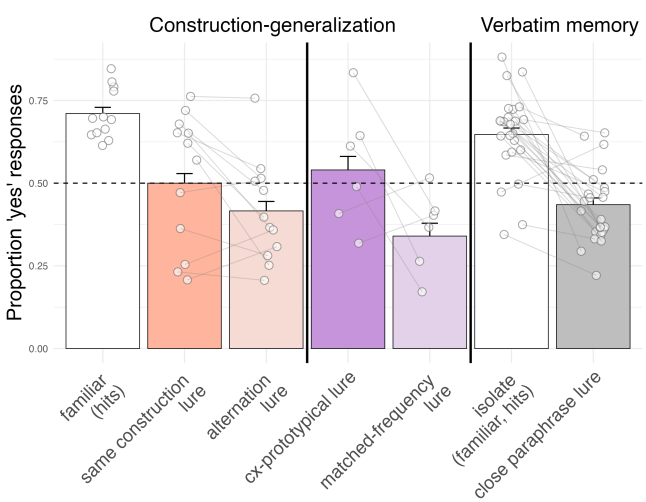

Figure 3: Exp. 2: Proportion of 'yes' (familiar) responses to witnessed items; 4 lure types: alternation vs. same-cx lures (orange), prototypical vs. frequency matched lures (purple), and isolate sentences vs. close paraphrases (white and grey). Points are items; segments indicate comparisons b/w items.

Same-Construction Lures Figure 3 (orange bars) displays numerically higher false alarm rates–lower accuracy—for same-construction lures compared to alternation lures, although the difference is not significant ($\beta = -.53$, $z = -1.36$, $p = .17$). Because the size and direction of the effect are comparable to Exp. 1, we probe this comparison again while setting a higher bar for evidence, to guard against spurious confirmatory results in Exp. 3.

False Memory for Cx-Prototypical Lures Comparing the lures that included the prototypical head word with a frequency matched head word (purple bars), Exp. 2 replicates Exp. 1 finding more false memories for the prototypical lures than frequency-matched lures ($\beta = -.84$, $z = -2.51$, $p = .01$).

Accuracy Participants marginally underestimated their own accuracy ($M_{estimated\_acc} = .57$, $M_{acc} = .61$; $\beta = -.04$, $t = -1.72$, $p = .09$), with no correlation between self-estimated accuracy and actual accuracy ($r = -.006$, CI =[-.17, .16]). Nor was accuracy predicted by participants' age ($r = -.04$, CI = [-.20, .11]) or education level ($r = .006$, CI =[-.15, .17]).

Exp. 3: Setting a Higher Bar for Same-Construction vs. Alternation Lures

Exp. 3 is primarily designed to test whether the significant difference in Exp. 1 between same-construction lures and alternation-lures is valid. Since the difference was not significant in Exp. 2, we intentionally redesigned materials to make construction-level generalizations *less likely*. We did this by further diluting exposure to each construction and increasing potential interference: instead of witnessing 12 instances of a single construction, Exp. 3 exposes participants to only six instances of each of 3 constructions (see Table 2). And instead of witnessing 8 isolate sentences as in Exp. 2, we randomly interspersed 12 isolate sentences, which allows us to replicate the verbatim memory task from Exp. 2. Additionally, we no longer use the same open-class words witnessed during exposure in the lure sentences, allowing us to attribute false-alarms fully to memory for the construction. If we detect significantly more false memories to same-construction lures than alternation lures in this final experiment, we will conclude generalized memory for the witnessed construction exists. Note, since open-class words are new, we expect the proportion of false-alarms to be lower across the board.

Method

A new group of 105 participants from Prolific completed the task (71 F; 26 M; 8 non-binary) and were paid for their time.

Six constructions were chosen in pairs: <double-object, prepositional dative>, <gossip construction, predication>, and <XerYer, conditional>. For each construction, two sets of 6 sentences were created for exposure and testing purposes (= 72 stimuli), so that two sentences across constructions in each pair were highly similar in content. Cosine semantic similarity between paired items was confirmed to be high using Word2vec ($M = .98$). Which construction in each pair was used as the target construction vs. for alternation lures was counterbalanced across participants, as were the lists used for exposure vs. test. No one witnessed the same content more than once at test.

Results

Verbatim Memory Evidence for significant verbatim memory is replicated (Figure 4). Comparing responses to isolate sentences (white bar) and nearly synonymous lures (grey bar), signal detection theory confirms significant sensitivity: *Mean d'* = .17, *CI* [.12, .23]), confirmed by the nonparametric Wilcoxon signed-rank test ($V = 3283, p < .0001$).

False Memory for Same-Construction Lures Importantly, we replicate evidence for construction-level generalization, despite participants being exposed to half as many instances of each of three constructions rather than one. This was confirmed by the same generalized linear model used to compare same-cx and alternation lures in experiments 1 and 2: response is the outcome, same-cx lure vs. alternation lures is the fixed effect, and random intercepts included for participants, items, and constructions ($\beta = -.33$, $z = -2.79$, $p = .005$). The size of the effect is smaller than in Exp. 1 or 2, making it less clearly visible in Figure 4 (orange bars). This isn't unexpected, given the more diluted exposure to each of three constructions and the lower false memories overall, presumably due to the inclusion of content words unwitnessed during exposure in test items. Since Exp. 3 included more items, we were able to confirm the effect in the same model with lure type added as a random slope on participants (additional complexity led to convergence failure) ($\beta = -.44$, $z = -3.00$, $p = .003$).

Accuracy Participants again underestimated their accuracy ($M_{estimated} = .44$ *vs.* $M_{acc} = .73$; $\beta = -.29$, $t = -13.66$, $p < .0001$). Accuracy was again uncorrelated with age ($r = -.005$, CI = [-.20, .19]) and education level ($r = .13$, CI =[-.06, .32]).

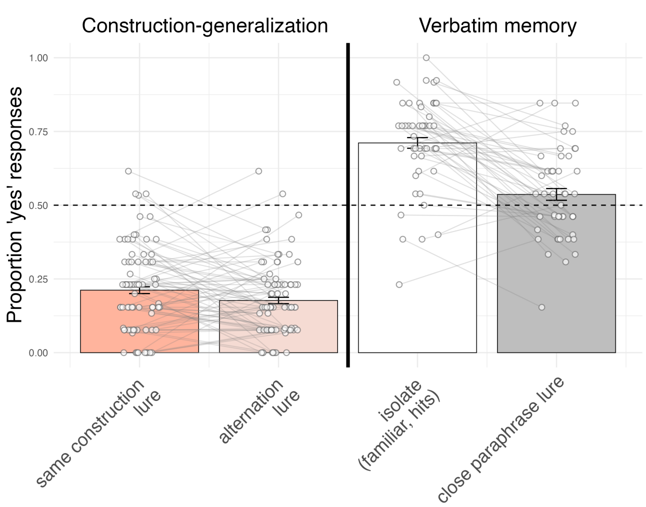

Figure 4: Exp. 3: Proportion of 'yes' (familiar) responses to same-cx lures vs. alternation lures (orange bars); and isolate sentences vs. close paraphrases (white and grey bars). Points are items; segments indicate comparisons b/w items.

General Discussion

The present work investigates whether and how speakers generalize beyond the specific utterances they witness to form structured construction-level generalizations in long-term memory. Results reveal illusory memories for unwitnessed but distributionally prototypical instances (Exp. 1 and 2) indicating that such structured construction generalizations are centered around a prototypical head word. This is potentially a useful cue for learners (Goldberg et al., 2004; Ellis & O'Donnell, 2014) and perhaps partially the source of constructions' prototypical meanings.

We also found that participants were more likely to falsely remember a new instance of a construction witnessed during exposure than a close paraphrase expressed by a different construction, despite a delay and without warning a memory task was to be given. This was significant in Exp. 1, numerically but not significant in Exp. 2, and importantly again when input is even more diluted and complex than in Exp. 2 (Exp. 3). A control study confirmed that false memories are due to construction-level generalization rather than semantic gist from the head word.

We additionally tested whether the existence of generalizations presupposes the loss of item-specific memories. Results show they do not, since people simultaneous retain significant verbatim memory for individual sentences in long term memory (Exps. 2 and 3). Finally, we show that memory for language is implicit, as participants consistently underestimated their performance, which was not predicted by age or education level.

Limitations

Current results situate construction learning within general memory and learning paradigms that allow for generalizations to emerge from episodic memories (Nosofsky, 1986; Hintzman, 1986). We presume similar episodic memories partially overlap in representational space, however, so that individual episodes need not be wholly unique. Future work is needed to model the combination of current results. It will also be illuminating to determine how Large Language Models capture the specifics and generalizations related to our network of constructions. Future work should replicate the current designs using auditory materials to determine whether the use of written stimuli influenced the results. Although we found no evidence of age effects, we cannot rule them out, since our population did not include elderly people nor children. Developmental extensions are of particular interest, since we cannot conclude that construction-level generalizations are initially formed by generalizing across memories of form-function pairings, given that we used familiar constructions. Exposing young children to novel or unfamiliar constructions would be highly valuable.

References

Anderson, J. R. (1974). Verbatim and propositional representation of sentences in immediate and long-term

memory. *Journal of Verbal Learning and Verbal Behavior*, *13*(2), 149-162.
Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, *62*(1), 67-82.
Boas, H. C. (2003). A constructional approach to resultatives. *CSLI publication.*
Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current directions in psychological science*, *11*(5), 164-169.
Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69-94). KNAW.
Bybee, J. L. (2010) *Language, Usage and Cognition*. Cambridge, UK: Cambridge University Press.
Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, *90*(1), 29-49.
Dąbrowska, E. (2014). Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics*, *25*(4), 617-653.
Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology*, *58*, 17–22.
Divjak, D. (2019). *Frequency in language: Memory, attention and learning*. Cambridge University Press.
Ellis, N. C., & O'Donnell, M. B. (2014). Construction learning as category learning: A cognitive analysis. *Constructions Collocations Patterns. Berlin: De Gruyter Mouton*, 71-97.
Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & cognition*, *38*(7), 833-848.
Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive psychology*, *17*(3), 324-363.
Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.
Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning Argument Structure Generalizations. *Cognitive Linguistics.*, *15*(3), 289-316. https://doi.org/10.1515/cogl.2004.011
Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. OUP Oxford.
Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press.
Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition 2*(1), 45–78.
Goldberg, A. E., & Herbst, T. (2021). The nice-of-you construction and its fragments. *Linguistics*, *59*(1), 285-318.
Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological review*, *93*(4), 411.
Israel, M. (1996). The way constructions grow. *Conceptual structure, discourse and language*, *217*, 230.
Jackendoff, R. (1990). *Semantic structures*. Cambridge, MA: MIT Press.
Jacobs, C. L., Dell, G. S., & Bannard, C. (2017). Phrase frequency effects in free recall: Evidence for redintegration. *Journal of memory and language*, *97*, 1-16.
Johnson, M. A., & Goldberg, A. E. (2013). Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and cognitive processes*, *28*(10), 1439-1452.
Johnson-Laird, P., & Stevenson, R. (1970). Memory for syntax. *Nature*, *227*(5256), 412-412.
Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory, 3*(2), 150–159.
Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*(6), 713-733.
Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5-27.
Murphy, G. & A. M. Shapiro. 1994. Forgetting of verbatim information in discourse. *Memory and Cognition*, *22*, 85-94.
Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*(1), 39.
Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*(6), 633-654.
Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of memory and Language*, *38*(3), 265-282.
Reder, L. (2014). The Costs and Benefits of Verbally Rehearsing Memory for Faces. *Basic and Applied Memory Research: Volume 1: Theory in Context; Volume 2: Practical Applications*, 51.
Roediger, H., & McDermott, K. B. (1995). Creating False Memories*:* Remembering Words Not Presented in Lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814.
Roediger III, H. L., & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science*, *9*(4), 123-127.
Schacter, D. L., & Dodson, C. S. (2001). Misattribution, false recognition and the sins of memory. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1385-1393.
Tomasello, M. (2003). *Constructing a language: a usage based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, *151*, 173.

Yi, E., Koenig, J. P., & Roland, D. (2019). Semantic similarity to high-frequency verbs affects syntactic frame selection. *Cognitive Linguistics*, *30*(3), 601-628.

Ziegler, J., Bencini, G., Goldberg, A., & Snedeker, J. (2019). How abstract is syntax? Evidence from structural priming. *Cognition*, *193*, 104045.

Ziegler, J., Snedeker, J., & Wittenberg, E. (2018). Event structures drive semantic structural priming, not thematic roles: Evidence from idioms and light verbs. *Cognitive Science*, *42*(8), 2918-2949.

Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 340-352.