Incidental verbatim memory for language

OLGA GUREVICH, MATTHEW A. JOHNSON AND ADELE E. GOLDBERG*

Princeton University

Abstract

It is widely believed that explicit verbatim memory for language is virtually nonexistent except in certain circumstances, for example if participants are warned they are to receive a memory test, if the language is 'interactive' (emotion-laden), or if the texts are exceedingly short and memory is tested immediately. The present experiments revisit the question of verbatim memory for language and demonstrate that participants do reliably recognize and recall full sentences that they are exposed to only once at above chance rates (Experiments 1 and 3). The texts are 300 words long, non-interactive, and no advanced warning of a memory test is given. Verbatim memory is demonstrated even when lexical content and memory for gist are controlled for (Experiments 2 and 4). The most striking finding is one of incidental recall: even after a six-day delay, participants reliably reproduce sentences they have heard before when asked to describe scenes, even though they are not asked to recall what they had heard (Experiment 5).

Keywords
verbatim memory, recognition, recall, implicit recall, usage-based model

* Correspondence address: Adele Goldberg, Department of Linguistics, Princeton University, 33 Washington Rd, Princeton, NJ 08544-0001, USA. E-mail: adele@princeton.edu.
Acknowledgements: The work described here was done while Gurevich was a postdoctoral fellow at the Educational Testing Service in Princeton, NJ and at Princeton University. She is currently at Microsoft Corp. We are grateful to NSF grant #0613227 to Goldberg for sponsoring this research. We are also very grateful to Charlotte Wieskittel for her senior thesis work on this topic under Goldberg's direction (Weiskittel 2006). Charlotte ran the first recall study, using the Spiderman story, and began the collection of data on Gramps as well. We are also grateful to Devin Casenhiser and Laura Suttle for helping us run subjects, and to Morten Christiansen and an anonymous reviewer for excellent feedback on an earlier version of this paper. Greg Murphy deserves special mention for generously offering astute feedback on two earlier drafts. The authors have only themselves to blame for any remaining infelicities.

Language and Cognition 2–1 (2010), 45–78
DOI 10.1515/LANGCOG.2010.003
1866–9808/10/0002–0045
© Walter de Gruyter

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

1. Introduction

Conventional wisdom holds that, except in limited circumstances, people almost immediately forget the surface details of the language they hear and only explicitly remember the 'gist.' This idea is an old one (Bartlett 1932; Binet and Henri 1894) and has received support from several studies that seemed to demonstrate that verbatim memory is lost as soon as an utterance has been understood (Anderson 1974; Gernsbacher 1985; Johnson-Laird and Stevenson 1970; Sachs 1967).

In a classic study by Sachs (1967), subjects heard a series of passages and were tested at varying intervals for their level of retention of sentences embedded within the passage. Sachs reported that "the original form of the sentence is stored only for the short time necessary for comprehension to occur." She thus concludes, "the meaning of the sentence is derived from the original string of words by an active, interpretive process. The original sentence which is perceived is rapidly forgotten, and the memory then is for the information contained in the sentence" (Sachs 1967: 442).

In another experiment, Bransford and Franks (1971) asked participants to recognize which sentences they had heard before by choosing among a new set of closely related sentences. All of the sentences presented during the experiment were combinations of component parts of a series of complex sentences or ideas. They found that participants "spontaneously integrate the information expressed by a number of non-consecutively experienced (but semantically related) sentences into holistic, semantic ideas" and thus participants falsely 'recognized' novel sentences that contained the combined meaning of multiple individual sentences (1971: 331).

Bock and Brewer (1974: 841) argued on the basis of subjects' tendency to falsely recall related but unwitnessed sentences that "an abstract representation of the meaning was remembered *rather than* the exact words, and . . . in recall the surface structure was reconstructed from this abstract representation" (emphasis added). This idea is echoed in much early work, e.g. "Listeners do not ordinarily retain the syntax of a sentence for longer than is necessary to grasp its meaning" (Johnson-Laird et al. 1974: 704). This idea remains the current wisdom on the topic, e.g. "There is no explicit memory for the surface syntactic structure of a perceived sentence" (Potter and Lombardi 1998: 267); or "Research on memory for verbal materials has demonstrated that sentences are quickly transformed into an underlying abstract meaning and that the original surface structure is lost" (Holtgraves 2008: 361). Loebell and Bock (2003: 793) state, "One of the most robust findings in psycholinguistics is that people cannot reliably recall sentence structures."

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

1.1. *Surface syntax is remembered in certain circumstances*

At the same time, it is recognized that surface syntax *can* be remembered under certain limited circumstances, in particular, when subjects are warned they will have to remember the sentences verbatim, when the utterance is emotion-laden, or when sentences are not presented as part of a coherent text.

Memory for surface form has been demonstrated to exist when subjects are told in advance that they will be tested on this information, particularly when the texts consist of only a few sentences. In a recognition task, Johnson-Laird and Stevenson (1970) systematically varied whether subjects anticipated that a memory test would follow exposure to a short orally presented passage (exactly how short the passage was is not stated); they found that only those who were warned that the memory test would follow demonstrated above-chance performance. They conclude by noting that "The results suggest that subjects tend to retain syntax . . . *only if* they know that they are to receive a memory test" (cf. also Johnson-Laird et al. 1974: 412). The authors go on to suggest that some form of verbal rehearsal may underlie the above-chance scores when the memory test was anticipated. Such rehearsal would clearly be an atypical comprehension strategy. Reyna and Kiernan (1994) also found evidence of verbatim memory, in six- and nine-year-old children in a paradigm in which participants were warned and tested immediately after having heard each of 8 three-sentence texts. The immediacy of the testing, the brevity of the texts involved, and the forewarning involved make these studies ill-suited to addressing the question of whether subjects retain verbatim memory in more naturalistic contexts.

We have all had the experience of having a particular hurtful, flattering, colorful or funny comment 'ring in our ears' weeks or months after the event. In fact, verbatim recognition memory has been demonstrated for such 'highly interactive' or socially important utterances. Murphy and Shapiro (1994) had one group of subjects read a rather bland, non-emotional letter, while another group read a biting, sarcastic letter. They found that subjects were better able to remember a given sentence when it received the sarcastic interpretation than when it received a bland interpretation in the context of the other letter.[1] Keenan et al. (1977) likewise tested the memory of subjects exposed to an academic lecture after a 30-hour delay. In a recognition task, the experimenters found that listeners

1. Whether or not subjects were significantly above chance at verbatim recognition in the 'bland' condition is not stated, although they were said to correctly identify old sentences 71% of the time, while incorrectly labeling paraphrases as old 54% of the time, which is suggestive of some verbatim recognition.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

were better at retaining surface information for utterances with a high 'interactional' content than they were for utterances with a low interactional content (cf. also Murphy 1992). But this type of high emotion language does not account for the majority of the language we hear, which is more emotionally neutral.

Finally, researchers have documented that verbatim memory is improved when subjects are exposed to isolated sentences or unrelated items that do not form a semantically coherent passage (Anderson and Bower 1973; Peterson and McIntyre 1973; de Villiers 1974). Gernsbacher (1985) suggests that integrating sentences into a coherent semantic representation seems to be causally involved in the loss of verbatim memory, insofar as coherent texts facilitate the creation of a summary gist or mental model. But again, this laboratory-induced phenomenon does not reflect our normal experience with language. While input is occasionally perceived as incoherent, we are quite adept at imbuing conversations with coherence, even when the utterances appear to be unrelated (Grice 1975).

Kintsch and Bates (1977) represents the rare study that has demonstrated some verbatim recognition memory in a naturalistic context (cf. also Gibbs 1981). Subjects listened to a regular lecture in which, unbeknownst to them, a subset of stimuli sentences were embedded. In the following class, two or even five days later, subjects were given a recognition memory test on the set of sentences and found to recognize old sentences significantly better than paraphrases. The old sentences and paraphrase sentences had overlapping but distinct open-class lexical items, however, leaving open the possibility that the appearance of verbatim memory of sentences was driven by memory for particular words (this is also true of the related experiment by Gibbs 1981); we return to this point when Experiment 2 is introduced. The studies reported here collectively serve to support and extend these early findings in a number of ways.

1.2. *Organization*

In this paper we revisit the long-standing issue of whether verbatim memory for language is above chance in naturalistic contexts of language use: when speakers are not warned that they will receive a memory test and when sentences are presented as part of coherent texts and are not interactive or emotion-laden. We investigate explicit verbatim memory in both *recognition* (Experiments 1 and 2) and *recall* of texts (Experiments 3 and 4). Experiments 2 and 4, moreover, control for possible lexical memory effects in order to determine whether memory for verbatim language extends beyond memory for individual words. In a final study (#5), we investigate whether hearing a story version once affects the retelling of the

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

story after a six-day delay, when participants are not asked to explicitly recall the story. This last experiment is a rare investigation into *incidental recall:* the tendency to use the same clauses that had been witnessed previously, without explicit direction to do so.

Previous results that had been offered as evidence against verbatim memory are reconsidered in light of the present results. We find that various factors have conspired to downplay the importance of memory for language in past research. When looked at in depth, previous work can be seen to support the view that verbatim memory for language exists.

1.3. *Desiderata*

Our experimental designs satisfy four important criteria. First, they involve realistic, connected text since lists of words or other disconnected text are not likely to be processed and remembered in the same way as natural discourse. We use text and illustrations from children's storybooks as stimuli, which allow us to approximate natural speech, as both natural discourse and children's books primarily use simple sentence patterns, shorter sentences and higher frequency words. One of the storybooks was written as a first person narrative and the other included dialogue, both of which are standard forms in spoken language.

A second criterion was that participants were given no indication that their memory of the text would be tested. Speakers' naiveté was confirmed during debriefing. Forewarning was avoided because it could cause subjects to rely on a special purpose encoding strategy instead of allowing comprehension to occur naturally.

A third criterion was that the texts were long, at around 300 words. This ensured that artificial forms of encoding were unlikely; the texts were too long to support auditory or other forms of conscious rehearsal, for example. The longer texts also ensured that memory would be tested more than a few syllables downstream.

A final criterion was to have a way to determine, in the case of free recall (Experiments 3 and 5), whether subjects were actually remembering texts verbatim or whether they simply happened to formulate the same content in the same way as the author of the story had. Insofar as language is formulaic, this is a real concern. That is, while it is possible to greet someone with any one of a number of possible utterances, including *how are you?, hello, nice to see you, you look familiar*, these phrases could be uttered due to priming by an interlocutor or they could just happen to be produced. To control for this issue, we created two versions of each story; each clause from version 1 had a corresponding clause in version 2 with the same content but different surface form. The research question

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

was therefore whether subjects who heard a particular formulation were more likely to use that formulation than subjects who had heard a paraphrase of the same content. This design allows us to distinguish verbatim memory from memory for gist.

The next two sections describe two experiments using different stories, designed to assess verbatim recognition of a coherent, fairly long and non-'interactive' text, for which no warning that memory would be tested was given.

2. Experiment 1: *Spiderman* recognition

Experiment 1 involves a visually presented recognition task of an orally presented naturalistic story. Subjects are presented with one of two versions of the story and are then tested on clauses that appeared in either of the two versions. Their task is simply to decide if the item had been witnessed before (old) or whether it is new (paraphrase). For each subject, half of the test items came from the version they had heard and the other half from the alternative version.

2.1. *Participants*

Twenty-four Princeton University undergraduates participated in the study, fulfilling a mandatory experimental requirement for a psychology course.

2.2. *Materials*

The test materials were based on an illustrated children's book, *Spider-Man: I am Spider-Man* (2002). The book is 32 pages long, 300 words, and each page contains its own picture and between one and four sentences of text. The text is written in the first person and does not include any dialogue.

While the pictures serve as cues for advancing the story, the accompanying sentences are not descriptions of the content of the pictures (see Appendix for the two versions of the story and the full set of pictures).

In order to test specifically for memory of surface syntactic structure, we created a second version of the story with the same content and the same word count. Sentences across the two versions were matched for the same gist content as those in the original story, but had different syntactic structures. Some examples of the sentences in the two versions are below.

Story version 1: My fingers can stick to anything.
Story version 2: I can stick my fingers to anything.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

Story version 1: I am strong enough to fight four bad guys at once!
Story version 2: Fighting four guys at once is easy for someone as strong as me.

The pictures were scanned and the text was digitally removed. A male narrator recorded the stories. The total playing time was 4:02 minutes for the original version of the story and 4:12 minutes for the alternate version.

For purposes of subsequent analysis, we subdivided the two stories into clauses, with one main predicate per clause. The analysis only considers those clauses that differed by more than one word between the two versions. There were a total of 37 such clauses, although there were 54 total clauses in each story.

As a between-subjects variable, we varied whether pictures from the original text were used as prompts during the recognition task (cued-recognition: group 1a) or whether random pictures appeared before each prompt (non-cued-recognition: group 1b). We hypothesized that pictures from the text would prompt participants to reinstate the context, which might increase recognition accuracy. Therefore, in addition to the pictures from the comic book for group 1a, we used unrelated distracter pictures from the Internet, depicting various scenes of nature, for group 1b.

2.3. *Procedure*

The experiment consisted of two stages: listening, and performing the recognition task. Before the start of the experiment, subjects were not warned about the purpose of the study. They were simply asked to listen to a story while looking at the accompanying pictures on a desktop computer.

In the listening stage, subjects heard one of two versions of the story while looking at the corresponding pictures. The story was advanced automatically, as soon as the accompanying sound was finished playing for each picture.

The recognition testing phase began immediately after the story was presented. Written clauses were displayed on the computer screen one at a time and subjects were asked to answer 'yes' or 'no' to whether or not they heard that exact clause in the story. Before each clause, a corresponding picture was shown for 1000 milliseconds; either the picture that had appeared when that content had been heard (in the case of group 1a subjects) or an unrelated picture (group 1b subjects). Subjects recorded their answers by pressing one of two keys. The selection of a key caused the program to progress to the next clause.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

We created two mixed lists consisting of clauses from the two story versions, each containing half the clauses from one story and half from the other, chosen at random, thus assuring that half the clauses were from the story version they heard (i.e. matching, with a veridical answer of 'yes'), and half were from the alternative version (non-matching, with a veridical answer of 'no'). The clauses were presented in the same order as in the original story. Each subject was given one of the two lists in the recognition stage. In several instances, the same picture was accompanied by more than one recognition clause. In these instances, the corresponding distracter picture was also presented before each clause, such that the presentation was exactly parallel with either set of pictures.

Thus the design was mixed-factorial. The between-subject factors were 2 (story 1 or story 2) $\times$ 2 (original or distracter picture) $\times$ 2 (recognition clause list). Matching or non-matching was a within-subjects factor.

2.4. *Results*

The measured dependent variable was accuracy. In order to smooth the effects of some participants having inherently better memory than others, we aggregated our data by subject and used the by-subject data points in measures of statistical significance.

The mean percentage of correct responses with the original picture was 0.72, and the mean with the distracter picture was 0.73: a non-significant difference (one-way ANOVA, $F(1, 22) = 0.22,\ p > 0.5$). Thus results in which pictures from the story were displayed and those with distracter pictures are collapsed below.

We can compare the probability of a 'yes' response to a matching sentence (true positive) and a non-matching sentence (false positive) (Table 1):

Table 1. *Spiderman: Mean probabilities of 'yes' and 'no' responses*

	Matching	Non-matching
Probability of 'yes'	**0.86 (Hits)**	0.41 (False alarms)
Probability of 'no'	0.14 (Misses)	**0.59 (Correct rejections)**

Note. Correct responses are indicated by boldface.

Using signal detection theory, we calculated the sensitivity rate, *d-prime*, for distinguishing matching and non-matching sentences. A *d'* of zero would indicate that there is no difference between signal and noise in the subjects' responses. In fact, the mean *d'* across individual subjects

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

was 1.42 ($N = 24$). A one-sampled t-test demonstrated that this *d'* is significantly greater than zero ($t(23) = 14.08$, $p < 0.01$).

2.5. *Discussion*

The results of Experiment 1 reveal that verbatim memory in recognition was significantly above chance, even though the experimental context was not highly interactive, the text was a cohesive one, and subjects were not warned that they were doing a memory task. In debriefings, subjects confirmed that they did not know they would be asked to retell the story, and they had not attempted to memorize the story when they heard it.

One concern about Experiment 1 was that the effect could have been driven, at least in part, by lexical effects (Reyna and Kiernan 1994). If a speaker who heard Story version 2 remembered that the word *easy* was used to describe a particular content, then he would be more likely to correctly recognize that the second and not the first utterance had occurred in the text:

Story version 1: I am strong enough to fight four bad guys at once!
Story version 2: Fighting four guys at once is easy for someone as strong as me.

The Kintsch and Bates (1977) experiment that had shown evidence for verbatim memory also involved differing open-class words in their old and paraphrase sentences. It is worth investigating whether verbatim memory relies crucially on memory for open-class words. Therefore, in a second experiment, we created two versions of a story that differed only in closed class words (prepositions, determiners, copula verb, light verbs, auxiliaries and morphology); all open-class content roots in each story were identical. It is generally accepted that closed-class words are not significantly affected by additional exposure, most likely because they are so frequent that their base activation rate does not change measurably. The fact that the two story versions differed only in closed class items should essentially eliminate lexical effects on the memory task. If we still find evidence of verbatim recognition memory, it is stronger evidence that memory traces for full utterances are retained.

Another potential issue with Experiment 1 stems from the fact that the Spiderman story was written in the first-person. This could be argued to increase the 'interactivity' of the context, and thus the tendency to remember the language verbatim. Also, although our subjects had not seen this particular comic book before, the general outline of the Spiderman story is well-known in popular culture, conceivably making it easier to remember and recognize verbatim sentences; alternatively, one could argue

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

that because the Spiderman script was known it would be easier to create an overall gist without relying on verbatim recall.

The story used in Experiment 2 was written in the 3rd person and the story line was unfamiliar to our subjects. The two versions used, one heard by each of two groups of subjects, differed only in closed class items. The use of a second story also guards against item-effects.

3. Experiment 2: *Gramps and the Fire Dragon* recognition

3.1. *Participants*

Twenty-four Princeton University undergraduates participated in the study, fulfilling a mandatory experimental requirement for a psychology course. None of the subjects in Experiment 2 had participated in Experiment 1.

3.2. *Materials*

We chose another children's story of comparable length to *Spiderman*, with illustrations that would act as cues but that were not depictions of the text. The story was *Gramps and the Fire Dragon* (2000). Some of the two-page illustrations were reduced to just one page for size consistency, so there were a total of 25 pages in the experiment. The story is told in third person but includes dialogue. As in Experiment 1, the words were digitally removed from the scanned pictures.

The original story consisted of 39 sentences and 272 words, which we initially divided into 40 clauses for coding purposes. As before, we also created an alternative version of the story, controlling for number of words and content, as in the example below:

Story version 1: But that only made the dragon angry.
Story version 2: But that only angered the dragon.

In other cases, we altered a sentence into a fragment (or vice versa), or changed conjoined verb phrases into two sentences. (See Appendix for the two versions of the story.)

The second version of the *Gramps* story had slightly more words than the original (N = 299 vs. N = 272). For purposes of analysis, we selected only those clauses that differed by more than one word in the two versions of the story, resulting in 37 clauses. However, the full story was recorded in each case so as not to interrupt the narrative flow.

Both versions of the story were recorded by a male speaker. The sound clips were presented with each picture using E-prime. Total running time for Story 1 was 3:02 minutes; for Story 2, it was 2:54 minutes.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

We again presented one group of subjects with distracter pictures instead of the storybook pictures. The distracter pictures were the same as those used in Experiment 1.

3.3. *Procedure*

The method of presentation and testing was exactly the same as in Experiment 1.

3.4. *Results*

The measured variable was accuracy.

The mean correct response when subjects viewed the original pictures was 0.60; the mean correct response when subjects viewed the distracter pictures was 0.59. As in Experiment 1, these are not significantly different ($F(1, 886) = 0.74, p > 0.5$), so the two groups are collapsed in the following discussion.

We compared the probability of a 'yes' response to a matching sentence (true positive) and a non-matching sentence (false positive) (Table 2).

Table 2. *Gramps and the Fire Dragon: Mean probabilities of 'yes' and 'no' responses*

	Matching	Non-matching
Probability of 'yes'	**0.81 (Hits)**	0.63 (False alarms)
Probability of 'no'	0.19 (Misses)	**0.37 (Correct rejections)**

Note. Correct responses are indicated by boldface.

The mean sensitivity measure *d'* was 0.70 across all subjects (N = 24). A one-sample t-test demonstrated that it was significantly greater than zero ($t(23) = 5.25, p < 0.01$). These results show that, as in Experiment 1, subjects were consistently able to distinguish between sentences they heard verbatim and sentences that had the same content but different surface form.

In debriefing, subjects stated that they did not realize their memory would be tested, and they did not try to memorize the story as they were listening to it.

3.5. *Discussion*

In both Experiments 1 and 2, recognition accuracy was above chance. Experiment 2 replicated Experiment 1 in demonstrating that there is verbatim recognition of texts that are fairly long (roughly 300 words) and involve non-interactive, coherent texts, even when subjects do not expect a

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

memory test. In addition, Experiment 2 demonstrates that the effect is not due simply to memory for individual open class items, as the same open-class roots were used in both story versions. Moreover, it is not necessary that subjects already have a general 'script' for the storyline (as they may have in the case of the Spiderman story used in Experiment 1), nor is a 1st person narrative essential.

4. Verbatim memory in previous literature revisited

How are studies #1 and #2 to be reconciled with previous findings that argued for a lack of verbatim recognition memory? It turns out that in several of the early studies, there are in fact indications that verbatim structures are not completely forgotten, even though the papers tended to emphasize (and be remembered as having demonstrated) a lack of verbatim recognition. It seems that on a time-scale of decades, there is some truth to the idea that only the gist (of the articles) is remembered. The present studies strengthen earlier findings in that our subjects received no forewarning; study #2 also controlled for possible lexical effects.

In the original study by Sachs (1967), although the findings were discussed by Sachs herself as providing evidence that verbatim memory was lost, there was actually some evidence for verbatim memory traces. At delays of 80 and 160 syllables, the difference between true and false positives (i.e. verbatim memory) was significant (reported as $p < .01$, and $p < .05$ respectively). At the same time, this previous evidence must be qualified in that subjects were instructed that they would be asked to recognize original sentences, and they were provided with four practice passages as a warm-up. As noted by the Johnson-Laird and Stevenson (1970) study, such instructions enhance performance. Also, subjects were not tested at delays longer than 160 syllables, thus providing no direct comparison with our 300-word stories.

Jarvella (1973) tested recall and recognition of sentences that came close to the end of 100-200 word segments of dialogue. The test sentences were followed by a sentence that either repeated the test sentence, implied it (as with a tag question), or was novel. Immediately after this last sentence, subjects were asked to recognize or recall the test sentence. Recognition was correct 96% of the time for repeated sentences, 82% for implied sentences, and 76% for novel sentences, with significant differences between the three. However, as in Sachs' study, subjects were warned that they were participating in a memory task, and the time elapsed between exposure and test was quite short (the length of one sentence).

Zimny (1987), as described in Kintsch et al. (1990), constructed two alternative versions of several 150-200 word stories. Subjects were not ex-

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

plicitly warned that the task was memory-related, but should have been able to guess it since each subject was tested on multiple stories. After hearing a story, subjects were asked to recognize one sentence, which was presented verbatim, with only surface changes (paraphrase), with meaning implied from the original, or with different meaning. Surface memory was measured as the difference in recognition between verbatim sentences and paraphrases. On immediate testing, verbatim recognition (hits) was at 74%, and paraphrase recognition (correct rejections) was at 50%, giving an average correct rate of 62% (similar to our findings for *Gramps and the Fire Dragon*). Again, however, subjects knew memory of the texts would be expected. Zimny found that when subjects were tested after a 40-minute delay, the difference between true and false positives became insignificant, leading her to conclude that verbatim memory was lost. However, the foils included not only different surface form, but in many cases also different meanings; knowing from previous tests what kinds of foils to expect may have predisposed subjects towards meaning-based memory.

Johnson-Laird and Stevenson (1970) found no evidence of verbatim recognition memory unless subjects were warned they were to receive a memory test. But, as was more common at the time, the study involved few subjects (12) and only a single stimulus sentence.

More recently, Holtgraves (2008), quoted in the introduction as presupposing a lack of verbatim memory, actually provides some relevant data, although it is not analyzed explicitly for verbatim recognition or recall. Instead, the studies demonstrated that people are quite sensitive to whether sentences perform speech acts; for example, there is an increased likelihood of falsely remembering a speech act verb after being exposed to a sentence that performs an (indirect) speech act than after a control sentence. In Holtgraves' Experiment 2, participants were asked to read 24 short scenarios, each one 3–7 sentences in length; without advanced warning, their memory for certain sentences was tested after a five minute delay. Although comparisons to chance are not provided, when participants were given a choice of four sentences including a verbatim sentence, a semantically matched foil, and two semantically distinct foils, and asked to choose an exact match, they correctly chose the verbatim sentence 78% of the time (Holtgraves' Experiment 2). As discussed more below, another group of participants also recalled 19% of the sentences heard verbatim (Holtgraves' Experiment 4), when cued only with a single word (naming the speech act involved); a discussion of recall memory follows in Section 5.

To summarize, memory for content or 'gist' has been emphasized in these studies, and in their later interpretations by others. Clearly, when compared directly, memory for meaning always trumps memory for structure, but there seems to be no solid evidence that verbatim memory

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

disappears entirely, and in fact previous studies offer numerical indications that some verbatim memory does remain. Our Experiments 1 and 2 demonstrate clearly that verbatim recognition is significantly above chance, even when subjects are not warned in advance that they will receive a memory test and when lexical effects are controlled for.

5. Recall experiments

Much less research has been done on the type of retention that occurs in recall. Since recall is generally thought to be more difficult than recognition, and since recognition of verbatim language has been believed to be poor or nonexistent, it has likely been taken for granted that recall memory for verbatim language does not exist, except possibly in the sorts of limited contexts outlined in Section 1.1.

And yet it is important to consider the ability to recall verbatim language, because the absolute number of correct responses in recognition tasks is not terribly meaningful. The number of correct responses clearly can be made to vary depending on how close lure sentences are to target utterances. Lures that differ only in tiny inconsequential ways (such as using contractions or not) are likely to decrease the number of correct responses, while lures that are insufficiently similar to the target sentence could artificially inflate the number of correct responses.

One of the first researchers to test recall in natural discourse was Bartlett (1932), who had subjects read a story about an unfamiliar culture and then retell it. He did not quantify his results, but he found that subjects tended to change or forget details that were inconsistent with their cultural frame of reference. This research led to the formation of Schema Theory, which was intended to explain and account for the incompleteness of memory. The theory claimed that incoming information is reduced to an integrated set of only the most relevant and important information, and that all irrelevant information and details, including the specifics of structure, are lost.

Some verbatim recall has been documented in earlier studies where very specific cues were used, and where subjects were warned about the memory task ahead of time. Jarvella (1973) was able to prompt recall by accompanying test sentences by audio tones and asking for recall after one intervening sentence was heard. He found verbatim recall rates of between 26% and 51%, depending on the nature of the intervening sentence. However, there was no control for whether the subjects remembered the surface form of the sentence, or remembered the gist and just happened to express it using the same surface form as the original (see next section for a more detailed discussion of this issue).

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

In a more controlled study, Bock and Brewer (1974) constructed different surface forms for lists of 10 sentences. After hearing a list, subjects were asked to write down what they remembered given one-word cues for each sentence, which were neutral with respect to surface form. Subjects successfully recalled 41% of the sentences verbatim compared to 16% close paraphrases (the rest of the productions were errors or omissions). They report the difference as significant ($p < 0.01$), but the short, artificially constructed sentence lists and advance knowledge of the memory task make it difficult to infer the existence of general verbatim recall memory from this study.

As mentioned above, recent work by Holtgraves (2008) demonstrated that 19% of the sentences heard were recalled verbatim (his Experiment 4). Participants were exposed to 24 scenarios, and then asked to recall a subset of those sentences when cued with the speech act that the sentence had performed as well as the speaker's name. For example, for a target sentence such as *I like your new coat*, participants were cued with 'Heather; compliment.' For the target *What time is it?* Participants were cued with 'Jane; ask.' However, because only one version of each sentence was used and many of the target utterances were fairly formulaic, it is conceivable that subjects primarily remembered the gist and just coincidentally formulated the meaning in a way that matched the target sentences.

The studies described in the next two sections investigate verbatim recall by prompting participants to recall specific sentences of fairly long stories, without prior warning about the memory task. Two versions of each story were used, with each sentence of version #1 paired with a paraphrase in version #2; in this way we were able to compare matches for heard utterances with coincidental matches to the alternative version.

The pictures from the recognition experiments were used as cues. In addition to immediate recognition of Experiments 3 and 4 below, subjects were also tested after six days, as described in Experiment 5.

6. Experiment 3: Spiderman recall[2]

6.1. *Participants*

Twenty Princeton undergraduates participated in the study, fulfilling an experimental requirement for a psychology course. All were native English speakers.

2. This experiment is based on Charlotte Weiskittel's senior thesis (Weiskittel 2006) done under the supervision of Goldberg.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

6.2. *Materials*

We used the same Spiderman comic book as in the recognition study. As before, two versions of the story were created in order to compare the number of verbatim matches with the number of accidental matches to the alternative version of the story that the subject did not hear.

6.3. *Method*

Participants were tested one at a time. They were asked to listen to a story while looking at the pictures on the screen, without instructions as to what type of task would follow. Each participant heard one of the two story versions. Immediately after the entire story was finished, an instructional page was shown that asked the participant to 'read the story back' as accurately as they could while looking through the original pictures. The precise instructions are in the Appendix.

The retellings were recorded and transcribed. We analyzed each clause produced by a participant as a Match (which could be an exact match or a one word off match), or a Non-Match. We feel our coding scheme was conservative, even though we included utterances in which a single word was changed in the final tally of verbatim Matches because the differences involved were very minor. For example, we included as Matches changes from *I am* to *I'm*, or from *I can go places no one else can* to *I can go places that no one else can*. We did *not* count as verbatim matches those productions that differed by two words, even if they were otherwise reasonably close in form to the target sentence. For example, the following productions were not counted as matches:

- **Heard:**
  - The kids at school were amazed **at** my **new** strength.
- **Produced (counted as non-match):**
  - The kids at school were amazed **by** my new-**found** strength.
- Heard in alternative story:
  - My new strength amazed my classmates.
- **Heard:**
  - I **was** just like everyone else.
- **Produced (counted as non-match):**
  - I **used to** be like everyone else.
- Heard in alternative story:
  - I wasn't always a superhero.

In addition, we classified each clause uttered as to whether or not it constituted a Match (exact or one word off) to the alternative version of

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

the story that had not been heard by the participant. This was done so that we could compare Matches to the heard story to accidental Matches to the alternative version of the story that had the same content.

6.4. *Results*

There were a total of 48 clauses in either version, of which we considered 43 (because the remaining five only differed by one word). Participants who heard story 1 produced on average 33.8 clauses, or 70% of those heard; those who heard story 2 produced 31.6 clauses, or 65.8%. The results are summarized in Table 3 and Figure 1 below.

Table 3. *Spiderman: Cumulative numbers of Correct Matches (to heard story) and Coincidental Matches (to alternative version of the story)*

	Correct Matches	Coincidental Matches	Total # of clauses heard
Story version 1	96	49	430
Story version 2	48	5	430
Total	144	54	860

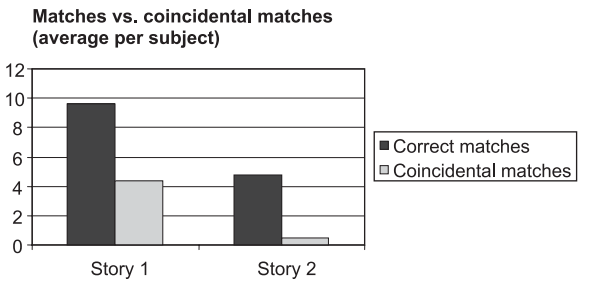


Figure 1. Spiderman: matches vs. coincidental matches

Results demonstrate clear evidence of verbatim recall. In particular, we compared mean verbatim matches across speakers, looking for a difference between matches to the heard story vs. coincidental matches. An overall 2×2 ANOVA (match/coincidental-match $\times$ story version 1/story version 2) revealed a significant main effect of match/coincidental-match ($F(1, 16) = 12.14, p < 0.01$), demonstrating evidence of verbatim memory. There was also a significant effect of story version ($F(1, 16) = 10.18, p < 0.01$), reflecting the fact that the number of coincidental

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

matches to story 1 was significantly greater than for story 2 ($F(1, 18) = 54.98, p < 0.01$). There was no significant difference in correct matches between version 1 and version 2 ($F(1, 18) = 3.22, p > 0.05$). There was also no significant interaction between match type and story version ($F(1, 16) = 0.11, p > 0.05$), indicating that the amount of verbatim memory for the two versions was not statistically different.

We also considered whether verbatim memory was evident for each story version, taken independently, setting a higher bar for the demonstration of verbatim recall. In fact, verbatim memory was evident for both story versions. Comparing the average number of correct matches to story 1 by the subjects who heard it (N = 10) with the average number of coincidental matches to story 1 by the subjects who heard the alternative story (N = 10), there was significant evidence of verbatim recall for version 1 ($F(1, 18) = 7.40, p < 0.01$). Similarly, the number of correct matches to story 2 by the subjects who heard it (N = 10) was significantly greater that the average number of coincidental matches to story 2 by the subjects who did not hear it (N = 10) ($F(1, 18) = 8.40, p < 0.01$).

6.5. *Discussion*

Subjects showed a remarkable ability to recall verbatim language heard in a fairly long story (300 words) without receiving any warning or indication that their memories for the stories would be tested. In line with earlier studies of verbatim memory, the majority of what participants recall is *not* verbatim matches to what they had heard: they by no means remember the entire stories verbatim. But the fact remains that a significant amount of verbatim memory is retained.

The fact that there were more coincidental matches to story version 1 than to the second story version demonstrates that our use of two versions of the same story is an important aspect of the design. If an analysis were to only consider the number of verbatim matches without comparing verbatim matches to coincidental matches, it would be subject to the criticism that the matches may be simply the result of remembering gist and formulating it in a natural way. It is probably fair to say that the language of the published story is more natural than our amateur attempt to write a children's story; the greater number of coincidental matches to the published version is humbling evidence of this fact. Yet importantly, both the subjects who heard story version 1 and those who heard story version 2 were more likely to produce verbatim matches to the story they heard than could be expected by the rate of coincidental matches to the same story version. This is strong evidence that subjects retain significant verbatim memory for the purposes of recall.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

7. Experiment 4: Gramps and the Fire Dragon recall

As discussed in preface to Experiment 2, it is useful to test memory for language on more than a single story. Moreover, the two versions of the Spiderman story involved some differing open class words and so it's conceivable that recall was encouraged by memory for particular open class words. For these reasons, we attempted to replicate the recall findings of Experiment 4 with the *Gramps and the Fire Dragon* story (used in Experiment 2). Recall that in the *Gramps* story, all open-class roots were shared across the two versions of the story.

7.1. *Participants*

Twenty Princeton undergraduate students participated for course credit. None of the subjects participated in any of the other verbatim studies.

7.2. *Materials*

One of the same two recorded versions of the children's book *Gramps and the Fire Dragon* used in the recognition Experiment 2 was played to each participant.

7.3. *Method*

The setup was the same as with the *Spiderman* recall experiment. Students listened to a version of the story while looking at the pictures. At the end, they were asked to repeat the story as best as they could while looking at the original pictures in the same order. Their responses were recorded, transcribed, and coded for whether they were Matches or Non-Matches, as described above.

7.4. *Results*

We again found evidence of verbatim recall. There were a total of 39 clauses, of which we considered 37 (because the other two differed only by a single word). Participants produced on average 31 clauses, or 82% of the number heard. Overall, of the clauses produced, 14% ($N = 85$) were verbatim matches to a heard clause. Since subjects produced fewer clauses than they had heard, this constitutes 11% of heard clauses being reproduced verbatim; by comparison, subjects produced only 4% coincidental matches to the alternative story they had not heard. The results are summarized in Table 4 and Figure 2 below.

We again compared mean verbatim matches across speakers, looking for a difference between matches to the heard story vs. coincidental

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

Table 4. *Gramps and the Fire Dragon*: *Cumulative numbers of Correct Matches (to heard story) and Coincidental Matches (to alternative version of the story)*

	Correct Matches	Coincidental Matches	Total # of clauses heard*
Story Version 1	52	20	370
Story Version 2	33	7	370
Total	85	27	740

* Total number of clauses = (37 × 10 people)

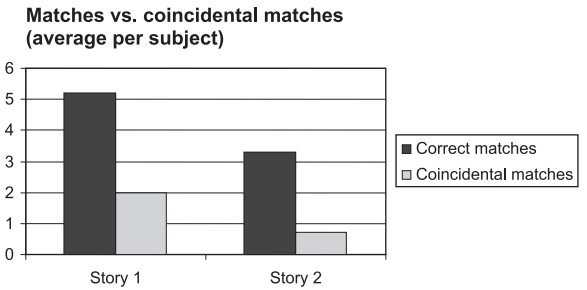


Figure 2. Gramps: matches vs. coincidental matches

matches. An overall 2×2 ANOVA (match/coincidental-match × story version 1/story version 2) revealed a significant main effect of match/coincidental-match ($F(1, 16) = 20.48, p < 0.01$), replicating the finding of verbatim memory in Experiment 1. There was also a small but significant effect of story version ($F(1, 16) = 6.24, p < 0.02$), again reflecting the fact that the number of coincidental matches to story 1 was marginally greater than for story 2 ($F(1, 18) = 4.21, p = 0.06$). The difference in correct matches between version 1 and version 2 was not significant ($F(1, 18) = 2.91, p > 0.05$). There was also no significant interaction between match and story version ($F(1, 16) = 0.22, p > 0.05$), indicating that the amount of verbatim memory for the two story versions was not significantly different.

We also considered whether verbatim memory was evident for each story version, taken independently, and again found that the effect was evident for each story version. That is, comparing the average number of correct matches to story 1 by the subjects who heard it (N = 10) with the average number of coincidental matches to story 1 by the subjects who heard the alternative story (N = 10), there was significant evidence of verbatim recall for version 1 ($F(1, 18) = 9.25, p < 0.01$). Similarly, the number of correct matches to story 2 by the subjects who heard it

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

($N = 10$) was significantly greater that the average number of coincidental matches to story 2 by subjects who did not hear it ($N = 10$) ($F(1, 18) = 12.62, p < 0.01$).

7.5. *Discussion*

Experiment 4 replicated the main finding of Experiment 3, namely that subjects' tendency to repeat utterances from the story they heard when asked to retell it was more common than could be expected by coincidence. As in Experiment 3, the story was fairly long and no warning of a memory test was given. Experiment 4, moreover, ruled out the possibility that the verbatim recall in Experiment 3 was due to memory for open class lexical items, since open class items were shared between the two versions of the story.

8. Experiment 5: Incidental recall after a six-day delay

The previous experiments demonstrated evidence of verbatim memory in naturalistic recall conditions, which is consistent with the findings in Experiments 1 and 2 involving recognition paradigms. The recall studies involved a delay over the time it takes to tell the story, roughly 3–4 minutes. This interval is considerably longer than what would be considered necessary for comprehension to occur, but not necessarily long enough to demonstrate long-term verbatim memory for language. We were interested in learning whether we would see effects of verbatim memory over a longer period and in an even more naturalistic context. In particular, Experiment 5 asks whether we might see incidental effects of verbatim memory over a six-day delay. We investigated whether the language speakers are exposed to when they initially hear a description influences their own subsequent description of the same scenes. Not only did we not warn them that their memory would be tested, but we also did not ask them to recall the sentences that they had heard.

8.1. *Participants*

Thirty-five subjects attended the initial session, and out of those, twenty-five returned for the additional session. One subject's data was lost due to technical difficulties with recording, leaving the study with twenty-four total participants, twelve for each script version. Participants were compensated with either $8 or experimental credit towards a course requirement for the first part of the study. For returning roughly a week later, they received either additional experimental credit or $15.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

8.2. *Procedure*

Participants were brought into the lab and asked to sit at a small table, across from what was ostensibly another participant. In reality, however, the other participant was a trained confederate. The experimenter provided a cover story, telling each subject they would be participating in a study that investigated how well people communicate when they cannot see each other. A barrier was then placed between the participants, occluding their vision of each other. The experimenter went on to explain that one of the participants would be watching two video cartoon clips, and would need to explain what occurred in each clip to the other participant. Based on these descriptions, the other participant would then be asked to take a short quiz about what happened in the cartoon. The participants were told that one of them had already been randomly chosen on the basis of their student ID number to be the 'describer', while the other was assigned to be the 'listener/quiz taker.' In reality, however, the confederate was always 'randomly' chosen to be the describer, and the true participant was always the listener. Participants were also informed that their descriptions were going to be taped for later analysis.

A laptop was placed in front of the describer (the confederate). The confederate was 'reminded' to wait until the clip was entirely over, and then to give his description to his partner. The confederate's description was a completely pre-determined script, memorized ahead of time, word for word. After each clip ended, he or she described the clip using this script, completely verbatim, while attempting to sound as natural as possible. At debriefing, no participant reported any suspicion of the description sounding in any way unnatural or non-spontaneous. After hearing the description, the other participant was given a corresponding quiz. The confederate described each of the two short cartoons.

The experimenter then collected the quizzes and left the room to quickly 'score' them. Upon returning, they would tell the participants that their participation in this study was complete, but that based on their good performance on the quizzes, they were invited back to participate in a separate, additional study. If the participants asked, they were told that it would be similar to this last experiment, but designed more for people who demonstrated a high facility in interpersonal communication (as measured by quiz performance). All efforts were made to maintain the illusion that this was in fact an entirely different study, and not simply a follow-up of the previous one. This was done to preclude the possibility of participants rehearsing the descriptions they received between the first and second sessions.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

In the second session, each participant was brought back roughly six days later (3–10 days; $\sigma = 6.14$) and paired with a second participant, who was in fact another confederate. The procedure was identical to the first session, except the roles were now reversed. The participant now watched the same video clips (which they had previously heard descriptions of), and then described them to the other confederate in their own words. The second confederate ostensibly hadn't seen the clips beforehand, which was why a second confederate was used instead of the original.

The dependent variable of interest was the number of clauses that the participants used in their description which were verbatim matches to what they had previously heard in the first session. We discuss in the scoring procedure section exactly how verbatim matches were determined.

Note that participants were not explicitly asked to recognize or recall previously heard utterances. Instead, we measured how the language they had heard affected their own later descriptions.

As in Experiments 1–4, we used two different versions of the description of each video, so that we could compare how likely participants were to match the description they heard and compare that to coincidental matches to the other version of the story. The two versions described the same scenes using different grammatical constructions, while conveying the same semantic 'gist.' For example, in script version 1, one clause was, *They bark and meow back and forth*, while the corresponding clause in version 2 was, *They take turns barking and meowing*.

8.3. *Materials*

Videos: The videos were from an old 'Felix the cat' cartoon: *Felix the cat finds out*. The cartoon lasted three minutes and was spliced into two parts. The first video lasted 1:14 minutes, the second video, 1:46. The storyline of the cartoon is fairly straightforward. A boy and a cat buy a hot dog, the hot dog comes to life (acting like a 'dog'), followed by a standoff between the hot dog and the cat, which the cat ultimately wins. Although the film is completely silent, there are a few instances in which words appear in dialogue boxes above characters' heads. As discussed in the scoring section, these phrases are excluded from analysis.

Quizzes: The mock quizzes involved showing participants a series of visual freeze frames; they were asked, 'based on the description you've just heard, which of the following screen shots was most likely from the cartoon that your partner has just watched?' One of the screen shots was actually from the relevant clip, all others were foils: screen shots from other, irrelevant Felix the Cat cartoons.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

Scripts: In order to compile the most natural sounding descriptions, pilot descriptions were gathered. In these pilot sessions, participants simply watched the clips, and, while being recorded, were instructed to describe them, as if they were describing them to someone who wasn't watching. From these, the most frequently occurring clauses were divided up between the two script versions, and adapted slightly in order to make the overall script versions sound as natural as possible.

As mentioned, the scripts differed in their underlying structure while keeping the basic content the same. There was also some overlap in some of the clauses across versions (henceforth 'overlap clauses'). For example, both script versions contained the clauses 'the cat gets upset' and 'finds a coin on the ground'. Overlap such as this was unavoidable, in that there are a limited number of ways that a given proposition can be naturally expressed. There were thirteen overlapping clauses. Scripts also differed slightly in the number of clauses they contained; one version contained 62 clauses, while the other contained 65. For each script in its entirety, see the Appendix.

8.4. *Scoring procedure*

Once the subjects' descriptions were recorded, they were transcribed manually. As in Experiments 3 and 4, descriptions were then scored clause-by-clause against the scripts for the two originals. The analysis examined true matches vs. coincidental matches for each story version, in order to determine how participants' descriptions were directly manipulated by the description they had heard six days prior.

As in the previous studies, a clause was counted as a match if it either matched the heard clause verbatim (exact), or if it differed by only a single word (one word off). To count as an exact match, clauses had to be completely identical, except for the connectives *and*, *so* and *but*, which could be added or omitted without consequence. The same allowances (*and*, *so* and *but*) applied to 'one-word-off' matches as well, where the utterance differed by exactly one (other) word.

There were 13 clauses which were the same in both stories, and two additional clauses occurred in dialogue boxes in the cartoon (*Move over fellows* and *split it 50-50*). These 15 clauses were excluded from analysis. Thus in version 1, there were 47 non-overlapping, available clauses, and in version 2, there were 50.

8.5. *Results and conclusion*

As in Experiments 3 and 4, we computed an overall 2×2 ANOVA (match/coincidental match $\times$ script version 1/script version 2), which

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

yielded a significant main effect of match over coincidental match ($F(1, 22) = 32.65, p < 0.001$). This effect was also present within each story version itself, with a significant advantage for matches over coincidental matches in both version 1 ($F(1, 11) = 18.25, p < 0.005$) and version 2 ($F(1, 11) = 13.24, p < 0.005$). There was no significant difference between story versions ($F(1, 18) = 1.16, p > 0.05$).

To evaluate the presence of verbatim memory with an even more strenuous criterion, an additional 2×2 ANOVA was run, this time considering only EXACT verbatim matches (correct vs. coincidental). Again a significant main effect ($F(1, 22) = 16.91, p < 0.001$) for correct matches over coincidental matches was found. And again, this pattern held true within each story version independently: version 1 ($F(1, 11) = 12.57$, $p < 0.01$) and version 2 ($F(1, 11) = 5.037, p < 0.05$). Again, there was no significant difference between story versions ($F(1, 18) = 0.239$, $p > 0.5$).

The results are summarized in Table 5 and Figure 3 below:

Table 5. *Felix the cat finds out: Cumulative number of correct matches (to heard script version) and coincidental matches (to alternative script version)*

	One word off + Exact matches			*Exact matches Alone*		
	Correct Matches	Coincidental Matches	Total # of available* clauses**	Correct Matches	Coincidental Matches	Total # of available* clauses**
Version 1	82	35	564	46	22	564
Version 2	51	19	600	28	14	600
Overall	133	54	1164	74	36	1164

* The total number of heard clauses minus both the clauses which overlap between versions, and the clauses which occurred in text during the clip

** Total # of available clauses = (version 1: 47×12 people; version 2: 50×12 people)

Thus, overall, and for each story version, matches were more likely to come from a participant who had previously been exposed to a given version (true matches), than from participants who were exposed to the other version (coincidental matches). This held true for when 'one word off' matches were considered in addition to exact matches, and also when exact matches were considered alone. These results indicate that participants retained a significant degree of the specific, underlying structure from a description which they heard almost a week prior, and listened to only a single time. Recall that participants returned for the second session

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

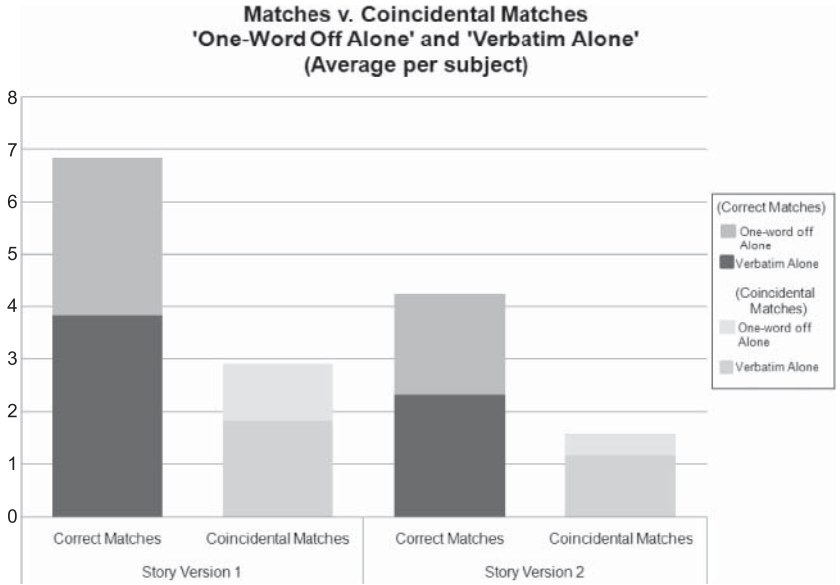


Figure 3. Felix: verbatim and one-word off - matches vs. coincidental matches

anywhere from three to 10 days after they participated in the first session. And yet we did not find a significant correlation between length of delay and true verbatim matches ($r^2 = -0.196, p = 0.4$). Thus it seems that verbatim memory for language exists and may be quite long lasting.

Participants displayed a natural tendency to reuse previously heard clauses in their own descriptions without any deliberate instruction, supporting the notion that this retention can be mediated through entirely implicit means. And yet the findings of Experiment 5 are not easy to pigeon-hole as implicit memory, which requires no conscious awareness. While participants were not asked to consciously recall verbatim utterances, the task did require that they consciously access the content of the earlier speech. The relationship between implicit and explicit memory is complicated and there may well be various levels of explicit memory (Karmiloff-Smith 1986) and different types of implicit knowledge; moreover, implicit and explicit memory may be more closely related than is sometimes assumed (Turk-Browne et al. 2006). In any case, the demonstration in Experiment 5 that retention can be both incidental (no deliberation), and implicitly mediated (no explicit recall), suggests that the retention of specific linguistic formulations is not only possible, but also an aspect of *natural* language processing.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

9. Conclusions and implications

We have seen that explicit verbatim memory for language exists in naturalistic contexts, in which subjects are exposed to fairly long, non-'interactive,' coherent stories, when no warning of a memory task is given, and when possible lexical memory effects are controlled for. Significant explicit verbatim memory has been demonstrated in recognition by above-chance performance. Explicit verbatim memory has also been shown in recall. Experiment 5 demonstrates that verbatim memory plays a role in *implicit recall*, or spontaneous production, even after an average delay of six days.

The emphasis in early studies on the lack of verbatim memory for language was based on comparisons between memory for meaning and memory for structure. When compared directly, it is very clear that meaning always trumps structure. At the same time, we also appear to have an impressive ability to recognize (Experiments 1 and 2), recall (Experiments 3 and 4), and reuse (Experiment 5) specific clauses verbatim.

Certain researchers have recognized that participants are able to recall full sentences shortly after they are heard without acknowledging the existence of verbatim memory. In particular, Potter and Lombardi have demonstrated that at least short-term *apparent* verbatim recall exists (Potter and Lombardi 1990, 1998; Lombardi and Potter 1992), but they presuppose that it is not actual verbatim memory for language. Instead, they argue that participants remember only meaning or gist, and appear to recall utterances verbatim because the relevant words (Lombardi and Potter 1992; Potter and Lombardi 1990) and structures (Potter and Lombardi 1998) have been primed. Therefore when participants attempt to reformulate the relevant meaning, the already activated words and structures result in the production of previously witnessed utterances. They state, "the surface syntax of the to-be-recalled sentence is not directly represented in memory, but is regenerated using normal mechanisms of sentence production" (Lombardi and Potter 1992: 713).

There is little doubt that lexical and structural priming exists, as does memory for gist, but these influences cannot account for the present findings. In particular, lexical priming cannot be appealed to because verbatim clauses are recognized and recalled even when primable (open-class) lexical items are controlled for (Experiments 2 and 4). Moreover, lexical priming is supposed to be short-lived, and yet participants reliably reuse clauses they witnessed an average of six days prior (Experiment 5). In addition, general meaning or gist was controlled across the two scripts used in each of the five studies. Structural priming alone cannot possibly account for the present findings either, because it cannot explain why

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

participants reliably used a particular structure in a particular context. That is, structural priming is by definition independent of propositional content, and yet participants in the present experiments reliably used a given structure in the clause where it had been originally witnessed and not in other clauses. There is no evidence that passives (or actives, or ditransitives, etc.) were primed overall; instead, a passive tended to be used to describe a particular scene only when the same scene had been described with a passive.

One way in which it is possible to view the results without appealing to verbatim memory for form is based on the idea that differences in form signal subtle differences in construal: participants may be remembering *They bark and meow back and forth* instead of *They take turns barking and meowing* because the two formulations construe the scene in slightly different ways. In fact, it can be argued that every difference in form indicates a slight difference in construal (Bolinger 1977; Langacker 1987). We attempted to mitigate such differences as much as possible; for example, viewing the cartoon in Experiment 5 made clear exactly who was doing what, how and when. To the extent that slight differences in construal remain due to the differences in the verbal descriptions, it is possible that participants remembered the very detailed construals—construals which are in effect isomorphic to the formal representations, rather than the formal representations themselves. Present results cannot adjudicate between these two possibilities. In any case, in order for the earlier-heard utterance to affect the way the scene is construed six days later, the earlier utterance—or the highly specific construal directly associated with it—must be retained.

A great deal of research within the 'usage-based' model of language learning has been aimed at understanding when and why learners make the particular generalizations they do. The present studies provide the somewhat belated finding that learners retain remarkably detailed memories for language, over which generalizations and abstractions may readily be formed. That is, the present findings lend credence to the possibility that language is learnable on the basis of categorizing over the input: possibly partially abstracted utterance tokens are retained. These may be categorized into more abstract constructional schemas (cf. e.g. Bybee 1985; Goldberg 2006; Langacker 1988; Tomasello 2003).

The present studies only open the door to future research in this area. For example, the relationship between implicit and explicit memory for language deserves further examination. Other questions are also outstanding. How much time is required before verbatim memory decays? Are certain types of constructions easier to recall or recognize than others? Factors such as construction type, lexical and construction fre-

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

quencies, sentence length and complexity could well be relevant factors. The present studies make clear, however, that there is something to be studied. Explicit and implicitly mediated verbatim memory for language exists.

References

Story books

Figueroa, A. 2002. *Spiderman: I am Spiderman*. Illustrated by Columbia Pictures.
Roberts, B. 2000. *Gramps and the Fire Dragon*. Illustrated by Melissa Iwai.

Movie stimuli

Felix the Cat (1924). *Felix the Cat Finds Out.* Sullivan Studios.

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. 149–162.
Anderson, J. R. & G. Bower. 1973. A propositional theory of recognition memory. *Memory and Cognition* 2(3). 406–412.
Bartlett, F. C. 1932. *Remembering: A study in experimental and social psychology*. London: Cambridge University Press.
Binet, A. & V. Henri. 1894. La mémoire des phrases (mémoire des idées). *L'Année Psychologique* 1. 24–59.
Bock, K. & W. F. Brewer. 1974. Reconstructive recall in sentences with alternative surface structures. *Journal of Experimental Psychology* 103(5). 837–843.
Bolinger, D. 1977. *Meaning and form*. London: Longman.
Bransford, J. D. & J. J. Franks. 1971. The abstraction of linguistic ideas. *Cognitive Psychology* 2. 331–350.
Bybee, J. L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
de Villiers, P. D. 1974. Imagery and theme in recall of connected discourse. *Journal of Experimental Psychology* 103. 263–268.
Gernsbacher, M. A. 1985. Surface information loss in comprehension. *Cognitive Psychology* 17. 324–363.
Gibbs, R. W. 1981. Memory for requests in conversation. *Journal of Verbal Learning and Verbal Behavior* 20. 630–640.
Goldberg, A. E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Grice, P. H. 1975. Logic and Conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics, volume III: Speech acts*, 41–58. New York: Academic Press.
Holtgraves, T. 2008. Conversation, speech acts, and memory. *Memory and Cognition* 36(2). 361–374.
Jarvella, R. J. 1973. Co-reference and short-term memory for discourse. *Journal of Experimental Psychology* 98. 426–428.
Johnson-Laird, P., C. Robins & L. Velicogna. 1974. Memory for words. *Nature* 251. 704–705.
Johnson-Laird, P. & R. Stevenson. 1970. Memory for syntax. *Nature* 227. 412.
Karmiloff-Smith, A. 1986. From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition* 23. 95–147.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

Keenan, J. M., B. MacWhinney & D. Mayhew. 1977. Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior* 16. 549–560.
Kintsch, W. & E. Bates. 1977. Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory* 3. 150–159.
Kintsch, W., D. Welsch, F. Schmalhofer & S. Zimny. 1990. Sentence memory: A theoretical analysis. *Journal of Memory and Language* 29. 133–159.
Langacker, R. W. 1987. *Foundations of cognitive grammar Volume I*. Stanford: Stanford University Press.
Langacker, R. W. 1988. A usage-based model. In B. Rudzka-Ostyn (ed.), *Topics in cognitive linguistics*. Philadelphia: John Benjamins.
Loebell, H. & K. Bock. 2003. Structural priming across languages. *Linguistics* 41(5). 791–824.
Lombardi, L. & M. C. Potter. 1992. The regeneration of syntax in short term memory. *Journal of Memory and Language* 31. 713–733.
Murphy, G. L. 1992. Comprehension and memory of personal reference: The use of social information in language processing. *Discourse Processes* 15. 337–356.
Murphy, G. & A. M. Shapiro. 1994. Forgetting of verbatim information in discourse. *Memory and Cognition* 22. 85–94.
Peterson, R. & C. McIntyre. 1973. The influence of semantic 'relatedness' on linguistic integration and retention. *American Journal of Psychology* 86(4). 697–706.
Potter, M. C. & L. Lombardi. 1990. Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29. 633–654.
Potter, M. C. & L. Lombardi. 1998. Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38(3). 265–282.
Reyna, V. F. & B. Kiernan. 1994. The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instruction, and retention interval. *Developmental Psychology* 30. 178–191.
Sachs, J. S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics* 2(9). 437–443.
Tomasello, M. 2003. *Constructing a language: A usage-based theory of language acquisition*. Boston: Harvard University Press.
Turk-Browne, N. B., D.-J. Yi & M. M. Chun. 2006. Linking implicit and explicit memory: Common encoding factors and shared representation. *Neuron* 49(6). 917–928.
Zimny, S. 1987. *Recognition memory for sentences form discourse.* Boulder, CO: University of Colorado PhD dissertation.

APPENDIX : COMPLETE STORY VERSIONS

Spiderman Story Version #1 Acton Figueroa (2002) *Spiderman: I am Spiderman*. Illustrated by Columbia Pictures.
Maybe you've heard of me. Maybe you've seen me. I am Spider-Man.

I wasn't always a superhero. I used to be a regular guy, just like everyone else. Well, not exactly like everyone else. I was sort of a geek. I really like school. But it wasn't always easy for me. I didn't always fit in. Some of the kids didn't like me. Everything changed when we went on a field trip to learn about spiders. Man-made spiders. Super-spiders. I used to be skinny. I used to have to wear glasses. Not anymore. Not since that spi-

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

der bit me. Now I am fast. Now I am strong. I don't like to fight, but sometimes I have no choice. The kids at school were amazed at my new strength. Like the spider that bit me, I have an extra sense. My spider sense warns me when someone needs my help. After the spider bit me, I began to make webbing. At first I couldn't control it. So I practiced. Now I can take aim and shoot my webbing wherever I want it to go. I swore to use my powers to help the people of this city. But I had to protect myself. If the bad guys knew who I was, my family would be in danger. And so I became Spider-Man. Like a spider, I can climb walls. My fingers can stick to anything. I can swing through the city on a strand of webbing. I can flip, dip, and whirl. I am strong enough to fight four bad guys at once! I can go places no one else can. As Spider-Man, I am always on the lookout for evil. Sometimes it comes looking for me. I'll always be Peter Parker. But when people are in danger, I am Spider-Man.

Spiderman Story Version #2

You may know me. You may have seen me. Spiderman is my name. I didn't used to have superpowers. I was just like everyone else. Actually, I guess I wasn't really ever just like everyone else. I was called a geek. School was interesting. But I had a hard time. Fitting in was the problem. At school, I wasn't liked by some of the kids. One day, we took a field trip to learn about spiders. That day everything changed. I used to be skinny and wear glasses. But not after that spider bit me. I am super speedy now. I'm also super strong. I don't like fighting, but I don't have any choice sometimes. My new strength amazed my classmates. I have an extra sense just like the spider that bit me. When my help is needed, I get a warning from my spider sense. I started making webs after the spider bit me. I didn't know how to do it at first. After long hours of practice, I am now able to shoot my webbing wherever I aim it. The people of this city needed me so I promised to help them. But if my identity were known by the bad guys, I could put my family in danger. So I became Spiderman. Just as a spider can climb walls, so can I. I can stick my fingers to anything. A strand of webbing helps me swing through the city. I can whirl and flip and dip. Fighting four guys at once is easy for someone as strong as me. No one else can go where I can go. Spiderman is always on the lookout for evil. Sometimes I come face to face with it. I'm still Peter Parker. But I'm Spiderman if anyone needs me!

Gramps Story Version #1 Bethany Roberts (2000). *Gramps and the Fire Dragon*. Illustrated by Melissa Iwai.

"It's bedtime," said Jesse. "But I'm not sleepy." "Me neither," said Gramps. And they rocked in front of the fire, back and forth, back and

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

forth. "Look into the fire," said Gramps. "Can you see pictures? I can see an old twisted apple tree." "Me, too," said Jesse. Together, they saw a flower garden, and a path leading to a castle. And then—they saw a fire dragon. "Yipes!" cried Jesse. The fire dragon roared and began to chase them, down the path, through the flower garden, and up the tree! Gramps tossed apples at the dragon. "Have a snack," he said. But that only made the dragon angry. Gramps and Jesse hitched a ride on a passing hot-air balloon. Roaring, the dragon followed. And they rocked in front of the fire, back and forth, back and forth. "Look into the fire," said Gramps. Jesse and Gramps zipped down a high mountain. The dragon zipped too. They raced through a jungle. The dragon raced, too. They crossed a wide river, The dragon crossed, too. Then they ran down a long dark tunnel. But the dragon ran, too. The dragon was so close, his flames licked Jesse's heels. "Run for your life!" shouted Gramps. The dragon moved towards Gramps. I'll save you, Gramps!" cried Jesse. Quickly, he flagged down a fire truck. Jesse grabbed a hose and sprayed the dragon. The dragon got smaller, and smaller, and smaller, until there was no dragon at all. "Just embers left," said Gramps, looking into the fire. "I guess I saved you, then," said Jesse. "You sure did," said Gramps. Gramps yawned. Jesse yawned. Then they rocked in front of the fire, back and forth, back and forth, until they fell asleep, together.

Gramps Story Version #2

"Time for bed" said Jesse. "But I'm not at all sleepy." "I'm not either" said Gramps. And they rocked and rocked, back and forth, in front of the fireplace. "Take a look at the fire" said Gramps. "Are you able to see pictures?" "I see a twisted old apple tree." "I do too" said Jesse. Together, they could see a flower garden, and a path that led to a castle. Just then—a fire dragon appeared. Jesse cried, "Yipes!" Roaring, the fire dragon started to chase them. He chased them down the path and through the flower garden and up the tree. Gramps tossed the dragon apples. "Snack on these" he said. But that only angered the dragon. A hot air balloon passed, and Gramps and Jesse hitched a ride on it. With a roar, the dragon followed them. And they rocked and rocked, back and forth, in front of the fireplace. "Take a look at the fire" said Gramps. Jesse and Gramps went down a high mountain, zipping along. The dragon went zipping along, too. They went racing through a jungle. The dragon went racing, too. They went across a wide river, The dragon went across, too. Then they ran through a long dark tunnel. But the dragon ran through the tunnel, too. The dragon got so close, his flames licked at Jesse's heels. Gramps shouted, "Run for your life!" The dragon ap-

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

proached Gramps. Jesse shouted, "Ill save you, Gramps!" Jesse grabbed a hose. He sprayed the dragon. The dragon got smaller, smaller, smaller ... until there was no dragon at all. "There are just embers left" said Gramps as he looked into the fire. "I saved you, didn't I?" said Jesse. "You surely did" said Gramps. Gramps yawned and Jesse yawned. Then they rocked and rocked, back and forth, in front of the fire. Until they both fell asleep.

Felix the cat finds out, script version #1

(Clip 1)
There's this boy and this cat. The boy is studying for classes, reading a book and looking frustrated like he doesn't understand it. The cat tries to get the boy to play with him but the boy refuses because he's studying for his school lessons. The cat gets upset and is just waiting for the boy to finish when he finds a coin on the ground. He shows the coin to the boy and they both get really excited. The boy thinks they should go to a hot dog stand and they buy a hot dog. They leave the hot dog stand and the boy has the hot dog and is about to eat it but the cat gets upset because he wants it. So they decide to flip a coin for it but it lands exactly on its rim so they decide to split the hot dog 50-50.

(Clip 2)
So they decided to split the hot dog 50-50 but all of a sudden the hot dog barks, jumps out of the bun, runs back to the hot dog vendor, he says "move over fellows" and lies down between the other hot dogs. So the cat and the boy are kind of upset and confused by this. So the cat goes over to the hot dog vendor and meows and the hot dogs bark back. The cat taunts them and they bark and meow back and forth until all of a sudden the hot dogs jump off the stand and start chasing the cat. They eventually chase him into what looks like a dog house and there's a big fight inside. The boy goes to the dog house to see what's going on. After a while the fighting ends and the cat pokes his head out and motions for the boy to come into the dog house and he goes inside. Then they come out happy and rubbing their bellies as if they had eaten all of the hot dogs.

Felix the cat finds out, script version #2

(Clip 1)
There's this cat and this boy. The boy is studying for classes but looking frustrated because he doesn't understand what he's reading. The cat wants to play with him but the boy says he can't since he has to study

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press

for his school lessons. This makes the cat upset and he just stands there waiting for the boy to finish when he finds a coin on the ground. He gets the boy's attention by showing him the coin and the two of them get really excited. The boy gets the idea to spend it on a hot dog so they go to a hot dog vendor and they buy a hot dog. They leave the hot dog vendor and the boy has the hot dog and is ready to eat it but the cat gets upset because he thinks he should have it. So they decide to flip a coin for it but it lands exactly on its edge. So they settle that they'll split it 50-50.

(Clip 2)
So they had decided on splitting the hot dog 50-50 but suddenly the hot dog goes woof, leaps from the bun and runs back to the hot dog vendor, tells the other hot dogs "move over fellows" and lies back down among the hot dogs. This makes the boy and cat upset and confused. So the cat goes over to the hot dogs and taunts them by meowing and the hot dogs bark back. They take turns barking and meowing until suddenly the hot dogs leap from the stand and start to chase the cat. Eventually they chase him into what appears to be a dog house and there's a huge fight inside. The boy comes up to the dog house to see what's going on. The fighting stops after a while, and the cat sticks his head out and waves the boy into the dog house and he goes inside. They come out rubbing their stomachs and looking happy like they had eaten all of the hot dogs.

https://doi.org/10.1515/langcog.2010.003 Published online by Cambridge University Press