

ORIGINAL ARTICLE

The role of linguistic factors in the retention of verbatim information in reading: An eye-tracking study on L1 and L2 German

Andreas Opitz¹ , Denisa Bordag^{1,2} and Alberto Furgoni¹

¹Herder-Institut, Leipzig University, Leipzig, Germany and ²University of Haifa, Haifa, Israel

Corresponding author: Andreas Opitz; Email: andreas.opitz@uni-leipzig.de

(Received 15 June 2023; revised 23 January 2024; accepted 5 March 2024)

Abstract

We investigated the retention of surface linguistic information during reading using eye-tracking. Departing from a research tradition that examines differences between meaning retention and verbatim memory, we focused on how different linguistic factors affect the retention of surface linguistic information. We examined three grammatical alternations in German that differed in involvement of changes in morpho-syntax and/or information structure, while their propositional meaning is unaffected: voice (active vs. passive), adverb positioning, different realizations of conditional clauses. Single sentences were presented and repeated, either identical or modified according to the grammatical alternation (with controlled interval between them). Results for native ($N = 60$) and non-native ($N = 58$) German participants show longer fixation durations for modified versus unmodified sentences when information structural changes are involved (voice, adverb position). In contrast, mere surface grammatical changes without a functional component (conditional clauses) did not lead to different reading behavior. Sensitivity to the manipulation was not influenced by language (L1, L2) or repetition interval. The study provides novel evidence that linguistic factors affect verbatim retention and highlights the importance of eye-tracking as a sensitive measure of implicit memory.

Keywords: eye-tracking; grammatical alternation; information structure; L1 vs. L2; reading; verbatim retention

Introduction

Recent research on the retention of low-level, surface linguistic information vs. meaning has challenged the previous dominant view that verbatim memory decays as soon as superior structures like propositional representation are built and that verbatim information is thus typically not retained in long-term memory. According to Gurevich and colleagues (Gurevich et al., 2010, p. 49), “various factors have conspired to downplay the importance of memory for language in past

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

research,” in particular since the focus had been on the memory for content or gist that always trumps the verbatim memory. Nevertheless, evidence gradually accumulates that our memory for surface linguistic information is not as negligible as assumed earlier.

In our study, we want to take one step further by showing not just that surface linguistic information is retained during reading, but also by revealing which linguistic qualities contribute to the verbatim retention of a given piece of text. In particular, we want to explore which types of syntactic structures¹ are more salient for verbatim retention. This question has importance not only for our understanding of how grammatical aspects of texts are processed during reading but also for language assessment.

Although all theories of language acquisition recognize the importance of memory for language processing and acquisition, there is a broader linguistic debate about the extent of its involvement in these processes, particularly with regard to the role of memory beyond the storage and retrieval of simple lexical units at the word or sub-word level. At one end of the spectrum of such approaches are those that assume that (working) memory primarily affects performance, with the core principles of language processing (e.g., grammatical operations such as feature checking and structure building) being fundamentally independent of memory (e.g., the performance-competence distinction in generative approaches, Chomsky, 2014). At the other end of the spectrum are approaches that emphasize the central role of memory in the development and processing of grammar.

According to the latter view (e.g., usage-based theories of grammar acquisition), sequences of words are stored verbatim in our memory and they are used to develop a mental grammar through extraction of abstract regularities about distributional and semantic-distributional relationships between words (Audring, 2022; Bybee, 1985; Ellis, 1996; Goldberg, 2006; Langacker, 1988; Tomasello, 2003).

From this perspective, the role of verbatim memory in language acquisition and representation goes beyond simply memorizing lexical units and chunks: Verbatim memory for sequences of words is the fundamental basis for the development of more abstract representations and grammar in general, which are stored in memory (e.g., in the form of schemas, as in constructional grammar; Audring, 2022; Bybee, 1985; Goldberg, 2006) and retrieved during processing.

Insights about which linguistic phenomena are more salient for verbatim retention can thus contribute to our understanding about their acquisition, in particular when viewed from the perspective of such theories.

Due to this relevance for understanding language acquisition and to recent observations of differences between verbatim memory of native speakers and non-native learners (Bordag *et al.*, 2021; Sampaio & Konopka, 2013), we tested both L1 speakers and L2 learners of German.

The claim that memory for surface linguistic structure fades off directly after a sentence has been processed goes back to the 1960s and 1970s, when several studies had been conducted that seemingly supported the conclusion (Anderson, 1974; Sachs, 1967, 1974; Soli & Balch, 1976 among many others). In these studies, participants listened to or read isolated sentences or texts and then decided whether a sentence typically presented on the screen was identical with the critical sentence presented previously. The manipulations on which the authors based their claims

about the absence of verbatim retention of surface linguistic information involved various alternations; such as the active/passive alternation (e.g., *He sent a letter about it to Galileo, the great Italian scientist* vs. *A letter about it was sent to Galileo, the great Italian scientist.*; Sachs, 1967, 1974; Anderson, 1974); a change in a position of an apposition (*He sent a letter about it to Galileo, the great Italian scientist.* vs. *He sent Galileo, the great Italian scientist, a letter about it.*; Sachs, 1967); double object constructions (*A group of interested people gave the film to some photo experts.* vs. *A group of interested people gave some photo experts the film.*; Soli & Balch, 1976); synonym substitutions (*A wealthy manufacturer, Matthew Bolton, sought out the young inventor.* vs. *A rich manufacturer, Matthew Bolton, sought out the young inventor.*; Sachs, 1974); or purely formal word order changes (*A wealthy manufacturer, Matthew Bolton, sought out the young inventor.* vs. *A wealthy manufacturer, Matthew Bolton, sought the young inventor out.*; Sachs, 1974). The ability to recognize the exact wording of such sentences was interpreted as evidence for surface form retention (verbatim memory). These alterations were then typically compared to manipulations in which the propositional meaning of the sentences was changed, such as swapping of the subject and the object (*He sent a letter about it to Galileo, the great Italian scientist.* vs. *Galileo, the great Italian scientist, sent him a letter about it.*; Sachs, 1974). The authors also manipulated the number of intervening syllables between the first and second presentation of the critical sentence (distance between 0 and ca. 160 syllables). The central finding of these studies was that while the memory for semantic information was preserved also at the longest tested distances, the memory for the surface linguistic information could not be reliably confirmed already after the intervention of only 40 syllables (i.e., ca. 20 words which makes ca. 2-3 sentences) or even earlier. Only when the second version of the critical sentence was presented immediately after the first version (syllable distance 0), both the retention of semantic and surface linguistic information were observed.

These findings were replicated in numerous variations (e.g., Bransford et al., 1972; Gernsbacher, 1985; Graesser & Mandler, 1975; Jarvella, 1971, 1979; Johnson-Laird & Stevenson, 1970) and taken for granted for a number of decades as illustrated by Von Eckardt and Potter (1985). In the reading research, conclusions about the content of mental text models constructed during reading of texts have been based on these and other studies as well (e.g., Construction-Integration Model by Kintsch, 1986, 1988, 1991, 1998). The general assumption is that “a primary property of mental models is that they represent what the text is about (the events, objects, and processes described in the text), rather than features of the text itself” (Glenberg et al., 1987, p. 69). The research on text models thus complies with the literature on memory in that it assumes that verbatim (surface) information is not replicated in the final mental representation (Garnham & Oakhill, 1996; Johnson-Laird & Stevenson, 1970).

Gradually, however, evidence started to accumulate that contradicts these findings. For example, already in 1977 Kintsch and Bates (Kintsch & Bates, 1977) observed that participants recognized sentences they had heard in a regular university lecture significantly better than their paraphrases. In 2010, Gurevich and colleagues (Gurevich et al., 2010) performed a study in which participants first listened to recorded texts from children’s storybooks while corresponding

illustrations were presented to them. In the recognition task, written clauses were displayed on the screen and subjects were asked to decide whether they had heard that exact clause (identical as presented or a paraphrase) in the story. In the recall task, participants were asked to retell the story from the pictures. The results demonstrated that explicit verbatim memory for language in naturalistic settings persists longer than had been believed and that it extends beyond memory for individual lexical items.

More recent research has confirmed that more information is indeed stored in long-term memory than what is commonly believed, even when it is processed without any attention or learning intention (Hutmacher & Kuhbandner, 2020; Kuhbandner *et al.*, 2017). In his 2020 study, Kuhbandner used a two-alternative forced choice recognition test to show that readers are able to remember exactly which word was written in a specific position in a book chapter both immediately and one week after testing, despite believing that they were just guessing. The author summarizes that phenomenal memory experience (*i.e.*, recollection or familiarity) is not indispensable (see also Craik *et al.*, 2015; Hutmacher & Kuhbandner, 2018; Voss *et al.*, 2008) and that verbatim information can be also retained through implicit memory. Detection of this type of retained information, however, requires paradigms that are “sensitive enough to also measure memory representations that are below the level of phenomenal memory awareness” (Kuhbandner, 2020, p. 2). It is conceivable that the earlier research paradigms employed in the 1960s and 1970s primarily tackled explicit memory for surface linguistic information that might decay at faster rates than our implicit memory for the same phenomena. This implication calls for the employment of methods more sensitive to the contents of unconscious and implicit memory, such as eye-tracking.

Interestingly, the only study that to our knowledge has employed eye-tracking to explore the retention of surface linguistic information during reading also addressed the question about possible differences in verbatim memory between L1 and L2 speakers (Bordag *et al.*, 2021). The assumption of potential differences between native speakers and language learners was motivated by a study by Sampaio and Konopka (2013). In a cued sentence recall procedure, the authors demonstrated that while L1 and L2 readers recall equally well the gist of a sentence like “*The bullet STRUCK/HIT the bull’s eye,*” the L2 readers showed advantages over L1 speakers in retention of the verbatim phrasing (*e.g.*, STRUCK vs. HIT). The authors thus concluded that L2 readers are more sensitive to synonymous lexical substitutions in such sentences than L1 readers. With respect to the lack of verbatim memory for lexical substitutions in L1, the study replicated the earlier results of Sachs (1974): Sachs revealed below chance ability of her participants to correctly identify changes in lexical substitutions (*e.g.*, *rich* vs. *wealthy*) as soon as 20 syllables after the presentation of the first version of a sentence and rather low ability to perform the task even directly after the presentation of the second version both in the auditory and visual versions of the experiment.

The eye-tracking study by Bordag and colleagues wanted to further investigate the putative differences between the memory for surface linguistic information in native and non-native speakers during reading. The participants read two versions of short passages in German. In the second version, critical sentences either stayed the same or were changed at the lexical level (contextual synonym substitutions,

e.g., *Theorie* vs. *Hypothese*—“theory” vs. “hypothesis”) or at the syntactic level (active/passive alternation). The reading behavior of corresponding critical sentences of the second text versions was compared with respect to whether they were in exactly the same version or in the alternative version of the manipulation that was presented during the first reading.

Overall, L2 readers displayed longer fixations in both lexical and syntactic conditions when the critical regions in the second text versions were changed compared to the first versions (e.g., “hypothesis”-“theory”; passive-active) than when they stayed the same (e.g., “theory”-“theory”; active-active). This finding indicated participants’ sensitivity to the change based on the fact that the verbatim memory of the first version of the sentence was still present in their mental text model. On the other hand, the L1 readers showed only a smaller effect for changes in the lexical condition and no effect in the syntactic condition. Bordag and colleagues hypothesized that the retention advantage for the lexical over the syntactic condition in L1 is due to the fact that substitutions of contextual synonyms affect conceptual-semantic representations. They argue that such representations become a part of mental text models both in L1 and L2. Surface linguistic representations of syntactic constructions (active vs. passive), however, become a part of mental text models in L2, but not in L1.

Importantly, the study delivers evidence of the superior ability of L2 readers to retain verbatim information compared to the L1 readers. The authors suggest explanations of this phenomenon based on several theoretical accounts, such as the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006, 2018), the Declarative/Procedural Model (Ullman, 2016), and the Fuzzy Trace Theory (FTT; Reyna & Kiernan, 1994), which suggest that the focus of the L2 learners on the surface linguistic information might compensate for their reduced ability to process and/or represent complex linguistic structures, such as syntactic hierarchies.

These results of Bordag and colleagues go beyond the findings by Sampaio and Konopka (2013) and against the results of Sachs (1974). In addition to confirming verbatim memory at the lexical level in L2 in an even broader scope than Sampaio and Konopka, Bordag and colleagues also found evidence for retention of surface information about lexical units in L1, albeit numerically smaller than in L2. This is even more astonishing since the effects were observed during reading texts and not just isolated sentences.

Previous research has shown that text reading facilitates the storage of gist representations rather than surface representations which seems to be negatively affected by the creation of coherent semantic representations (e.g., Gernsbacher, 1985; Glenberg et al., 1987). The results contradict the earlier findings of Sachs (1974), which indicated that the verbatim memory of her L1 participants for lexical substitutions was the worst of all tested conditions. The verbatim memory for the active/passive alteration also decayed rapidly in the experiment where the first version of the sentence was presented auditorily, but not when it was presented visually—in that case, the L1 registration of changes regarding the active/passive alternation was the same as for the semantic changes.

The comparison of the results of both the earlier studies with the contemporary ones and of the contemporary studies among themselves shows that there are still some unresolved issues. First, the question about retention of surface linguistic

information is far from being answered. Second, the established assumption about the fast decay of verbatim memory might be an artifact of research methods targeting phenomenal awareness instead of testing implicit memory. Sensitive methods capturing also unintentional processes, such as eye-tracking, seem to be more suitable to test the issue. Third, differences between the findings are most likely, at least to some degree, a result of various factors that affect the retention of surface linguistic information. Previous research has, among other things, indicated that retention of verbatim memory is probably affected by the presentation mode (advantage of reading vs. listening), the status of the critical region (advantage of sentences over texts Anderson & Bower, 1973; de Villiers, 1974; Peterson & McIntyre, 1973), the interval between the presentation and the testing (better verbatim memory at shorter intervals), the L1 vs. L2 language status (L2 advantage), and the type of the tested linguistic phenomenon (e.g., difference between lexical substitutions and active/passive transformations).

To contribute to the understanding of our memory for surface linguistic information, we decided to focus on the last aspect mentioned and ask the question: Which linguistic properties make the surface linguistic information more salient for verbatim retention? That such properties play a role had been speculated already in the early research. Soli and Balch (1976, p.676) for example wrote: “some wording changes which appear to be entirely formal can have semantic qualities which might affect memorability in other situations” and “The effects of such factors on memorability of linguistic information clearly deserves further investigation.”

Indeed, more contemporary research on the related question of syntactic priming has shown that its persistence does not equally generalize across different syntactic structures. For instance, studies that explored both the dative and active/passive alternations (e.g., Bernolet *et al.*, 2013; Bock *et al.*, 2007; Bock & Griffin, 2000) revealed that priming effects for active/passive were weaker and shorter-lived than for datives.

However, up until now linguistic factors of this kind have not been systematically explored in research on verbatim retention. In our current study, we want to contribute to closing this research gap by focusing on verbatim memory for several syntactic structures that differ in the degree in which they affect the grammatical and the semantic/functional level.

Since it is the first research of this sort, we decided to create conditions that are most favorable for the detection of verbatim memory effects, if they exist. Therefore, we decided to explore isolated sentences that are presented in a written mode (reading) and to test the verbatim memory with two relatively short gaps between presentation and testing. We also decided to test both L1 and L2 participants and to employ a similar eye-tracking paradigm as Bordag *et al.* (2021), since it proved sensitive to implicit verbatim memory effects during reading.

The present study

For the present study, we selected three German syntactic structures with different properties to explore their salience for verbatim retention in sentential context:

passive-active alternation (voice), alternation of the position of an adverbial adjunct, and alternation of two types of conditional clauses.

The first structure is the passive-active alternation, (e.g., *Ein Mädchen rettet den Touristen aus einer peinlichen Situation.* vs. *Der Tourist wird von einem Mädchen aus einer peinlichen Situation gerettet.*—“The girl saves the tourist from an awkward situation.” vs. “The tourist is saved by the girl from an awkward situation.”). The alternation is of morphosyntactic nature and provides guidance on how a sentence shall be constructed, i.e., it has a grammatical function. Importantly, argument structure plays a role, too. In active sentences, the abstract case (nominative vs. accusative), the semantic/theta roles (agent vs. patient), and grammatical functions (subject vs. object) are aligned differently than in passive sentences. In active sentences, the agent is expressed by the subject and the patient by the object. In passive sentences, the patient becomes the subject, and the agent is expressed by a prepositional phrase. Consequently, there is also a change in grammatical case which is reflected overtly morphologically: the semantic role of patient, which is expressed as a direct object in accusative case in active sentences, is in nominative case when it is in the subject position in the passive sentence (e.g., *den Touristen*(accusative) vs. *der Tourist*(nominative)). In parallel to English, the German active is a synthetic form and the passive an analytic form consisting of an auxiliary and a participle.

In addition, the alternation affects the functional relationship between the sentence elements, i.e., it also has a communicative function. The passive voice shifts the focus from the agent to the action and the patient and thus changes the information structure of the sentence (functional sentence perspective). The changes on the surface level include morphological changes (e.g., case marking), word order changes, and also the number of words (with additional functional words in passive voice: the auxiliary and the preposition).

The second alternation involves the position of adverbial adjuncts in a sentence (e.g., *Bei der Sitzung zählt die Vorsitzende die Stimmen.* vs. *Die Vorsitzende zählt bei der Sitzung die Stimmen.*—“At the meeting, the chairwoman counts the votes” vs. “The chairwoman counts the votes at the meeting.”). The adjuncts were located either at the very beginning of the sentence or in its central field before the object. Both positions are possible in the relatively free German word order. The alternation does not have any grammatical implication, it affects only the information structure of the sentence, i.e., it has a communicative function. The primary difference is whether the element is in the topic or in the comment. When the adverbial is at the first position, it is a part of the topic and thus in focus. When the adverbial is in the central field of the sentence, i.e., comment, the emphasis is removed, the adverbial is no longer in focus and has a less prominent communicative function in the sentence. The adjuncts in our present study consisted either of a prepositional phrase comprising a preposition, an article, and a noun (e.g., *bei der Sitzung*—“at the meeting”) or a one-word adverb such as *manchmal* (“sometimes”). On the surface level, the only change is in the word order.

The third alternation concerns two realizations of irreal conditional clauses (also called Type III conditional): *Wenn die Sängerin mehr geübt hätte, hätte sie im Wettbewerb den ersten Platz belegt.* vs. *Hätte die Sängerin mehr geübt, hätte sie im Wettbewerb den ersten Platz belegt*—“If the singer had practiced more, she would have come first in the competition.” vs. “Had the singer practiced more, she would

Table 1. Comparison across alternations of the most relevant linguistic characteristics

	Active/ passive	Adverbials	Conditional clauses
Change of word order	+	+	+
- of functional words	+	+	+
		(For adverbial PPs)	
- of lexical words	+	+	-
Morphological changes	+	-	-
Change in functional sentence perspective (information structure)	+	+	-
Number of additional (functional) words	2	0	1

have come first in the competition.” This alternation changes the grammatical structure of the sentence. However, it affects only syntactic aspects and not the morphological ones. The two versions of conditional clauses differ in the position of the finite verb. In the one that starts with the subordinating conjunction *wenn*—“if,” the finite verb (here the auxiliary *hätte*—“had”) is at the final position of the subordinate clause (which is typical for German subordinate clauses). The other version starts with the finite auxiliary verb *hätte(n)*—“had” or *wäre(n)*—“was/were.” The alternation does not affect the word order of lexical elements in the sentence, thus there is no change in the information structure. Neither German grammars nor the intuition of the native speakers offer any indication about any semantic or communicative differences between the two realizations of conditional sentences: they seem to be fully equivalent in their functions, with the one starting with the conjunction *wenn* being somewhat more frequent. The alternation, therefore, seems to be of purely formal nature. On the surface level, the word order of one function word (the auxiliary) changes as well as the number of words (additional subordinate conjunction *wenn*—“if”).

All three types of alternation affect word order. However, while the passive/active and the adverbial alternations involve word order changes of lexical words, the word order changes of the conditional alternation involve only functional words. This aspect can be significant, since recent eye-tracking research indicates that functional words are less likely to be registered than content words (Drieghe *et al.*, 2008; Staub *et al.*, 2019). Related to this difference, while the first two alternations involve changes in functional sentence perspective, the conditional alternation does not. Contrary to the adverbial and conditional alternations, the active/passive alternation as a change of the verbal category voice also involves morphological changes including that of case (object in accusative case in the active sentence becomes a subject in nominative case in the passive sentence) and the corresponding deeper level changes of argument structure alignment. On the other hand, as opposed to the passive/active and the conditional alternations, the adverbial alternation does not involve any change of the number of words in the sentence (see Table 1).

Based on these differences, we consider the passive/active alternation as most salient for retention. The possible difference in the retention salience between the

adverbial and conditional clauses depends on the role that formally grammatical and communicative aspects play in verbatim retention.

With respect to formal differences, both alternations involve word order changes, in particular different beginnings in the two alternation realizations. If verbatim memory is related to the form aspects, both structures should be retained equally well. If the communicative function plays a primary role, there should be an advantage in retention for the adverbial manipulation.

During the experiment, participants had to read sentences while their eye movements were tracked. Two critical sentences with the same propositional content were always presented either with 2–4 intervening sentences (short distance) or with 12–14 intervening sentences (long distance). Our long distance was thus slightly longer (ca. 180 syllables) than the usual longest tested distances in the 1960s and 1970s studies (about 160 syllables). The order of the presentation of the pairs of critical sentences with the same proposition was manipulated such that either one sentence realization was presented twice (“same” condition: e.g., active-active, passive-passive) or two different realizations were presented (“changed” condition: e.g., active-passive, passive-active). We compared reading behavior on the identical realization in the “same” vs. “changed” condition. The rationale of the manipulation was that if participants retained only the semantic content of the sentence (its proposition), there should be no difference in the reading times in the “same” vs. “changed” conditions. If, however, they also retained the surface linguistic information, longer reading times were expected in the “changed” than in the “same” condition. Similarly to Bordag et al. (2021), we assumed that the repeated lexical material acted as a retrieval cue for what has been retained during the first reading of the proposition. If the surface linguistic structure was retained as well, its repetition in the “same” condition should facilitate the reading while the lack of the overlap on the verbatim level should lead to processing costs that would be manifested in longer reading times in the “changed” condition.²

Methods

We performed Experiment 1 with native German speakers and Experiment 2 with L2 German learners. Since both experiments employed the same materials and methodology, we report them together both in the Methods and in the Analysis sections.

Participants

In Experiment 1, 60 L1 German speakers (mean_{age} = 25.9 years; SD = 5.3 years; range = 19–40 years) were tested. In Experiment 2, 60 L2 German speakers with a Slavic L1 participated in this study. The participants’ German knowledge was (pre)-advanced, and it ranged between B2 and C1 level of the Common European Framework of Reference (CEFR). These data were collected through questionnaires and a vocabulary test taken from the DIALANG project (Alderson, 2005). Due to tracking problems that resulted in too much data loss, data of two of the L2 participants had to be removed. There were thus 58 remaining L2 participants (47 Czech, 4 Russian, 2 Polish, 3 Ukrainian, 1 Bulgarian, 1 Slovakian; mean_{age} = 23.8

years, SD = 4.1 years; range = 18–29 years). The participants had normal or corrected to normal sight and reported neither reading nor cognitive impairments. They signed a consent form before the experiment and received monetary compensation.

Materials

The stimulus list consisted of 230 sentences: 144 were experimental sentences, 73 were filler sentences, and 13 were used for a familiarization phase (see supplementary material). The final set of experimental sentences was based on two pretests. In the first pretest, 16 L1 German speakers rated the materials on plausibility and naturalness using a seven-point Likert scale. In the second pretest with L2 German learners, we controlled that all vocabulary is well known to L2 learners at the B2 level. Neither the L1 nor the L2 participants of the pretests participated in the actual experiments. Filler sentences comprised various tenses and many different simple syntactic structures, including those of the experimental sentences. The experimental sentences addressed the three types of syntactic alternations that corresponded to three manipulations in our experiments (*Voice*, *Adverbial position*, *Conditional Clause*). Each manipulation comprised 24 sentence pairs which presented two surface realizations of the same proposition (48 sentences).

In the *Voice* manipulation, the propositions were realized in the active (see example in (1)) and in the passive forms (see example in (2)).

(1) *Die Händler belügen die Kunden.* “The traders deceive the customers.”

(2) *Die Kunden werden von den Händlern belogen.* “The customers are deceived by the traders.”

Half of the sentence pairs were semantically neutral, i.e., the sentence was equally plausible if the agent and the patient were swapped. The other half was semantically biased so that the given agent and patient assignment was clearly preferred to the reverse. Furthermore, both patients and agents were always expressed through full NPs and never through pronouns to avoid ambiguous bindings in the passive sentences.

In the *Adverbial position* manipulation, the propositions were realized with an adverbial adjunct either in the first position (i.e., first field) with consequent subject-verb inversion (see example in (3)) or in the central field (see example in (4)).

(3) *Jahrelang blieb die Krankheit unbemerkt.* “For years, the disease remained undetected.”

(4) *Die Krankheit blieb jahrelang unbemerkt.* “The disease remained for years undetected.”

Additionally, the critical adverbial was never the last element of the sentence. The adverbial adjuncts were realized either as single-word adverbs (12 sentences) or as prepositional phrases consisting of a preposition, an article, and a noun (12 sentences).

In the *Conditional clause* manipulation, the propositions were realized with Type III conditional clauses, realized with verbs in the irrealis mood. The clauses started either with the conjunction *wenn* “if” (see example in (5)) or with the reduced conditional (see example in (6)).

Table 2. Distribution of the sentences in each experimental condition per each experimental list

Manipulation	Combinations per experimental list	
Voice	Active	6 × active-active
	Passive	6 × active-passive 6 × passive-passive 6 × passive-active
Adverbial position	First field	6 × first-first
	Central field	6 × first-central 6 × central-central 6 × central-first
Conditional clause	Reduced	6 × conjunction-conjunction 6 × conjunction-reduced
	Conjunction	6 × reduced-reduced 6 × reduced-conjunction

(5) *Wenn der Schüler die Lehrerin gefragt hätte, hätte sie ihm geholfen.* “If the student had asked the teacher, she would have helped him.”

(6) *Hätte der Schüler die Lehrerin gefragt, hätte sie ihm geholfen.* “Had the student asked the teacher, she would have helped him.”

In both instances of the alternation (conjunction present or absent), the canonical and least marked order, in which the conditional clause preceded the main clause, was used. Modal verbs were excluded to avoid too complex constructions. Pronouns were permitted only as resumptive pronouns in the second part of the clause that referred back to an NP already introduced in the conditional clause. Similarly, negation was allowed to appear only in the invariant main clause.

For each pair in each manipulation, four combinations were constructed. In the “same” combination, each pair member was repeated (e.g., passive-passive; active-active). In the “changed” combination, the two different members were combined in the two possible orders (e.g., active-passive; passive-active). Four experimental lists were constructed that differed in the combination in which the pair members were presented. Each proposition appeared once in each experimental list, but only in one of the four possible combinations. Conditions and items were distributed over these lists according to a Latin square design.

Overall, a participant read six instances of each combination (24 for each alternation), 73 filler sentences and 13 familiarization items, which sums up to 158 sentences (see Table 2). A list with all critical sentences and fillers is provided as supplementary material.

Apparatus and procedure

The experiment was programmed with the software EyeLink Experiment Builder (version 2.3.38; SR Research, 2020), and it was run on a Asus ROG Zephyrus S17 laptop connected to a Lenovo T480 laptop via a LAN cable. The eye tracker was an EyeLink Portable Duo with a 500 Hz sample rate setting installed on a laptop mount and participants positioned their chin on the EyeLink Table Clamp Chin Cup to stabilize the head.

The participants were tested individually in a silent, darkened room, and they sat comfortably in front of the presentation laptop about 60cm from the eye tracker. After reading the instructions, the eye tracker was calibrated for both eyes, and the calibration was successively validated. Next, the participants underwent a familiarization phase, and then, they would start the actual experiment which was divided into 4 blocks of roughly 50 sentences each.

The order of the trials was pseudo-randomized with the restriction that no more than two instances of the same type of alternation/condition (i.e., Voice, Adverbial position, or Conditional clauses) could follow each other. In each block, there were three instances of same/change combinations for each of the three alterations. That means that in each block, the same proposition appeared twice, either in the “same” or in the “changed” combination.

The distance between the first and second presentation of each proposition was controlled within each block in such a way that they were either in short distance (with 2–4 intervening sentences) or in long distance (with 12–14 intervening sentences).

This implies that there were no breaks between presentations of two realizations of the same proposition. The manipulation of distance was completely cross-balanced over subjects and items. Within each of the four experimental lists, half of the items in each condition was presented with a short repetition interval and half with a long interval. Then, for each of the four experimental lists a second version was created which differed from the initial list version in the assignment of short-versus long-distance repetition of the critical items and conditions. This resulted in a total of eight experimental lists in which each participant read half of the critical sentences in each condition with a short and half with a long-distance repetition and each item appeared equally often in each condition with a short and long repetition interval across all lists.

To check the participants’ sustained attention, a one-back question was pseudo-randomly presented six times in each block. The participants had to answer whether they had already read the given sentence earlier in the block. Half of the questions referred to trials with propositions that appeared the first time in the experiment and thus differed in their meaning from all previous sentences (always a filler sentence, expected No-response), while the other half referred to trials that were exact repetitions of previously presented sentences (expected Yes-responses). The questions of the one-back task were thus constructed such that they kept the participants’ attention focused on the meaning (i.e., the proposition) of the sentences, but did not attract it to their verbatim (i.e., surface form) information. After each block, there was a short break, and the eye tracker was re-calibrated at the beginning of the following block while the first trial after each pause was a filler trial. Overall, the experimental session lasted around 45 minutes.

The screen had a 17” size with a 1920 × 1080 pixels resolution. The text of the sentences was in black with a monospaced, 20-point size presented in sans-serif with a proportional font (“consolas”) and on a light gray background. The text was vertically centered and left-aligned with a 200-pixel (ca. 4 cm) margin.

Each trial started with EyeLink’s drift correction point that functioned as a fixation point.

It was positioned horizontally centered on the left of the screen where the first word of the next sentence would appear. As soon as it was fixated on by a participant, the experimenter pressed a button from the control laptop to display the sentence. After reading the sentence, the participants moved their gaze to an arrow at the bottom-right corner of the screen that triggered the next trial if fixated upon for at least 350 ms. If the one-back question came up, participants responded by pressing the upper right button of a response box for “yes” and the upper left button for “no.”

Data pre-processing and analysis

Prior to statistical analyses, all eye-tracking data were pre-processed with the software DataViewer (version 4.1.211; SR Research, 2019) for the detection of fixations and saccades with the software’s default settings. Additionally, the software’s automatic 4-stage fixation cleaning with standard settings for minimum and maximum fixation durations was performed.³ Interest areas were automatically defined with the software’s default settings for text with 30-pixel margins around the text. For the analyses reported below, an interest area was defined as the whole clause. For the Voice manipulation, as can be seen in examples (1) and (2), both alternating versions differ starting from the initial NP (agent vs. patient/theme of the sentence) till the end of the sentence (patient/theme in active voice vs. participle of the lexical verb in passive voice). Similarly, for the Conditional manipulation, the auxiliary is in the last position, preceded by the lexical verb when it is realized with the *wenn* (“if”) subjunction (*Wenn Paul Spanisch gelernt [VERB] hätte [AUX], [. . .]*—“If Paul had learned Spanish, [. . .],” see also example (5)). If the conditional clause is realized with the reduced form, the auxiliary verb is at the very first position while the lexical verb is at the very last position (*Hätte [AUX] Paul Spanisch gelernt, [. . .]*—“Had Paul learned Spanish, [. . .], see also example (6)). Interest areas smaller than a clause would thus not be meaningful for the analysis.

All experimental trials were further scanned manually, and drift correction was performed if necessary. As a result of the scanning, the data of two L2 participants were excluded from further analyses due to too much loss of gaze tracking and incomprehensible data. The tool “Get Reading Measure” provided by DataViewer was used to obtain total reading times for the areas of interest. The reading times for the second reading of each sentence were analyzed. According to the rationale of the study, the second reading of a sentence could either be the same realization of a given proposition as in the first reading or its alternation. Thus, the main experimental manipulation of interest was the same vs. changed condition (coded as factor “Change”), and the analyses focused on differences in total reading time between the same and changed conditions. For statistical analyses reported below, the software R was used (version 4.3.0; R Core Team, 2023). The data were analyzed with mixed-effects regression modeling using the R package *lme4* (version 1.1-33, Bates et al., 2015).

A maximal model structure was pursued (Barr et al., 2013). All models included fixed effects for all variables of interest and their interactions: Change (main manipulation, i.e., either the same or changed repetition of the sentence), Language

(L1 vs. L2 participants), and (Surface)Form (corresponding to the two surface realizations of the three respective alternations, e.g., active vs. passive).

As mentioned in the Methods section, we controlled for the interval between the first and second occurrence of a sentence (short interval with 2-4 inter-sentences or long interval with 12-14 inter-sentences). All models thus also included Distance as a fixed effect. However, all analyses showed that the factor Distance does not significantly affect the results, and, moreover, including it into the models does not change the pattern of results for the other factors of interest. We therefore refrain from reporting results for Distance in detail and report below only analyses without this factor.⁴

Beyond fixed effects and their interactions, the model structures included error terms for items and participants. Because models with the maximum structure for error terms (including random intercepts and random slopes for all predictor variables and their interactions as justified by the data structure) did not converge, we used R package *buildmer* (version 2.9, Voeten, 2023) to identify the maximal structure for error terms that was still capable of converging. Starting with an empty random effects structure, terms were added stepwise to the model until convergence of the model could no longer be achieved. The adding of terms was ordered based on the significance of the change in log-likelihood. The result of this procedure was considered the final (i.e., the maximal feasible) model. Final model structures are reported for each of the analyses below. Significance of fixed effects was evaluated using the R package *lmerTest* (version 3.1-3, Kuznetsova *et al.*, 2017), and ANOVA (type III) tables with Satterthwaite approximation for degrees of freedom are reported below. All bivalent categorical predictor variables were effect coded (i.e., as $-0.5/+0.5$).

Results

Manipulation of voice (active-passive)

Results are summarized in Table 3 and Figure 1. Analyses of total reading times revealed the main effects of Form ($F(1, 22.4) = 18.46, p < .001$), Language ($F(1, 116.7) = 35.71, p < .001$), and Change ($F(1, 115.8) = 37.4, p < .001$). There was also a significant interaction of Form:Language ($F(1, 2568.6) = 7.08, p = .008$). (see Table 4). Statistical analyses thus indicate that L2 participants generally read all sentences more slowly than native participants. Also, sentences in active voice were read faster than sentences in passive voice (that were also longer). The interaction shows that the difference between active and passive voice was larger in L2 than in L1. Importantly, it was also the case that sentences that were presented in a “changed” condition (active-passive, passive-active) were read more slowly than sentences in the “same” condition. The absence of any significant interaction between Language and Change indicates that both populations were equally sensitive to the experimental manipulation. In addition, the effect of Change was identical irrespective of the direction of change (active-passive vs. passive-active), as indicated by the absence of any interaction of Form:Change ($p = .908$), see Figure 1.

As mentioned in the Methods section, we additionally controlled our items for semantic bias. For 50% of sentences, the roles of agent—patient could plausibly be

Table 3. Results of active-passive alternation (L1 and L2). Total reading times: Means and (SD) in ms

	L1			L2			L1&L2
	Active	Passive	Mean	Active	Passive	Mean	Mean
Changed	1639.1 (809.4)	1816.8 (902.0)	1728.0	2305.3 (1073.1)	2587.5 (1225.1)	2446.4	2087.2
Same	1523.4 (716.8)	1681.3 (744.9)	1602.4	2074.7 (876.7)	2393.9 (1118.6)	2234.3	1918.3
<i>diff.</i>	115.7	135.5	125.6	230.6	193.6	212.1	168.85

Table 4. Mixed model ANOVA (type III) table for active-passive alternation

Effect	Sum Sq	NumDF	DenDF	F value	Pr(>F)
Change	17,157,718.214	1	115.8	37.44	<.001
Form	8,459,180.925	1	22.4	18.46	<.001
Language	16,364,891.015	1	116.7	35.71	<.001
Change:Form	6,068.329	1	2,569.1	0.01	.908
Change:Language	1,118,651.327	1	115.8	2.44	.121
Form:Language	3,244,698.124	1	2,568.6	7.08	.008
Change:Form:Language	135,086.697	1	2,568.7	0.29	.587

Note: Final model structure TOTAL_DURATION ~ Change×Form×Language + (1+Change|participant.id) + (1+Form|item.id).

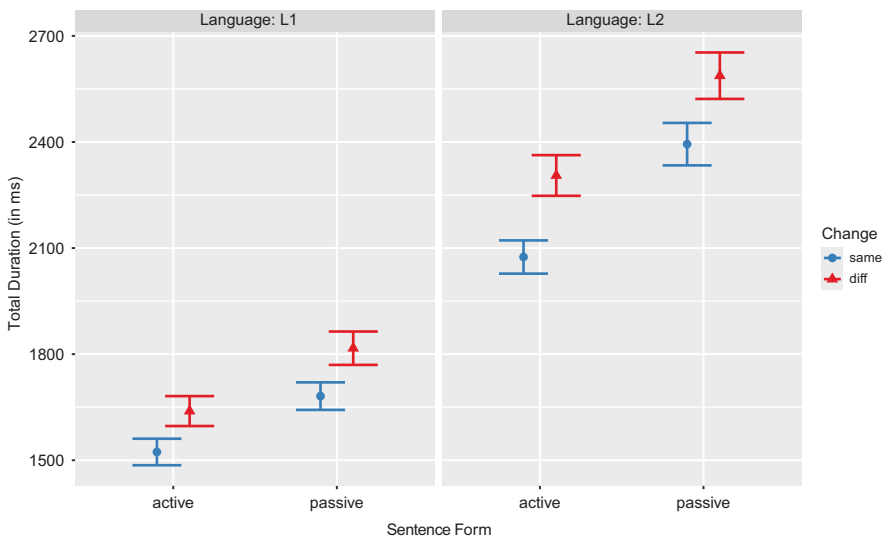
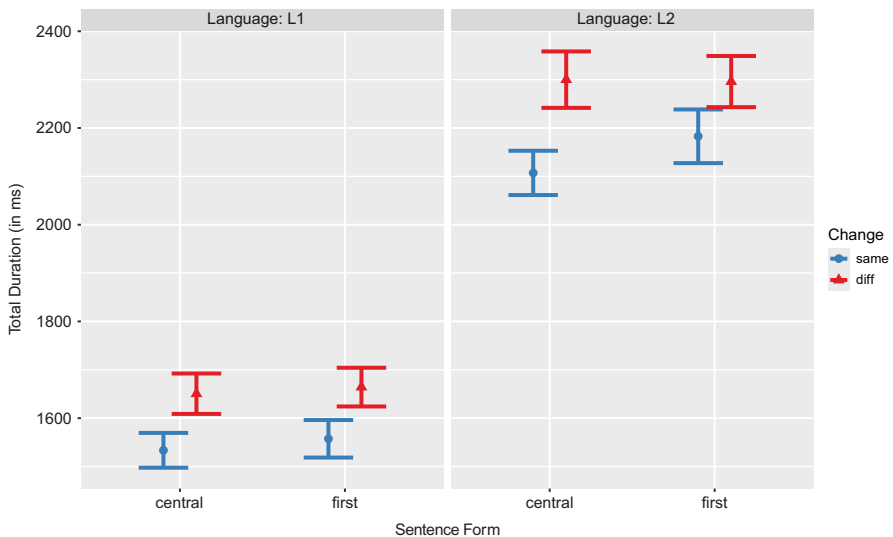


Figure 1. Results of voice alternation: mean total reading times for identical (“same”) or changed (“diff”) presentation of the sentence, grouped for language (L1 vs. L2) and sentence form (active vs. passive).

Table 5. Results of alternation of adverbial position (L1 and L2). Total reading times: Means and (SD) in ms

	L1			L2			L1&L2
	Front	Central	Mean	Front	Central	Mean	Mean
Changed	1664.2 (763.6)	1650.6 (798.1)	1657.4	2296.0 (987.0)	2300.0 (1086.6)	2298.0	1977.7
Same	1557.3 (742.1)	1533.5 (688.0)	1545.4	2182.8 (1034.9)	2107.2 (854.5)	2145.0	1845.2
<i>diff.</i>	106.9	117.1	112.0	113.2	192.8	153.0	132.5

**Figure 2.** Results of adverbial position alternation: mean total reading times for identical (“same”) or changed (“diff”) presentation of the sentence, grouped for language (L1 vs. L2) and sentence form (adverbial at the central vs. first position).

reversed (e.g., *Ein Mädchen rettet den Touristen aus einer peinlichen Situation.*—“The girl saves the tourist from an awkward situation.”), for 50% reversal of semantic roles was not plausible (e.g., *Die Tierschützer füttern die Koalas.*—“The animal rights activists feed the koalas.”). Subsequent analyses showed that semantic bias did not affect results (all $p > .188$, see [Appendix](#)).

Manipulation of adverbial position

Results for the manipulation of adverbial adjuncts position are summarized in [Table 5](#) and [Figure 2](#). Analyses revealed similar effects as for the active-passive manipulation (see [Table 6](#)). There were main effects of Language ($F(1, 129.1) = 29.34, p < .001$) and Change ($F(1, 2680.5) = 33.22, p < .001$) but no significant interaction of both factors ($F(1, 2681.4) = 0.54, p = .464$). Again, L1

Table 6. Mixed model ANOVA (type III) table for alternation of adverbial position

Effect	Sum Sq	NumDF	DenDF	F value	Pr(>F)
Change	12,665,760	1	2,680.5	33.22	<.001
Form	636,532.9	1	2,680.8	1.67	.196
Language	11,187,049.6	1	129.1	29.34	<.001
Change:Form	279,776.1	1	2,680.6	0.73	.392
Change:Language	204,241.3	1	2,681.4	0.54	.464
Form:Language	20,408.3	1	2,680.8	0.05	.817
Change:Form:Language	241,362.8	1	2,681.4	0.63	.426

Note: Final model structure $\log(\text{TOTAL_DURATION}+1) \sim \text{Change} \times \text{Form} \times \text{Language} + (1|\text{participant.id}) + (1+ \text{Language}|\text{item.id})$.

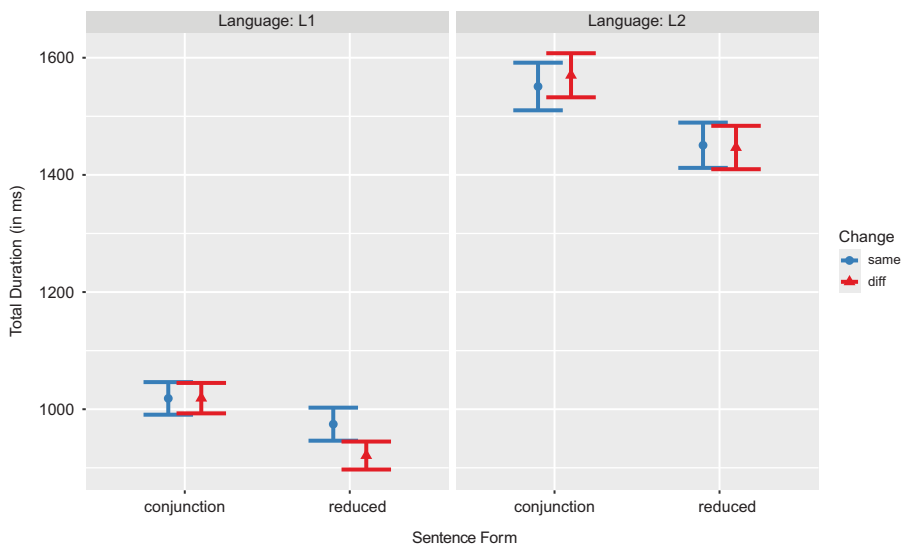
participants were faster in reading all sentences. Importantly, it was again the case that sentences in the “changed” condition were fixated longer than identical sentences in the “same” condition. The absence of any significant interaction shows that both populations were equally sensitive also to this type of manipulation. In contrast to the active-passive manipulation, this time the factor Form (adverbial in first vs. central position) did not influence reading times significantly ($F(1, 2680.8) = 1.67, p = .196$). As mentioned above, we also controlled the syntactic type of the adverbials (12 sentences with single-word adverbs, 12 sentences with an adverbial PP). We therefore run an additional analysis that included the factor “Syntactic Type of Adverb.” Results showed a significant main effect ($F(1, 22.1) = 10.42, p = <.001$) indicating that sentences with single-word adverbial were read faster than sentences with adverbial PP. There was also a significant interaction of Syntactic Type : Language ($F(1, 22.0) = 6.50, p = .018$) showing that the difference between single-word adverbs and adverbial PPs was larger for L2 than for L1. Crucially, however, the lack of any significant interaction involving Syntactic Type and Change indicated that the critical manipulation was not influenced by the syntactic form of the adverbial (for details see [Appendix](#)).

Manipulation of conditional clauses

As indicated by the numerical results (see [Table 7](#) and [Figure 3](#)), statistical analyses revealed that the pattern of results for the conditional clause manipulation was different from the two other investigated alternations. This time, there were only main effects for Language ($F(1, 127.4) = 45.95, p < .001$) and Form ($F(1, 2680.6) = 7.02, p = .008$). But neither was there any main effect of Change ($F(1, 2680.5) = 0.001, p = .983$) nor did this factor interact with any other fixed effects (all $p > .580$). Thus, no indications can be observed that manipulation of the conditional clauses would affect the reading behavior of the participants. The main effect for Form shows that conditional clauses introduced with *wenn* (“if”) took longer to read than reduced clauses, which is most likely due to the fact that they were also longer. The fact that the factor Form was not involved in any interaction

Table 7. Results of alternation of conditional clauses (L1 and L2). Total reading times: Means and (SD) in ms

	L1			L2			L1&L2
	Conjunct	Reduced	Mean	Conjunct	Reduced	Mean	Mean
Changed	1018.9 (496.9)	921.0 (456.6)	969.7	1570.1 (725.5)	1446.7 (713.1)	1508.4	1239.2
Same	1018.5 (532.3)	974.5 (537.8)	996.5	1550.9 (783.1)	1450.6 (744.3)	1500.8	1248.6
<i>diff.</i>	0.4	-53.5	-26.8	19.2	-3.9	7.6	-9.4

**Figure 3.** Results of conditional clause alternation: mean total times for identical (“same”) or changed (“diff”) presentation of the sentence, grouped for language (L1 vs. L2) and sentence form (conditional clause starting with conjunction or reduced (without conjunction)).

shows that its influence on reading times was the same in all conditions (see also Table 8).

Summary of results and discussion

The analyses revealed very clear patterns. First, both L1 and L2 participants manifested the same sensitivity to the crucial manipulation. Throughout the experiment, reading behavior in both populations was similarly affected when the second presentation of a proposition was either changed or identical compared to its first presentation (no interactions with predictor Language in any of the analyses).

Table 8. Mixed model ANOVA (type III) table for alternation of conditional clauses

Effect	Sum Sq	NumDF	DenDF	F value	Pr(>F)
Change	231.801	1	2,680.5	0.00	.983
Form	3,527,123.940	1	2,680.6	7.02	.008
Language	23,101,198.904	1	127.4	45.95	<.001
Change:Form	60,201.853	1	2,680.5	0.12	.729
Change:Language	2,352.687	1	2,681.3	0.00	.945
Form:Language	153,378.785	1	2,680.6	0.31	.581
Change:Form:Language	112,565.633	1	2,681.3	0.22	.636

Note: Final model structure $\log(\text{TOTAL_DURATION}+1) \sim \text{Change} \times \text{Form} \times \text{Language} + (1|\text{participant.id}) + (1+\text{Language}|\text{item.id})$.

Beyond generally slower reading times for L2 participants, we do not find any indications for L1-L2 differences.⁵

Second, however, a different pattern of results is observed for the three types of alternation. In the Voice and Adverbial Position manipulations, participants reacted with longer reading times for changed sentences than for identical repetitions. For the Conditional alternation, however, neither L1 nor L2 participants showed any indications of different reading behaviors for the changed versus identical repetitions. Thus, it seems that there are characteristics inherent to the three types of alternation that have an impact on their retention. Moreover, the results showed no evidence that differences in the distance between the first and second encounter with the sentence affected reading behavior (see [Appendix](#) for additional statistical analyses).

General discussion

The experiments in our study delivered several results that enhance our knowledge about the information that is retained in memory during reading and the factors that influence it. Unlike most previous studies, we did not focus on comparing retention of meaning versus form. Instead, we zoomed in on the language features that are considered formal and investigated whether differences between them affect their retention. We chose three alternations in German which show a graded degree of grammatical and information structural change between their two realizations but leave their propositional meaning unaffected: The active/passive alternation affects the grammatical, the argument-structural, and the information structural level, the adjunct position alternation only the information structure level, and the conditional clause alternation only the grammatical level. This differentiated approach to the surface linguistic information turned out to be very productive. Before we discuss it in detail, we must first address the results that are not directly related to the crucial manipulation.

First, we obtained a comparably similar pattern of results for L1 speakers and advanced L2 learners. In this respect, our study differs from both Bordag et al. (2021) and Sampaio and Konopka (2013) which both observed advantages for L2

learners in retention of surface linguistic information. However, there were important differences in the design of the studies. In the experiments by Bordag and colleagues, the critical sentences with the active/passive alternation were parts of coherent text passages. In contrast, we presented the critical sentences in isolation. Previous research has shown that memory for surface linguistic information is negatively affected by semantic integration processes during the construction of mental text models and that verbatim memory for isolated sentences is thus better than the memory for sentences presented in coherent contexts (e.g., Anderson & Bower, 1973; Peterson & McIntyre 1973; de Villiers, 1974). The fact that we observed retention effects for the active/passive alternation in the current L1 experiment while they were not observed in the text reading experiment by Bordag *et al.* (2021) is thus in agreement with previous research and highlights the role of coherent texts in the decay of verbatim memory. In the study by Sampaio and Konopka (2013), the L2 advantage for the surface form retention in synonym substitutions (and here also restricted only to one condition) was observed for isolated sentences. However, the authors employed a cue recall method that might require retention that goes beyond implicit memory. Therefore, we assume that the fact that we were able to observe more prominent surface form retention effects in L2 and effects of the same order of magnitude in L1 might be due to the employment of the eye-tracking methodology which is more sensitive to implicit memory effects than recall procedures. Overall, our results do not contradict the assumption of Sampaio and Konopka (2013) and Bordag and colleagues (Bordag *et al.*, 2021) for better retention of surface form in L2 than in L1:

The difference between the results of their studies and ours could be explained by the L1 effects being detectable only under optimal testing conditions, whereas the more robust L2 effects can be observed also under less favorable conditions.

Second, our results also differ from earlier studies with respect to the time span for which effects of surface-level retention are observed. As discussed in the introduction, studies from the 1960s and 1970s had led to the long-lasting assumption that we store only propositional content of sentences and texts, whereas the surface level fades off as soon as we complete reading a sentence. Although, for example, Sachs (1967) and Anderson (1974) observed that the memory for non-semantic manipulations is available basically only directly after the critical sentence was presented, our results show that it can last in unchanged strength for 2-14 intervening sentences. At the same time, it must be acknowledged that inconsistent indications that memory for non-propositional contents might last longer than usually observed were reported already in the early studies. For example, Sachs (1974) observed retention effects for the active/passive alternation in her reading experiment, although she did not observe them either in her 1967 study or in her 1974 listening experiment. For other alternations, typically subsumed by the authors under “form” manipulations in contrast to the “meaning” manipulations (see Introduction), no retention effects were usually observed except immediately after the critical sentence presentation.

The main finding of our study, however, is that the type of alternations affects retention. As mentioned in the introduction, the argument that “some wording changes which appear to be entirely formal can have semantic qualities which might affect memorability in other situations” (Soli & Balch, 1976, p. 676) had been

presented already very early, but with no subsequent systematic research. Our current study represents a step forward precisely in this direction.

Our results show that while alternations that affect information structure are associated with better retention of surface linguistic information (active/passive alternation and adverbial position alternation), no evidence for such retention could be found for surface manipulations that are purely formal (conditional clause alternation). In other words, grammatical transformations that do not have any communicative function do not seem to trigger better retention. This is most prominently demonstrated by the absence of any retention effects in the conditional clause alternation. As soon as there is no functional difference between alternation variants, we do not observe any differences in reading behavior.

This point is further illustrated by the fact that there is no difference between the retention effect of the active/passive and adverbial position alternations. If the grammatical change per se contributed to the retention effect, we would expect an interaction with a larger effect for the active/passive alternation (which involves the strongest grammatical changes, including different alignments of semantic roles, abstract case, and grammatical functions) than for the adverbial position alternation (which involve no changes beyond linearization and information structures). This was not the case.⁶ We therefore conclude that it was the linguistic property shared by the two manipulations, i.e., changes in information structure, that caused the effect. It should be noted, however, that information structural changes in the active/passive manipulation are always confounded with changes in the alignment of arguments in German. In this regard, disentangling these two levels and assessing possible differences in their effects on verbatim retention (which could lie not only in the strength of the effect but also in its durability) would be a valuable topic for future research.

It is particularly interesting that information structure plays such a strong role even in L2. Throughout our study, L2 participants show a similar reading pattern with respect to the same/changed manipulations as native speakers. This is even more surprising in the case of the conditional clause alternation. This type of alternation involves rather large surface-level changes including changes in the position of the finite verb (i.e., clause-initial vs. clause-final). That neither L1 nor L2 participants noticed such changes suggests that these purely formal grammatical surface form properties have not been retained at all.

This finding is the more interesting as it shows that rather large changes in word order and/or morphological changes do not seem to be the most critical factors that trigger better verbatim retention, but that it is indeed the information structure. In contrast to the conditional clause alternation, the adverbial position alternation, which does not involve any morphological changes, but which affects the information structure, led to robust retention effects in L1 and L2.

However, with respect to the lack of L1-L2 differences in the present study, it must be noted that the learners' L1s (mostly Czech and other Slavonic languages) have an even more free word order and that information structure is also encoded via word order in them. A strong reliance on information structure in L2 German could thus be a result of transfer from participants' native languages. It would be interesting to investigate learners with different combinations of L1 and L2 that are more distinct with respect to how fixed word order is and how information

structure is encoded (e.g., including languages in which word order primarily encodes grammatical functions rather than a functional perspective, e.g., Chinese or to a lesser degree also English).

In terms of the debate about the role of memory in language acquisition and processing, our results suggest that different linguistic features trigger different degrees of verbatim memory for the sentence surface structure. Crucially, this verbatim retention goes beyond the scope of working memory (up to 14 intervening sentences in our study) and beyond memory for propositional content. Further research is needed to specify the extent to which the retention of such properties contributes to both language acquisition and processing.

To conclude, our research has revealed two main novel findings. First, information structure is another type of information that is retained during reading in addition to the lexical and propositional meaning. It may not be available to phenomenal memory awareness, but it seems to be retained in the implicit memory. Interestingly, advanced L2 German learners with Czech as their L1 seem to be equally sensitive to changes in information structure encoding at the surface level (word order) as the native speakers. It remains a question for further research what role learners' L1 plays in this context. Are differences in the relevance and in the encoding of information structure between L1 and L2 reflected in differences in verbatim retention?

Second, our findings call for a revision of the concept of verbatim memory. Clearly, readers are able to retain surface linguistic information, but not all types of it. Purely formal surface information, which is arbitrary in the sense that it bears no relation to the meaning of the sentence, its information structure, or its communicative function, does not seem to be retained. More alternations and structures need to be tested to verify this claim and to explore the role of pragmatics in information retention in general.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0142716424000122>.

Replication Package. The materials, the data, and the analysis codes are available on Open Science Framework (OSF): <https://osf.io/kp7jg/>.

Acknowledgments. The authors wish to express their gratitude to the LABLES lab and the Faculty of Arts at Charles University in Prague for providing the necessary facilities to conduct the experiments abroad. Special thanks are extended to Lisa-Naomi Meller for her invaluable assistance in preparing the experimental materials, testing the participants, and editing the manuscript.

Funding statement. This work was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (grant number BO 3615/9-1 to Denisa Bordag).

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Notes

1 We use the word “structure” in a more general meaning that also includes, e.g., information structure realizations like word order.

2 A partly similar procedure has been employed by Sanford and colleagues (Sanford *et al.*, 2006; Sanford *et al.* 2009). However, there are significant distinctions, both in terms of methodology (e.g., different tasks)

and the scope or research question. Sanford and colleagues use a word-change paradigm to explore how linguistic properties of focus, subordination, focalization, and clefting can control the depth of processing, leading either to more fine-grained or underspecified representations. Our study centers on the retention of the surface linguistic structures per se.

3 The first step applies thresholds for durations of 80 ms and a 0.5° distance, the second and third steps have threshold durations of 50ms and 140ms, respectively, and a distance threshold of 1.25°. The fourth step sets minimum durations to 140ms and maximum durations to 800ms.

4 Detailed analyses which also include this factor can, however, be found in the [Appendix](#).

5 It should be noted that all L2 participants were on a relatively high, (pre-)advanced language level (B2-C1). In their proficiency assessments, they usually scored at the B2 level in one test and C1 in another. They thus formed a relatively homogenous group in terms of proficiency. Therefore, further analysis considering different proficiency levels is not possible for the present data.

6 This was statistically evaluated by computing a joint model of both alternation types (final model structure: TOTAL_DURATION ~ Alternation*Change*Language + (1|participant.id) + (1|item.id)). Results showed no main effect for factor Alternation type ($F(1, 46.0) = 1.36, p = .250$). More importantly, there was no significant interaction of Alternation type with Change ($F(1, 5534.2) = 0.96, p = .326$).

References

- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum.
- 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. [https://doi.org/10.1016/S0022-5371\(74\)80039-3](https://doi.org/10.1016/S0022-5371(74)80039-3)
- Anderson, J. R., & Bower, G. H. (1973). *Human associative memory* (pp. xiv, 524). V. H. Winston & Sons.
- Audring, J. (2022). Advances in morphological theory: Construction morphology and relational morphology. *Annual Review of Linguistics*, *8*(1), 39–58. <https://doi.org/10.1146/annurev-linguistics-031120-115118>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, *127*(3), 287–306. <https://doi.org/10.1016/j.cognition.2013.02.005>
- Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458. <https://doi.org/10.1016/j.cognition.2006.07.003>
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*, 177–192. <https://doi.org/10.1037/0096-3445.129.2.177>
- Bordag, D., Opitz, A., Polter, M., & Meng, M. (2021). Non-native readers are more sensitive to changes in surface linguistic information than native readers. *Bilingualism: Language and Cognition*, *24*(4), 599–611. <https://doi.org/10.1017/S1366728920000772>
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*(2), 193–209. [https://doi.org/10.1016/0010-0285\(72\)90003-5](https://doi.org/10.1016/0010-0285(72)90003-5)
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. John Benjamins Publishing.
- Chomsky, N. (2014). *Aspects of the Theory of Syntax* (no. 11). MIT press.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. <https://doi.org/10.1017/S0142716406060024>
- Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. <https://doi.org/10.1017/S0272263117000250>

- Craik, F. I. M., Rose, N. S., & Gopie, N.** (2015). Recognition without awareness: Encoding and retrieval factors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **41**, 1271–1281. <https://doi.org/10.1037/xlm0000137>
- de Villiers, P. A.** (1974). Imagery and theme in recall of connected discourse. *Journal of Experimental Psychology*, **103**, 263–268. <https://doi.org/10.1037/h0037608>
- Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K.** (2008). The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 1552–1560. <https://doi.org/10.1037/a0013017>
- Ellis, J.** (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 1–22). Lawrence Erlbaum Associates Publishers.
- Garnham, A., & Oakhill, J.** (1996). The mental models theory of language comprehension. In B. K. Britton, & A. C. Graesser (Eds.), *Models of understanding text* (pp. 313–339). Lawrence Erlbaum Associates, Inc.
- Gernsbacher, M. A.** (1985). Surface information loss in comprehension. *Cognitive Psychology*, **17**(3), 324–363. [https://doi.org/10.1016/0010-0285\(85\)90012-X](https://doi.org/10.1016/0010-0285(85)90012-X)
- Glenberg, A. M., Meyer, M., & Lindem, K.** (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, **26**(1), 69–83. [https://doi.org/10.1016/0749-596X\(87\)90063-5](https://doi.org/10.1016/0749-596X(87)90063-5)
- Goldberg, A. E.** (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Graesser, A., & Mandler, G.** (1975). Recognition memory for the meaning and surface structure of sentences. *Journal of Experimental Psychology: Human Learning and Memory*, **1**, 238–248. <https://doi.org/10.1037/0278-7393.1.3.238>
- Gurevich, O., Johnson, M. A., & Goldberg, A. E.** (2010). Incidental verbatim memory for language. *Language and Cognition*, **2**(1), 45–78. <https://doi.org/10.1515/langcog.2010.003>
- Hutmacher, F., & Kuhbandner, C.** (2018). Long-term memory for haptically explored objects: fidelity, durability, incidental encoding, and cross-modal transfer. *Psychological Science*, **29**(12), 2031–2038. <https://doi.org/10.1177/0956797618803644>
- Hutmacher, F., & Kuhbandner, C.** (2020). Detailed long-term memory for unattended, irrelevant, and incidentally encoded auditory information. *Journal of Experimental Psychology: General*, **149**, 222–229. <https://doi.org/10.1037/xge0000650>
- Jarvella, R. J.** (1971). Syntactic processing of connected speech. *Journal of Verbal Learning & Verbal Behavior*, **10**, 409–416. [https://doi.org/10.1016/S0022-5371\(71\)80040-3](https://doi.org/10.1016/S0022-5371(71)80040-3)
- Jarvella, R. J.** (1979). Immediate memory and discourse processing. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (vol. 13, pp. 379–421). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60088-2](https://doi.org/10.1016/S0079-7421(08)60088-2)
- Johnson-Laird, P. N., & Stevenson, R.** (1970). Memory for syntax. *Nature*, **227**(5256), Article 5256. <https://doi.org/10.1038/227412a0>
- Kintsch, W., & Bates, E.** (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 150–159. <https://doi.org/10.1037/0278-7393.3.2.150>
- Kintsch, W.** (1986). Learning from text. *Cognition and Instruction*, **3**(2), 87–108.
- Kintsch, W.** (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, **95**(2), 163.
- Kintsch, W.** (1991). The role of knowledge in discourse comprehension: A construction-integration model. *Advances in Psychology*, **79**, 107–153.
- Kintsch, W.** (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kuhbandner, C.** (2020). Long-lasting verbatim memory for the words of books after a single reading without any learning intention. *Frontiers in Psychology*, **11**, 541538. <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01780>
- Kuhbandner, C., Rosas-Corona, E. A., & Spachtholz, P.** (2017). High-fidelity visual long-term memory within an unattended blink of an eye. *Frontiers in Psychology*, **8**, 296875. <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01859>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>

- Langacker, R. W.** (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics* (pp. 127). John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.50.06lan>
- Peterson, R. G., & McIntyre, C. W.** (1973). The influence of semantic "relatedness" on linguistic integration and retention. *The American Journal of Psychology*, *86*, 697–706. <https://doi.org/10.2307/1422077>
- R Core Team.** (2023). *R: A Language and Environment for Statistical Computing* (4.2.2). R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reyna, V. F., & Kiernan, B.** (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*, 178–191. <https://doi.org/10.1037/0012-1649.30.2.178>
- Sachs, J. S.** (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*, 437–442. <https://doi.org/10.3758/BF03208784>
- Sachs, J. S.** (1974). Memory in reading and listening to discourse. *Memory & Cognition*, *2*(1), 95–100. <https://doi.org/10.3758/BF03197498>
- Sampaio, C., & Konopka, A. E.** (2013). Memory for non-native language: The role of lexical processing in the retention of surface form. *Memory*, *21*(4), 537–544. <https://doi.org/10.1080/09658211.2012.746371>
- Sanford, A. J., Sanford, A. J., Molle, J., & Emmott, C.** (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*(2), 109–130.
- Sanford, A. J. S., Price, J., & Sanford, A. J.** (2009). Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition*, *37*, 880–888. <https://doi.org/10.3758/MC.37.6.880>
- Soli, S. D., & Balch, W. R.** (1976). Performance biases and recognition memory for semantic and formal changes in connected discourse. *Memory & Cognition*, *4*(6), 673–676. <https://doi.org/10.3758/BF03213233>
- SR Research.** (2019). *DataViewer* (4.1.211). SR Research Ltd.
- SR Research.** (2020). *Experiment Builder* (2.3.38). SR Research Ltd.
- Staub, A., Dodge, S., & Cohen, A. L.** (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review*, *26*(1), 340–346. <https://doi.org/10.3758/s13423-018-1492-z>
- Tomasello, M.** (2003). *Constructing a language: A usage-based theory of language acquisition* (pp. viii, 388). Harvard University Press.
- Ullman, M. T.** (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 953–968). Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00076-6>
- Von Eckardt, B., & Potter, M. C.** (1985). Clauses and the semantic representation of words. *Memory & Cognition*, *13*(4), 371–376. <https://doi.org/10.3758/BF03202505>
- Voss, J. L., Baym, C. L., & Paller, K. A.** (2008). Accurate forced-choice recognition without awareness of memory retrieval. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *15*(6), 454–459. <https://doi.org/10.1101/lm.971208>
- Voeten, C.** (2023). *builder: Stepwise elimination and term reordering for mixed-effects regression*. R package version 2.9. <https://CRAN.R-project.org/package=builder>.

Cite this article: Opitz, A., Bordag, D., & Furgoni, A. (2024). The role of linguistic factors in the retention of verbatim information in reading: An eye-tracking study on L1 and L2 German. *Applied Psycholinguistics*. <https://doi.org/10.1017/S0142716424000122>