Bilingualism: Language and Cognition

cambridge.org/bil

Non-native Readers Are More Sensitive to Changes in Surface Linguistic Information than Native Readers*

Denisa Bordag[1,2], Andreas Opitz[1], Max Polter[1] and Michael Meng[3]

[1]Leipzig University, Leipzig, Germany; [2]University of Haifa, Haifa, Israel and [3]Hochschule Merseburg – University of Applied Sciences, Merseburg, Germany

Research Article

Cite this article: 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**, 599–611. https://doi.org/10.1017/S1366728920000772

Received: 25 August 2019
Revised: 12 November 2020
Accepted: 16 November 2020
First published online: 17 February 2021

Keywords:
reading; SLA; surface linguistic information; memory; eye-tracking

Address for correspondence:
Denisa Bordag, Universität Leipzig, Beethovenstr. 15, 04107 Leipzig, Germany.
E-mail: denisav@uni-leipzig.de

Abstract

In the present study we challenge the generally accepted view based primarily on L1 data that surface linguistic information decays rapidly during reading and that only propositional information is retained in memory. In two eye-tracking experiments, we show that both L1 and L2 adult readers retain verbatim information of a text. In particular, the reading behaviour of L2 German learners revealed that they were sensitive to both lexical (synonyms) and syntactic (active/passive alternation) substitutions during a second reading of the texts, while L1 exhibited only reduced sensitivity to the lexical substitutions. The results deliver an important piece of evidence that complies with several current processing (e.g., Shallow Structure Hypothesis), acquisition (Declarative/Procedural Model) and cognitive (e.g., Fuzzy Trace Theory) approaches and adds a new dimension to their empirical and theoretical basis.

Introduction

The long history of research on memory for gist vs. surface linguistic information seems to converge on the assumption that linguistic information is not retained verbatim, but converted to conceptual form, which is then stored in long-term memory. Unless required by an additional task (e.g., an announced memory test), the low-level, surface information is assumed to be available immediately after the sentence processing, but to decay as soon as hierarchically superior structures (e.g., propositional representation) are built (e.g., Anderson, 1974; Caplan, 1972; Gernsbacher, 1985; Jarvella, 1971; Jarvella & Herman, 1972; Johnson-Laird & Stevenson, 1970; Sachs, 1967, 1974; for more recent studies, e.g., Holtgraves, 2008; Just & Carpenter, 1992; Lombardi & Potter, 1992; Potter & Lombardi, 1998; Rummer & Engelkamp, 2001; Rummer, Schweppe & Martin, 2013).

At the same time, usage-based theories of grammar acquisition assume that grammatical knowledge is derived from a database of memorised chunks (Bybee, 1985; Ellis, 1996; Goldberg, 2006; Langacker, 1988; Tomasello, 2003). Within this theoretical framework, it is assumed that sequences of words, which are stored verbatim in memory, are used to abstract regularities and that mental grammar develops through gradual assembling of knowledge about distributional and semantic-distributional relationships between words.

The two above-mentioned lines of research, which reach conclusions that might seem difficult to reconcile at the first sight, differ in populations on which each of them focuses. While research on memory for the two types of information (gist vs. verbatim) typically addresses performance of adult native speakers, usage-based theories focus on language acquisition, i.e., on children and non-native speakers. In the present study, we want to test the hypothesis that adult L2 learners outperform L1 speakers in memory for surface linguistic information. We assume that this ability would subserve the assembly of the chunk database necessary for the development of implicit grammatical knowledge. In particular, we want to compare the degree of retention of surface lexical (synonyms) and structural (active/passive alternation) information during reading in L1 and L2 German.

Since we explore verbatim memory in reading, the research question can also be formulated within the context of text comprehension. During reading, readers construct a mental text model, a higher-level mental representation of affairs conveyed by a text. It contains both propositional meaning and additional information contributed by the reader (inferences, world knowledge) (Field, 2004, p. 176). The model is dynamic, both because more information from the text is gradually integrated into it, and some of the earlier information fades from the memory (Bernhardt, 2011; Kintsch, 1988, 1998; Stanovich, 2000; van Dijk & Kintsch, 1983). It is considered 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, Meyer & Lindem, 1987, p. 69). The research on text models thus complies with the literature on memory in that it assumes that verbatim information is not replicated in the mental representation (Garnham, 1996; Johnson-Laird, 1977). However, also in this

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

area comparisons between L1 and L2 speakers with respect to retention of verbatim information are missing. In our study, we investigate to what degree surface linguistic information becomes a part of the mental text representation of these two populations and whether there is a difference in the degree in which L1 or L2 speakers rely on this type of information during reading.

Despite the general consensus that verbatim information is retained in memory only very briefly, several recent studies indicate that surface linguistic information does not fade from memory as fast and completely as previously believed. Gurevich, Johnson, and Goldberg (2010) carefully reviewed the previous literature on the topic and concluded that "various factors have conspired to downplay the importance of memory for language in past research" (Gurevich et al., 2010, p. 49). The emphasis of the studies had been on memory for content or gist which always trumps the verbatim memory; the latter, however, does not seem to be as uninvolved in naturalistic contexts as claimed. As an example, Gurevich et al. (2010) mention a study by Kintsch and Bates (1977) which showed that participants recognise sentences they had heard in a regular university lecture significantly better than their paraphrases.

In their own recognition and recall experiments, Gurevich et al. (2010) used texts and illustrations from children's storybooks as stimuli. Participants first listened to the recorded texts while corresponding pictures were presented to them. In the recognition task, written clauses were displayed on the screen and subjects were asked to answer 'yes' or 'no' to whether or not they had heard that exact clause in the story. In the recall task, participants were asked to retell the story from the pictures. The story had two versions; the second being a paraphrase of the first using different formulations to express the same content. The research question was whether participants who had heard a particular formulation were more likely to recognise/produce that formulation than subjects who had heard a paraphrase of the same content. This turned out to be the case. Gurevich and colleagues could demonstrate 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 (cf. Gibbs, 1981; Kintsch & Bates, 1977). Their results challenge the previous finding that downplays the significance of verbatim memory.

Interestingly – and perhaps because these L1 results have remained unchallenged for so many years – there is to our knowledge only one study that has directly addressed the retention of surface information in L2. In a cued sentence recall procedure, Sampaio and Konopka (2013) showed that while L1 and L2 speakers recall equally well the gist of a sentence like "*The bullet STRUCK the bull's eye*", the L2 speakers recall better the verbatim phrasing of sentences with non-preferred lexical items (e.g., *STRUCK* vs. *HIT*) and are thus more sensitive to synonymous lexical substitutions in such sentences. According to the authors, the L1 vs. L2 difference arises due to the nature of the lexical-semantic pathways that support processing in L1 and L2. They interpret their finding within the Revised Hierarchical Model (RHM) of Kroll and Stewart (1990, 1994) and claim that the engagement of the L2-L1 route (as opposed to the route that links the words directly with the conceptual store) leads L2 speakers to more intensive lexical processing, which benefits the retention of the L2 surface form (i.e., verbatim memory). However, due to the design of the study, it does not address the results of the L1 participants in the same way: it shows that L2 recall of verbatim phrasing is better than recall in L1, but due to a missing baseline for L1, it does not answer the question, whether also recall within L1 is above chance, or not (i.e., whether also L1 participants retain surface forms in L1).

Other recent studies also support the assumption that L2 learners engage more intensely with low-level text information than native speakers. As an example, in a series of studies on incidental vocabulary acquisition, Bordag and colleagues (Bordag, Kirschenbaum, Tschirner & Opitz, 2015; Bordag, Kirschenbaum, Opitz & Tschirner, 2014; Bordag, Kirschenbaum, Opitz, Rogahn & Tschirner, 2016a, 2016b) revealed that syntactically complex texts trigger a shift of attention from the global aim of text comprehension to individual lexical items and their properties in L2 learners, but not in L1 speakers. In contexts that were demanding (i.e., syntactically complex) for L2 learners, these acquired better both the meaning and the grammatical properties of unknown words during reading compared to L1 speakers and compared to the condition when they read syntactically undemanding texts. Such findings indicate that the focus of less proficient readers/speakers might be more directed at lower-level information than that of adult L1 speakers and that this behaviour does not have to be limited to individual lexical items as suggested by Sampaio and Konopka (2013).

In addition to the recent research studies mentioned above, existing cognitive, acquisition, and processing theories make assumptions that support the hypothesis of verbatim retention in L2 readers, too.

The Fuzzy Trace Theory (FTT), a cognitive theory originally proposed by Ch. Brainerd and V. F. Reyna in the 1990s (Reyna & Brainerd, 1995), is employed especially in research on decision making and false memories, but it makes assumptions that are very relevant for the present purposes.[1] As a dual process theory, it assumes two types of representation of past events: meaning-based gist representations, which support fuzzy (yet advanced) intuition, and superficial verbatim representations of information, which support precise analysis (Reyna, 2012, p. 332). Both types of representations are encoded in parallel, can be retrieved independently from each other and have different forgetting rates (with verbatim traces becoming inaccessible at a faster rate than gist traces). Studies have shown that the dominance of one of the two representation types depends on a number of factors, including expertise. Reyna (2012) reports several studies (e.g., Reyna, 2004; Reyna & Lloyd, 2006; Reyna, Lloyd & Brainerd, 2003) showing that experts rely on simpler gist-based representations, use less information, and process it less precisely to make decisions. At the same time, they have better discrimination than novices regarding predictions of outcomes in their area of expertise (Lloyd & Reyna, 2009; Reyna & Lloyd, 2006). In the area of second language acquisition, a study of Arndt and Beato (2017) not only shows that non-native speakers are less likely to produce false memories for critical words in Deese-Roediger-McDermott paradigm lists than native speakers, but also that less proficient L2 speakers produce fewer false memories than highly proficient L2 speakers (cf. equivalent results for younger vs. older children, e.g., Metzger, Warren, Shelton, Price, Reed & Williams, 2008). Such results support the assumption of FTT that older children, adults, and proficient L2 speakers are more susceptible to false memories because their encoding is more based on gist.

[1]Delaney, Verkoeijen, and Spirgel (2010) point out explicitly that Kintsch's (Mannes & Kintsch, 1987; Kintsch, 1994) assumptions about mental text models are consistent with FTT.

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

As suggested by studies of Bordag and colleagues (Bordag et al., 2014; 2015; 2016a, 2016b), differences in L1 and L2 reading and the related success in incidental language acquisition may crucially depend on the degree of difficulty that the reading process poses for readers or, in other words, on their novice vs. expert status to the task. Consequently and in accordance with the hypothesis that we test, we might expect that not fully proficient L2 learners might deal with L2 reading as novices and rely more on verbatim-based analysis than proficient adult speakers. Interestingly, this assumption is also supported by a study by Keysar, Hayakawa, and Sun Gyu (2012) who have shown that gain vs. loss framing effects, which are known to arise due to gist-based intuition, disappear when decisions are made in an L2, thus supporting the notion that L2 text processing may depend more on verbatim-based analysis.[2]

Among acquisition approaches, it is especially the Declarative/Procedural model (DPM) (Ullman, 2004) and its claims about the differences in L1 and L2 processing that are complementary to our hypothesis. According to DPM, procedural memory that subserves the rule-governed sequential and hierarchical computation of complex linguistic structures (i.e., crucial aspects of mental grammar) declines with age (Ullman, 2004). Its decline is accompanied by improvement of declarative memory, which underlies the mental lexicon that contains word-specific knowledge and memorised complex forms (Ullman, 2004, 2016; Ullman & Lovelett, 2018). Along with usage-based theories, the DPM assumes that children initially learn both idiosyncratic and complex forms in declarative memory, while the procedural system gradually acquires the grammatical knowledge underlying rule-governed combination. Due to the decline of the procedural system, L2 learners tend to rely more strongly on declarative memory, even for functions that depend on the procedural system in L1. Reliance on verbatim storage in declarative memory and on associative generalisations over them could thus compensate for the limited L2 ability to acquire and process grammar procedurally.

Along the lines sketched above, the overrepresentation of verbatim information in L2 mental text models also complies with the Shallow Structure Hypothesis (SSH) by Clahsen and Felser (2006, 2017). Clahsen and Felser (2017) explicitly classify SSH as a multiple-pathways model that assumes at least two processing routes that operate in parallel similar to the two types of representation encoding in FTT. One of the routes involves creating detailed syntactic structures; the other one is syntactically shallower. Shallow processing (characterized by deficiencies in hierarchical syntactic organisation) is accompanied by reliance (or overreliance) on "semantic, pragmatic, probabilistic, or **surface-level information**" (Clahsen & Felser, 2017, p. 694, bold by the authors). Similar to FTT, SSH also assumes that the dominance of one of the two routes may depend on various factors such as task or language proficiency. According to the hypothesis, L2 learners dispose of the same processing architecture and mental-processing mechanisms as L1 speakers, but they have "problems building or manipulating abstract syntactic representations in real time" (Clahsen & Felser, 2017, p. 2) and underuse syntactic information in online processing. Consequently, they are more likely to rely, e.g., on surface linguistic information than native speakers, which is in accordance with our hypothesis.

[2]The authors attribute the results to greater emotional and cognitive distance in L2 which is not in contradiction to FTT, rather contrary: The emotional/cognitive distance can be related to increased surface level processing.

We surveyed three theoretical approaches to show that there are cognitive, acquisition-related, and processing grounds, based on which we might expect that despite previous findings focussing predominately on L1, L2 learners do retain verbatim information and represent it in their mental text model during reading. The following study delivers evidence that L2 learners are more sensitive to such information than native speakers.

The present study

In the present study, participants read six short texts twice while their eye movements were tracked. There were eight regions of interest (ROIs) in each text. Four ROIs involved nouns (lexical conditions) and the other four ROIs involved two sentences in active and two in passive voice (syntactic conditions). The second version (V2) of each text differed from the first version (V1) in that two ROIs always remained the same and two were changed. In the lexical conditions, the noun was exchanged for its near-synonym. In the syntactic conditions, active was transformed into passive and vice versa. The reading behaviour was evaluated at the critical ROIs in same (i.e., no change) vs. changed conditions in V2. The rationale of the design was that if readers represent verbatim surface information in their mental text model, they should be sensitive to the changes in V2 and respond to them with longer fixation times.

Contrary to most previous studies based on single sentences or very brief texts, the participants in our study read texts of c. 300-400 words. This aspect is important because according to, e.g., Gernsbacher (1985), coherent texts are assumed to facilitate the creation of a summary gist and therefore the integration of sentences into coherent semantic representations is causally involved in the loss of verbatim memory. At the same time, our participants knew they would have to answer questions about the contents of the texts, but the task did not include anything that could draw their attention to the surface form of the texts. For the participants, the focus on content questions justified the second reading. Thus, though being indeed different from a leisure reading situation, the testing situation in our study was close to a natural situation when students read and reread a text for a studying purpose. Importantly, the experimental measurements took place during a phase preceding the assumed testing phase (i.e., answering the comprehension questions). In addition, the data collection through eye-tracking did not require any explicit recall or sentence generation as is the case in most previous studies in this area. Thus, it was possible to address the type of knowledge which might be stored in the mental text model but is not accessible for explicit recall. Research on vocabulary acquisition has repeatedly shown that more lexical knowledge is stored (often associated with the so-called passive vocabulary) than what is revealed through explicit vocabulary tests (active vocabulary) (Bordag et al., 2014, 2015, 2016a, 2016b; Ellis, 2004, 2005; Laufer, 1998).

The critical syntactic manipulation of the present study (active vs. passive alternation) resembles manipulations applied in structural priming. Structural priming studies have demonstrated that the processing of a linguistic unit is facilitated by the recent processing of a structurally equivalent form (see Pickering & Ferreira, 2008, for a review) and that the priming effect persists also when the sentence prime and the target picture are separated by unrelated filler sentences (Bock & Griffin, 2000; Bock, Dell, Chang & Onishi, 2007, for replication). Interpretations of this phenomenon are based on activation networks that represent phrasal constituents as nodes and explain the priming effects either as facilitation

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

due to recency of activation (Pickering & Branigan, 1998; Roelofs, 1992, 1993, for lexical priming), or as a form of implicit learning (Bock & Griffin, 2000; Chang, Dell & Bock, 2006; Chang, Dell, Bock & Griffin, 2000). However, the crucial difference between a structural priming manipulation and the manipulation in the present study is that while the prime and the target sentences are separated by (both structurally and semantically) unrelated filler sentences in the syntactic priming studies, in the present study the critical sentence in V1 is separated from the critical sentence in V2 by coherent text passages that include sentences with the structure counterparts (e.g., other passive and active sentences were read before the target active/passive sentence in V2 was reached). Consequently, the expected effect in the present study is unlikely to be attributed to syntactic priming, which is assumed to persist only until its syntactic counterpart is encountered (e.g., Bernolet, Collina & Hartsuiker, 2016, p. 100). In addition, the lexical overlap between the prime and the target sentence in syntactic priming is either absent or only partial (typically the verb, cf. lexical boost effect) to reduce the effect of explicit memory for the prime sentence. In the present study, the lexical material in the critical sentence in V1 and in V2 is identical (as all non-critical sentences in V1 and V2); only its syntactic structure alternates. Similar to syntactic priming studies, we assume that the repeated lexical material acts as a retrieval cue, so that the reader is more likely to recall the critical sentence from V1 (from the prime sentence in the syntactic priming studies) from memory and anticipate or predict it – including its structure – during reading of V2 (reuse it in syntactic priming studies). However, such an anticipation effect can be expected only if participants stored verbatim the surface linguistic information of the sentence, not just its gist or lexical content (cf. Freunberger & Roehm, 2016, who show that L1 readers predict conceptual-semantic information but not specific surface-form information).

Method

In the present study, we tested two groups of participants, L1 native German speakers and advanced L2 German learners. Since the method was the same, we will present it together for both groups, highlighting the differences where relevant.

Participants

L1: Twenty-four students at the University of Leipzig participated in the experiment. They were all native speakers of German who had normal or corrected-to-normal vision. Their mean age was 24.5 years (min = 20, max = 33).

L2: Twenty-four exchange students at the University of Leipzig participated in the experiment. They were native speakers of Slavic (8) and Romance (16) languages[3] with German knowledge at B2-C1 level (typically B2 in productive and C1 in comprehension skills). They had normal or corrected-to-normal vision. Their mean age was 26.3 years (min. 21, max. 39).

Materials

Six texts were constructed, which each participant read in two versions, V1 and V2. The topics of the texts were Languages of the World, Beethoven, Coffee, Depression, Museums, and Nutella. The average length of each text was 350 words (min. 319, max. 391). They were of average difficulty with LIX index (following Björnsson, 1968) between 46.4 and 55.2 (see Table 1). The texts contained only vocabulary and grammar familiar to German learners at B2 level as assessed in a pre-test.[4]

Table 1. Overview of texts statistics

Text Topic	Word Count	Mean Length of Sentence (in words)	LIX	'predicate'
Languages of the World	347	16.3	53.2	intermediate
Beethoven	341	14.6	46.7	intermediate
Coffee	334	15.7	48.3	intermediate
Depression	319	15.8	55.2	high
Museums	369	14.1	46.4	intermediate
Nutella	391	13.2	47.5	intermediate

There were eight ROIs in each text, none of them appeared in the two first sentences or in the last sentence of the text. The ROIs of the four lexical conditions consisted of single nouns; the four syntactic ROIs were sentences. The items in the lexical conditions were 24 synonym or contextual-synonym pairs; the items in the syntactic conditions were 24 active/passive sentence pairs. Critical lexical items or sentences were always separated by at least one line on the screen. The synonym pairs were constructed such that the two nouns were fully exchangeable in the given context. One pair member always had lower frequency compared to the other pair member. The difference between the frequencies between the pair members was always at least two frequency classes according to Deutscher Wortschatz[5] (Goldhahn, Eckart & Quasthoff, 2012), see Appendix S1 (Supplementary Material). At the same time, the low-frequency nouns were also longer (high 7.67 (*SD* = 2.91), low 9.79 (*SD* = 4.29) letters on average). The critical nouns always appeared only once in all texts.

Given the analytic way of constructing German passives, the passive sentences were always longer than active sentences (i.e., they contained an auxiliary verb, see (1) and (2)). The readability of the texts was assessed in the same pre-test in which the interchangeability of the synonyms was assessed and there was no statistical difference in the readability ratings between the text versions.

(1) active:
Jedes Jahr am 1. Oktober veranstaltet die European Depression Association (EDA)
[every year at 1.October organizes the European Depression Association (EDA)]
deswegen den Europäischen Tag der Depression.
[therefore the European Day of Depression]
"The European Depression Association (EDA) thus organizes European Day of Depression every year on October 1."

(2) passive:
Jedes Jahr am 1. Oktober wird deswegen von der European Depression
[every year at 1. October AUX.PASSIVE therefore of the

[3]Explorative analyses of the Slavic vs. Romance data did not indicate any differences between the populations. The number of participants in the groups was however too small to draw any conclusions. Further research is necessary to assess the role of L1 in retention of verbatim information during reading in L2.

[4]All materials/texts are publicly accessible via the IRIS database.

[5]Public access via: http://wortschatz.uni-leipzig.de/

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

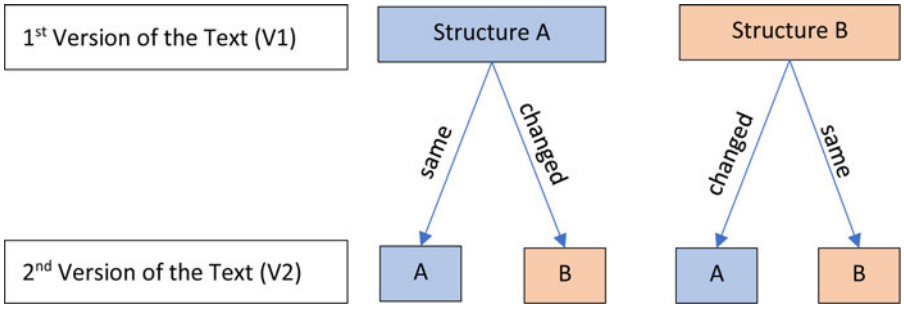


Fig. 1. Scheme of the Experimental Manipulation.

Table 2. Examples for Same/Changed Alternations for the Lexical Conditions

First Version of the Text (V1)		Alternation		Second Version of the Text (V2)
Talent 'talent'	→	Same	→	*Talent* 'talent'
Begabung 'giftedness'	→	Changed	→	*Talent* 'talent'
Begabung 'giftedness'	→	Same	→	*Begabung* 'giftedness'
Talent 'talent'	→	Changed	→	*Begabung* 'giftedness'

European Depression]
Association (EDA) der Europäische Tag der Depression veranstaltet.
[Association (EDA) the European Day of Depression organized.PASSIVE]
"European Day of Depression is thus organized by the European Depression Association (EDA) every year on October 1."

In each text and in each version, there were always two items of each type in each condition, i.e., two high-frequency synonyms and two low-frequency synonyms; and two active sentences and two passive sentences. ROIs of the lexical conditions did not overlap with ROIs of the syntactic conditions. Versions V1 and V2 differed in which ROIs were changed and which remained the same. In each V2, two syntactic and two lexical ROIs remained the same as in V1; the two other ROIs were exchanged for the other pair member (see Figure 1). Which ROIs remained the same and which were changed was determined by rotations according to the Latin square design and cross-balanced over all participants.

The critical comparisons of reading behaviour were at the same vs. changed ROIs in lexical and syntactic conditions during second reading (V2), e.g., total fixation durations at the word "Talent" after "Talent" in V1, or after "Begabung" in V1 (see Table 2) and at the active sentence after the active sentence in V1, or after its passive counterpart in V1.

Procedure

Participants were told that they would read a text twice, solve some mathematical tasks before the second reading and answer comprehension questions afterwards. The sequence of the main events during the experiment was as follows: 1) V1 text was presented; 2) participants solved two non-trivial mathematical tasks; 3) V2 text was presented; 4) participants answered two comprehension questions. Both the mathematical tasks and the comprehension questions were multiple-choice questions in which participants had to choose from four (mathematical questions) or three (comprehension questions) options. An example of a mathematical question would be "What is the value of x in: x = 5–2 * 5 + 2?" with possible answers 21, 17, -3, and -5. The purpose of the mathematical questions was to distract participants from the text and prevent them from consciously rehearsing parts of the text. The purpose of the comprehension questions was primarily to justify the second reading of the text and to check participants' comprehension. The questions targeted both text details and inferencing that required close reading. There was a short break after the third text was read. The order of text presentation rotated according to Latin square design.

During the experiment, participants were seated approximately 60 cm from a monitor. Both eyes were tracked using either an SMI eye-tracker REDm or RED 250mobile. The calibration took place at the beginning of the session and before the fourth text (after the break). Each text was split into parts respecting paragraph beginnings and ends, and presented on 6-7 successive screens. All texts were presented double-spaced, in 21-point Courier New font, left aligned on the display in black letters on a light-grey background. Gaze direction was sampled at a rate of 250 Hz (RED 250mobile) for half of the participants and at 120 Hz (REDm) for the other half.[6] Apart from the use of two different eye-trackers, the experimental material was presented to all participants identically on the same screen (Hannspree JC199D, 19 inches) with the same screen resolution (1280 by 1024 pixels, 60Hz). The session started with a short practice text (77 words, two screens) presented in the same format as the critical texts, but with no manipulations between V1 and V2. At the end of the experiment participants filled in a questionnaire. The L2 questionnaire also included language-history questions and a detailed self-assessment questionnaire, which was shown to correspond well with proficiency tests in previous studies (Bordag et al., 2014, 2015, 2016a, 2016b). The entire experimental session took approximately 75 minutes for the L1 and c. 90 minutes for the L2 participants.

Data analysis

The same procedures for cleaning and analysing the eye-tracking data were performed for the L1 and L2 data. For the lexical conditions (synonyms), the critical word was defined as ROI and the following two words were defined as potential spill-over regions.

[6]This approach was taken because eye-trackers had to be borrowed from two institutions. Though the RED 250mobile eye-tracker provided a more fine-grained temporal resolution, factor "Eye-tracker type" when taken into the statistical analyses did not interact with any other factors.

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

For the syntactic conditions (active-passive alternation), the entire clause was defined as ROI. All ROIs were defined using the software BeGaze (version 3.6). In addition, data in ROIs were scanned manually and drifts were corrected prior to further analyses (8% of fixations). For detection of fixations and saccades, BeGaze 3.6 was used with default settings. Fixations with durations shorter than 80 ms were removed. We also removed fixations with durations exceeding a participant's mean plus 3.5 standard deviations (excluding c. 0.5% and 0.7% of the ROI data in the L1 and L2 experiment, respectively).[7] In addition, we excluded three participants (two L1 and one L2) who reported that they noticed a word or two had been changed in the texts at the debriefing.

Since the experiment had an identical design in L1 and L2, both language groups were analysed in one statistical analysis, with the factor Language comprising two levels (L1 vs. L2). The analyses were performed using linear mixed-effect models employing the software R (R Core Team, 2018 with packages *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2017) and *afex* (Singmann, Bolker, Westfall, Aust & Ben-Shachar, 2020). By default, categorial independent variables were effect coded. All models included a maximum random effect structures with all fixed effects and their interactions as both intercepts and slopes for participants and items (Barr, Levy, Scheepers, and Tily, 2013). Fixation times as dependent variables were log-transformed for the analyses (model structures, i.e., the R model call, are given in corresponding Table legends). Fixed effects were Language (L1 vs. L2), Alternation (same vs. changed), and Form (high versus low-frequency member of the synonym pair; active versus passive voice of the syntactic alternation). Significance of fixed-effects was determined by model comparisons using likelihood ratio tests; fixed effect terms were computed with Satterthwaite method for denominator degrees of freedom for *t* tests. Relevant outcomes of the analyses are reported in the following section (Results), a full report of all statistics is provided in Appendix S2 (Supplementary Material) [8].

According to the rationale of the study, the main focus of the critical analyses was on reading behaviour of the second presentation of the texts, comparing ROIs that were either 'changed' or 'same' with respect to corresponding ROIs of the first presentation (factor 'Alternation').[9] We expected that participants should fixate critical ROIs in the 'changed' conditions more than in the 'same' conditions as a result of a surprisal effect. Under the assumption that participants stored the previous verbatim wording, such an effect could be caused by the fact that a new, previously unencountered word or structure was perceived. Thus, we expected changes in the reading behaviour especially in the total fixation durations measure and in the number of fixations for the syntactic conditions given the length of the ROIs (in the lexical conditions, we considered differences in the number of fixations less likely, since only single words were concerned). These were thus the two measures that we analysed. Lexical conditions (synonyms) and syntactic conditions (active-passive alternation) were analysed separately and are reported in two subsections.

To assess baseline differences between the two tested populations with respect to the first and second reading of the texts, we additionally compared reading behaviour for all unchanged ('same') regions of V1 and V2.

Though the primary purpose of the content questions was to justify the second reading of the text and to keep participants attentive to the content, we analysed the responses to these questions as well. They are reported at the end of the Results section.

Results

Lexical conditions (synonyms)

Results for lexical conditions are summarised in Table 3. Analysis of mean NUMBER OF FIXATIONS in each ROI revealed main effects of Language ($\beta = -0.187$, $SE = 0.067$, $p = .008$)[10] and Form ($\beta = -0.219$, $SE = 0.077$, $p = .009$) indicating that, in general, the L2 participants produced more fixations than the L1 participants (2.19 vs. 1.82) and that low-frequency, longer members of the synonym pairs elicited more fixations than their high-frequency, shorter counterparts (2.23 vs. 1.78). However, there was no effect for the factor Alternation ($\beta = -0.053$, $SE = 0.039$, $p = .182$), nor was there any significant interactions of the fixed effects (all $p > .20$)[11].

The analyses of TOTAL FIXATION DURATIONS yielded main effects of Language ($\beta = -0.156$, $SE = 0.028$, $p < .001$), Form ($\beta = -0.125$, $SE = 0.032$, $p = .001$) and Alternation ($\beta = -0.049$, $SE = 0.016$, $p = .003$). Thus, beyond the fact that the L2 participants fixated the critical regions longer than the L1 participants (514.2 ms vs. 379.6 ms) and that total fixation durations were also longer for low-frequency items compared to high-frequency counterparts (518.5 ms vs.389.2 ms), the statistical analyses provided evidence that the main experimental manipulation (Alternation) had an influence on total fixation durations in the expected direction. Participants fixated regions in the changed conditions longer (475.5 ms) than in the same conditions (428.0 ms). Although the numerical difference for same vs. changed regions was larger in L2 (16.7 ms in L1 and 78.4 ms in L2), the corresponding Interaction Language vs. Alternation failed to be significant ($\beta = 0.013$, $SE = 0.018$, $p = .473$).[12] The results for total fixation durations are illustrated in Figure 2.

Results for the lexical conditions (synonym alternations) can thus so far be summarised in that the experimental manipulation had an influence on participants' fixation latencies. Participants fixated regions with exchanged lexical material significantly longer than unaltered regions. Although this effect was numerically more pronounced for L2 participants, statistically no significant interaction of Language and Alternation was observed.[13] As expected, participants also showed longer fixation durations for

[7]Note that we thus opted for an individual upper bound for each participant when defining a fixation instead of setting a fixed value, e.g., 1500 ms. We did so in order to account for the relatively large inter-individual differences between participants. However, most data points removed had fixation latencies > 2000 ms.

[8]Appendix S2 (Supplementary Material) provides the output of all statistic models for each measure as i) a joint analysis for both populations including the factor Language (i.e., L1 and L2) as well as ii) separate analyses for both groups (L1 and L2).

[9]See Supplementary materials (Supplementary Material) for a comparison with the first reading data and a discussion of tendencies that went beyond the main research question.

[10]For details of statistical results (including *t*-values) see Appendix S2 (Supplementary Material).

[11]Please see Appendix S2 (Supplementary Material) for a detailed report of statistic models.

[12]Separate analyses for L1 and L2 revealed that the factor Alternation was significant only for L2 participants ($\beta = -0.061$, $SE = 0.023$, $p = .012$), but not for L1 participants ($\beta = -0.034$, $SE = 0.027$, $p = .230$), indicating that the main effect of Alternation was mainly driven by L2 participants.

[13]Spill-over regions (the following two words) were also analysed, both separately and together with the critical regions. In general, effects were the same as for the critical word alone, but smaller in size. No effect emerged at spill-over regions that was not present on the critical word itself. Our results presented here thus focus on the critical region alone.

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

Table 3. Results of Lexical Conditions (Synonyms) in L1 and L2

		L1			L2			L1 & L2
		high	low	*mean*	high	low	*mean*	*mean*
Mean Number of Fixations (with SDs)								
	changed	1.63 (0.86)	2.10 (1.40)	***1.87 (1.19)***	1.99 (1.29)	2.53 (1.67)	***2.26 (1.51)***	***2.06 (1.38)***
	Same	1.63 (0.81)	1.92 (1.11)	***1.78 (0.98)***	1.87 (1.07)	2.36 (1.31)	***2.12 (1.22)***	***1.95 (1.12)***
Effect of Alternation	*diff.*	*0.01*	*0.18*	***0.09***	*0.12*	*0.17*	***0.14***	***0.11***
Total Fixation Durations: Means in ms (with SDs)								
	changed	340.5 (209.0)	435.3 (286.6)	***387.9 (255.7)***	471.9 (314.1)	654.2 (452.6)	***563.1 (400.1)***	***475.5 (350.3)***
	Same	326.4 (211.8)	416.0 (314.2)	***371.2 (272.0)***	418.0 (267.6)	551.4 (321.1)	***484.7 (303.6)***	***428.0 (294.0)***
Effect of Alternation	*diff.*	*14.1*	*19.3*	***16.7***	***53.9***	***102.8***	***78.4***	***47.5***

Note: high vs. low are abbreviations for the high- versus low-frequency member of a synonym pair, e.g., *Talent* "talent" (H, more frequent) vs. *Begabung* "giftedness" (L, less frequent)

the low-frequency member of the synonym pair than for its more frequent counterpart. But importantly, this main effect did not interact with the Alternation manipulation. This shows that the alternation effect was not driven by cases when high-frequency, shorter nouns were exchanged by their low-frequency, longer synonyms (which could be viewed as more salient for noticing) during the second reading. The significant frequency effect also indicates that both groups of participants were sensitive to the lexical properties of the critical words and that the smaller L1 effects cannot be attributed to e.g., skimming of the given text part. To sum up, the results of the analyses of the lexical conditions provide evidence that both groups of participants are sensitive to changes in surface linguistic information on the lexical level.

Syntactic conditions (active – passive alternation)

Results for the syntactic conditions are summarised in Table 4. In the analyses of the NUMBER OF FIXATIONS, main effects were observed for the factors Language ($\beta = -0.669$, $SE = 0.294$, $p = .027$) and Form ($\beta = -0.942$, $SE = 0.129$, $p < .001$): L2 participants produced more fixations than L1 participants and passive sentences led to more fixations than active sentences.[14] Although there was no main effect for Alternation ($\beta = -0.257$, $SE = 0.171$, $p = .151$), the interaction between Language and Alternation was significant ($\beta = 0.519$, $SE = 0.209$, $p = .024$). This interaction indicates, in contrast to the lexical conditions, that the syntactic alternation (i.e., the change from active to passive and vice versa) influenced reading behaviour differently for participants in L1 and L2. To resolve this interaction, both groups were subsequently analysed separately. This revealed that only L2 participants were sensitive to the Alternation manipulation (main effect of Alternation: $\beta = -0.814$, $SE = 0.348$, $p = .032$). They produced more fixations in the changed ROIs (12.71) than in the same ROIs (10.35). No such significant difference was observed for the L1 participants (10.63 vs. 10.34; $\beta = 0.243$, $SE = 0.151$, $p = .121$).

[14]Also, in this case, the results regarding the number of fixations with respect to the factor Form has little relevance, since passive sentences were always longer.

The analysis of TOTAL FIXATION DURATIONS revealed a similar pattern. Again there were main effects for Language ($\beta = -0.114$, $SE = 0.032$, $p = .001$) and Form ($\beta = -0.095$, $SE = 0.018$, $p < .001$), but no main effect for the factor Alternation ($\beta = -0.020$, $SE = 0.014$, $p = .180$). However, there was a significant interaction between the factors Language and Alternation ($\beta = 0.034$, $SE = 0.015$, $p = .040$). Separate analyses revealed that the main effect of Alternation was significant only for L2 ($\beta = -0.058$, $SE = 0.027$, $p = .046$), but not for the L1 participants ($\beta = 0.013$, $SE = 0.016$, $p = .400$). The L2 participants thus produced shorter fixation times in the unchanged ROIs (2328.8 ms) than in changed ROIs (2960.9 ms). The (numerically very small) difference (−48 ms) was not significant for the L1 participants. Additionally, there was no significant interaction of Form with Alternation ($\beta = 0.007$, $SE = 0.014$, $p = .611$), indicating that the effect of the experimental manipulation (i.e., increased fixation times in the changed conditions) was independent of the direction of that change, i.e., the same effect was seen for sentences in active and passive voice. The results for total fixation durations are also illustrated in Figure 3.

In sum, the results from the syntactic conditions revealed that the L2 participants produced more fixations and longer total fixation times than the L1 participants did in general. Most interestingly, the significant interactions of Language and Alternation show that only for the L2 participants there were more fixations and longer total fixation durations for ROIs in the 'changed' compared to the 'same' conditions. No such difference was observed for the L1 participants. This observation supports the initial hypothesis that the L2 learners are more sensitive to changes in surface form than the L1 speakers.

The pattern of significant effects for the factor Alternation (changed versus same conditions) is summarised in Table 5.

Comparison between first and second reading

For all unchanged ROIs, an additional comparison was carried out for the first versus second reading of the texts. This analysis can be regarded as a baseline and diagnostic to assess whether L1 and L2 participants benefitted to the same degree from the repetition of the texts in general. Analyses were carried out over

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

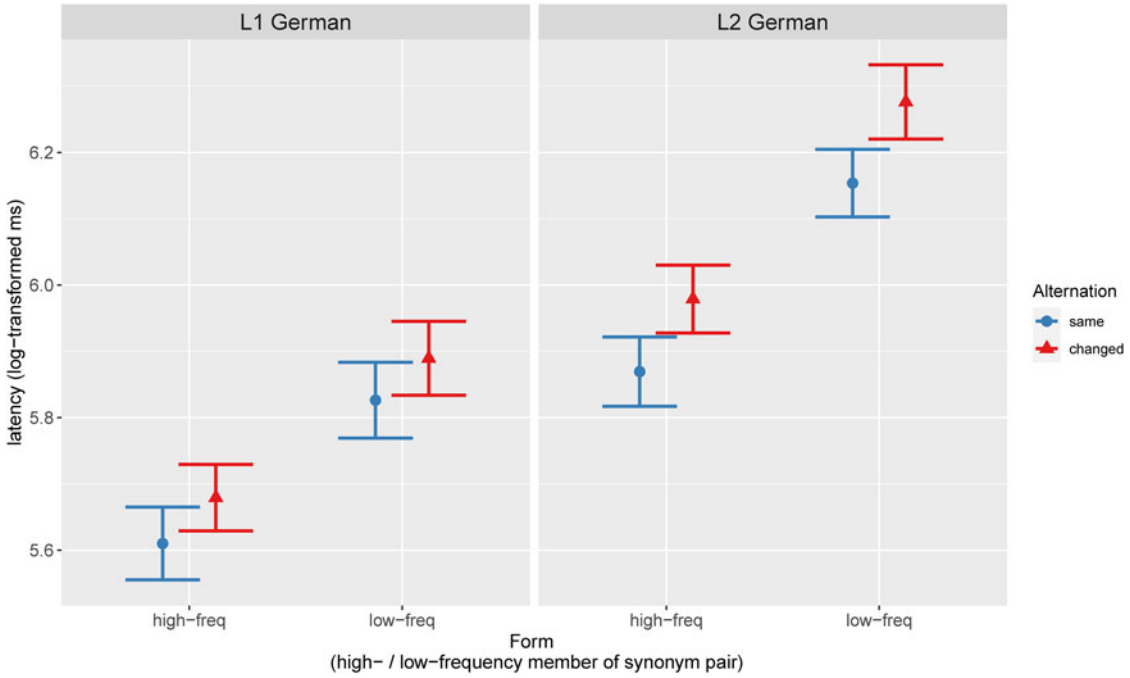


Fig. 2. Summary of total fixation durations during second reading for the lexical conditions (means of log-transformed latencies with *SE*).

all unchanged (i.e., 'same') ROIs with the additional factor Reading (V1 vs. V2). The changed conditions were not included in the analyses, because it involves the confounding factor of the changed linguistic material (that was expected to increase the reading times and to affect the L2 readers more than the L1 readers).

Lexical conditions

With respect to the NUMBER OF FIXATIONS, we observed a significant difference for Language ($\beta = -0.210$, $SE = 0.046$, $p < .001$) and Form ($\beta = -0.191$, $SE = 0.069$, $p = .011$), but no significant difference for Reading ($\beta = 0.018$, $SE = 0.034$, $p < .601$) nor any interaction of this factor with other factors (all $p > .100$). For TOTAL FIXATION DURATIONS, again effects for Language ($\beta = -0.167$, $SE = 0.026$, $p < .001$) and Form ($\beta = -0.116$, $SE = 0.027$, $p < .001$) were observed. This time, an additional effect for Reading was seen, too ($\beta = 0.047$, $SE = 0.017$, $p = .009$). However, the factor Reading did not interact with Language ($\beta = -0.022$, $SE = 0.017$, $p < .195$). Thus, for the lexical conditions, a general effect of shorter fixation times was seen for the second reading of the text, but this difference was statistically indistinguishable for both groups of participants.

Syntactic conditions

Analyses of the NUMBER OF FIXATIONS only revealed significant main effects for Reading ($\beta = 0.399$, $SE = 0.108$, $p = .003$), and Form ($\beta = -0.884$, $SE = 0.218$, $p = .002$). Regarding *total fixation times*, beyond a main effect of Language ($\beta = -0.087$, $SE = 0.032$, $p = .009$), the analyses showed also main effects of Form ($\beta = -0.079$, $SE = 0.022$, $p = .003$), and Reading ($\beta = 0.050$, $SE = 0.012$, $p < .001$). These results show that the active sentences were read faster than passive sentences and that both active and passive sentences were read faster during the second reading. Importantly, none of the interactions reached significance. Crucially, the non-significant interaction of Language and Reading ($\beta = -0.012$, $SE = 0.010$, $p = .241$) indicates that the effect of repetition on reading times was identical for L1 (diff. = 166.7 ms) and L2 (diff. = 154.9 ms). For detailed analyses including a Figure comparing all conditions of the first and second reading see Appendix S3 (Supplementary Material). Thus, a repetition effect of the same size was observed in both L1 and L2 in the second reading, which complies with the finding that when readers read the same text twice, they speed up on second reading (Hyönä & Niemi, 1990; Raney & Rayner, 1995).

Importantly, though, the pattern of results shows that the different effects for the factor Alternation in the syntactic conditions in L2 cannot be explained by, e.g., a combination of frequency and repetition effects: in the lexical conditions, the repetition effect was the same in L1 and L2, and there was no significant difference in the alternation effect for the two language groups. In the syntactic conditions, the repetition effect was again the same in L1 and L2, and yet we observed a robust alternation effect in L2 only. The alternation effect in the syntactic conditions thus cannot be attributed just to slower reading in L2 per se.

Content questions

Answers to the content questions were analysed using a generalised mixed-effect model with a binomially distributed dependent variable Answer (correct vs. incorrect). The factor Language was set as a fixed effect and text and participant as random effects with a maximum random-effects structure justified by the experimental design. Results revealed that, although both groups of participants differed numerically in their accuracy rates (L1: 76.8% vs. L2: 72.3%), the difference between the two groups was not significant ($\beta = -0.302$, $SE = 0.211$, $p = .154$). This result also shows that even though the L2 participants read more slowly, this did not result—in general—in better memory for information contained in the text. We thus assume that slower reading per se is not responsible for better memory for linguistics structure in

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

Table 4. Results of Syntactic Conditions (Active – Passive) in L1 and L2

		L1			L2			L1&L2
		active	passive	*mean*	active	passive	*mean*	*mean*
Mean Number of Fixations (with SDs)								
	changed	9.65 (4.68)	11.03 (5.11)	***10.34 (4.94)***	11.31 (5.99)	14.10 (6.63)	***12.71 (6.47)***	***11.53 (5.90)***
	same	9.86 (5.22)	11.39 (4.98)	***10.63 (5.15)***	9.30 (3.99)	11.39 (3.83)	***10.35 (4.04)***	***10.49 (4.65)***
Effect of Alternation	*diff.*	*−0.21*	*−0.36*	***−0.29***	***2.01***	***2.71***	***2.36***	***1.04***
Total Fixation Durations: Means in ms (with SDs)								
	changed	2065.9 (1195.1)	2316.1 (1160.8)	***2191.0 (1181.5)***	2657.1 (1543.3)	3264.6 (1651.9)	***2960.9 (1625.5)***	***2575.9 (1478.1)***
	same	2109.1 (1341.7)	2368.8 (1221.3)	***2239.0 (1286.7)***	2171.4 (915.6)	2486.1 (829.2)	***2328.8 (887.0)***	***2283.9 (1111.8)***
Effect of Alternation	*diff.*	*−43.2*	*−52.7*	***−48.0***	*485.7*	*778.5*	***632.1***	***292.0***

L2. Previous research also indicates that longer reading times in L2 are due to various factors such as e.g., slower word recognition processes rather than to a deeper or more thorough processing (Grabe, 2008: 122ff).

General discussion

Previous research in L1 generally converges on the claim that readers do not retain surface linguistic information verbatim during reading and this view has not been challenged until recently (Gurevich et al., 2010). At the same time, usage-based theories about language acquisition made assumptions that lead us to the hypothesis that the situation might be different for language learners. In particular, we hypothesised that L2 readers retain surface linguistic information and rely on it more than L1 readers. The results of our experiment supported this assumption.

In accordance with the previous L1 research, we did not find convincing evidence that verbatim information would be represented in L1 mental text models. Though the lack of an interaction between the factors Language and Alternation indicates that both L1 and L2 learners were sensitive to substitutions at the lexical level, the effect of Alternation in L1 was numerically rather small (16.7 ms compared to 47.5 ms in L2) and could be also explained by lexical-semantic differences between the synonyms (see below). The effect manifested itself only in the analyses of total readings times. Given that the ROIs comprised only one word, it is not surprising that no effect was observed in the number of fixations measure (as hypothesized).

In the syntactic conditions, the interaction between the factors Language and Alternation was significant and subsequent analyses showed that no differences in reading behaviour of the changed vs. same regions were observed for L1 participants. There was no evidence that L1 readers were sensitive to the introduced changes at the syntactic level.

Thus, we found only weak support for the finding of Gurevich et al. (2010) which indicated that surface information might be retained in L1. At the same time, our results do not rule out the possibility that also L1 speakers represent and retain surface linguistic information, but rely on it less than L2 learners. Readers (or listeners, as in Gurevich et al.'s study) can, for example, show different sensitivity to different language structures. Our study involved only (contextual) synonym substitutions and active/passive alternations. In Gurevich et al.'s study, the authors employed a wide range of structures and rephrasings. It is a question for future research as to what degree different structures or formulations are retained verbatim in L1 mental text models and which further factors (like, e.g., the modality of presentation, involvement of lexical semantics) possibly influence their retention and the role they play during processing.

The question why the sensitivity to the changes in L1 in our study is limited only to the lexical items and does not include syntactic structures needs to be addressed in detail, too. We hypothesize that if L1 readers represent only propositional contents of the text in their mental text model, mental representations of lexical units, but not particular syntactic structures, might be comprised in the model. In the experiment, some of the synonyms were fully exchangeable only in the given contexts, but would have only partially overlapping semantic representations as isolated words (e.g., THEORY vs. HYPOTHESIS). Lexical differences between the synonyms thus might result in different conceptual-semantic representations at the propositional level, which is a part of the mental text model. When the changes concerned only the surface information (while lexical semantics were identical) as it was in the syntactic conditions, L1 readers showed no indication that this information would be a part of their mental text model. This is in accordance with studies on prediction that provide evidence that native speakers predict conceptual-semantic information, but not form-specific information during reading (Freunberger & Roehm, 2016).

In contrast to the results of the L1 speakers, we found robust evidence that L2 learners retain surface linguistic information during reading and that they rely on its representation more than L1 speakers as evidenced especially by the results in the syntactic conditions, for which differences were significant only for L2 learners. The effect was observed both in the analysis of the total reading times and the number of fixations.

The results of the experiment support the initial hypothesis that verbatim information plays a different role in native and non-native reading. L2 participants were sensitive to the verbatim changes in both the lexical and syntactic conditions. This sensitivity implies that they retained the surface linguistic information of the first text version in memory, because only in this case they

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

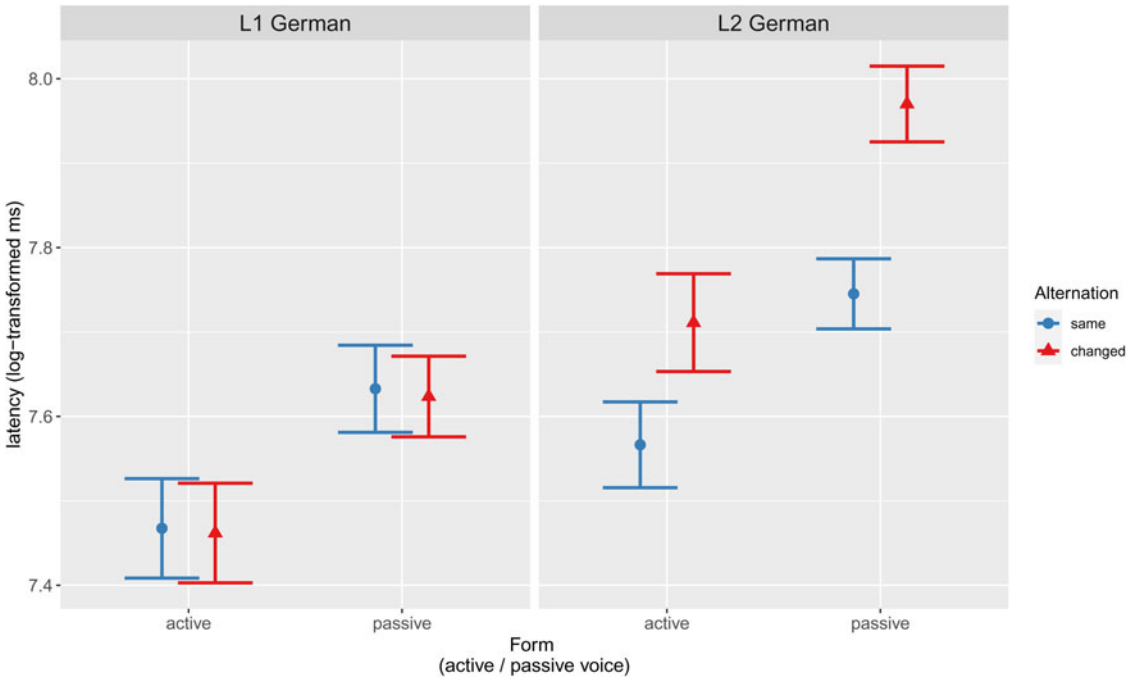


Fig. 3. Summary of total fixation durations during second reading for the syntactic conditions (means of log-transformed latencies with *SE*).

Table 5. Pattern of significant effects of Alternation (significance level $p < .05$)

	lexical		syntactic	
Measure	L1	L2	L1	L2
Number of Fixations	–		–	+
Total Fixation Durations	+		–	+

could show a surprisal effect at encountering an alternative lexical or syntactic form in the second text version (while the semantic and propositional content remained the same). The eye-movement data thus offer unambiguous evidence that L2 learners retain verbatim information in memory and that they are more sensitive to it than L1 speakers at least on the syntactic level. Importantly, being explicitly asked about whether they noticed anything interesting or worth mentioning after the experimental session, none of the participants included in the analyses reported noticing that words or formulations changed between the text versions. Typically, they commented on the mathematical tasks and the content questions. This indicates that participants were not explicitly aware of the text manipulation, though their eye-tracking data reveal that they reacted to it.

Our study thus supports the results of Sampaio and Konopka (2013), but advances them in two important directions. First, unlike Sampaio and Konopka or also, e.g., Kintsch and Bates (1977) and Gibbs (1981), we did not test verbatim memory of open-class lexical items only, but also of syntactic information. The inclusion of the syntactic conditions is important because it demonstrates that verbatim memory goes beyond a memory for particular words (cf. Gurevich et al., 2010, p. 48). Second, we did not test verbatim memory within isolated sentences, but within coherent texts. This aspect is crucial because, as pointed out (by, e.g., Gernsbacher, 1985), the integration of sentences into coherent semantic representations is causally involved in the loss of verbatim memory – coherent texts are assumed to facilitate the creation of a summary gist. The results of our study thus have direct relevance also for theories of mental text models: to our knowledge, our study is the first to show not only that L2 readers represent verbatim information and rely on this representation more than L1 readers, but that the retention of the surface linguistic information persists also during reading of coherent texts. A pioneering aspect of our study is the employment of eye-tracking to explore information retention during reading. In the previous research (also that including exploration of mental text models) participants have typically been asked to give explicit judgements (e.g., the recognition experiment by Gurevich et al., 2010, or the task in Sampaio & Konopka, 2013) or to productively recall the critical sentences/text (e.g., production experiment of Gurevich et al., 2010). Such approaches enable an assessment of explicit knowledge that participants are able to retrieve, yet the human brain stores much more than what can be consciously accessed or actively produced. In fact, we do not know how much of what we cannot say or report is actually lost, and what is still stored in memory but cannot be retrieved (cf. Pierce, Klein, Chen, Delcenserie & Genesee, 2014). Employing the eye-tracking technology might offer insights into processing and representation of linguistic information that go beyond findings accessible to methods which rely on overt knowledge retrieval.

In the introductory section we have shown that though our hypothesis regarding the overreliance on surface linguistic information by L2 learners compared to L1 speakers is novel, it can be predicted based on assumptions made by cognitive, acquisition, and processing theories. Our results comply with theories postulating different processing for individuals with a different degree of expertise. L2 learners can be seen as novices in L2 reading compared to L1 readers who – at least in the case of university students as typical participants in psycholinguistic studies with

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

adults – can be classified as experts. According to the FTT theory, the dominance of verbatim representations of past events (of which a recently read text is an instance) is associated with lower expertise, while experts rely more on gist-based information. This claim is in accordance with our result that while L2 learners demonstrate stronger reliance on surface linguistic information than L1 speakers, they do not outperform native speakers in answering the content questions. Further research is necessary to determine whether reliance on verbatim information is specific to L2 processing, or whether it is a more general cognitive strategy of novices as suggested by the FTT theory (Experiment 3 in Arndt & Beato, 2017, suggests that it is the latter).

The SSH links the L2 focus on "surface" to a deficit in building abstract syntactic representations. During processes like reading, online analysis of hierarchical syntactic structure may be very demanding or even impossible for L2 learners, navigating them towards a mental text representation that deviates from that of L1 speakers in the amount of surface linguistic information that it contains and/or in the degree on how much they rely on it. Verbatim representation and reliance on it would thus compensate for the (unavailable) results of a deeper syntactic analysis by synthesis of a complete and fully specified syntactic representation. It is an intriguing question for further research whether there might be a causal relationship between non-building a hierarchical syntactic structure and non-decaying of the low-level, surface information that is otherwise assumed to decay immediately after sentence processing (i.e., after the hierarchical syntactic structure had been built).

The results of our study also comply with the DP model: while morphosyntactic processing in L1 proceeds primarily via the procedural memory, L2 learners are assumed to employ the declarative system to process also grammar (Ullman, 2005). The difficulties L2 learners might have with building abstract syntactic structure due to their limited ability to use the procedural system might thus be compensated by means of representation of verbatim information in declarative memory and by relying more on it than on procedural processing. Interestingly, such compensation could not only ease the reading process by partial relocation of the cognitive load from the procedural module to the declarative system; it could also contribute to an enhancement of the L2 repertoire of stored verbatim chunks that can thus be used for acquisition and/or productive use of the L2. Consequently, a representation that is a result of a limitation would at the same time contribute to its overcoming.

Our study brought important evidence concerning the role of surface linguistic information in L1 and L2 reading. While our findings comply with current cognitive, acquisition and processing theories and add a new dimension to their empirical and theoretical bases, they also clearly reveal the need for further research on the topic. We indicated several novel questions that the findings open, and there are indeed more such as, e.g., the actual length of verbatim retention, whether L2 readers – contrary to native speakers – also employ form-specific information in prediction, or whether the fact that L2 readers retain more surface linguistic information has a positive effect on the L2 acquisition process, or whether it might actually hinder it. Also, while our data deliver evidence that L2 readers retain verbatim information and rely on it more than L1 readers, the question about the L1 retention of especially structural verbatim information during reading (and without surface information oriented tasks) still remains open. In the present study, we found no evidence for it, but different approaches or paradigms might be needed to decide whether native readers only rely less on such type of information, or whether they also underrepresent it compared to L2 learners.

Acknowledgements. The authors are grateful to the company SMI for lending their eye-tracker and to Hana Marschner and Raphael Münz for their help with running the experiments and preparing the data.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728920000772
List of Supplementary Materials
Appendix S1: Appendix S1 List of Lexical Items.pdf
Appendix S2: Appendix S2 Statistical Tables.pdf
Appendix S3: Appendix S3 Supplementary Analyses.pdf

References

Anderson JR (1974) Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning and Verbal Behavior* **13**(2), 149–162. doi: 10.1016/S0022-5371(74)80039-3

Arndt J and Beato MS (2017) The role of language proficiency in producing false memories. *Journal of Memory and Language* **95**, 146–158. http://dx.doi.org/10.1016/j.jml.2017.03.004

Barr DJ, Levy R, Scheepers C and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Bernhardt E (2011) *Understanding advanced second-language reading*. New York, NY: Routledge.

Bernolet S, Collina S and Hartsuiker RJ (2016) The persistence of syntactic priming revisited. *Journal of Memory and Language* **91**(6), 99–116. doi: 10.1016/j.jml.2016.01.002

Björnsson C-H (1968) *Läsbarhet (Readability)*. Stockholm, Sweden: Liber.

Bock K, Dell GS, Chang F and Onishi KH (2007) Persistent structural priming from language comprehension to language production. *Cognition* **104**(3), 437–458. doi: 10.1016/j.cognition.2006.07.003

Bock K and Griffin ZM (2000) The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* **129**(2), 177–192. doi: 10.1037//0096–3445.129.2.177

Bordag D, Kirschenbaum A, Opitz A, Rogahn M and Tschirner E (2016a) Der Einfluss syntaktischer Komplexität auf den beiläufigen Erwerb des Wortwissens beim Lesen in Deutsch als Fremdsprache. *Deutsch als Fremdsprache* **53**(3), 159–169.

Bordag D, Kirschenbaum A, Opitz A, Rogahn M and Tschirner E (2016b) Incidental acquisition of grammatical features during reading in L1 and L2. *Studies in Second Language Acquisition* **38**(3), 445–483. doi: 10.1017/S0272263115000315

Bordag D, Kirschenbaum A, Tschirner E and Opitz A (2015) Incidental acquisition of new words during reading in L2: Inference of meaning and its integration in the L2 mental lexicon. *Bilingualism: Language and Cognition* **18**(3), 372–390. doi: 10.1017/S1366728914000078

Bordag D, Kirschenbaum A, Opitz A and Tschirner E (2014) To store or not to store, and if so, where? An experimental study on incidental vocabulary acquisition by adult native speakers of German. In Torrens V & Escobar L (eds), *The processing of lexicon and morphosyntax* (pp. 127–147). Newcastle, United Kingdom: Cambridge Scholars Publishing.

Bybee JL (1985) *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.

Caplan D (1972) Clause boundaries and recognition latencies for words in sentences. *Perception & Psychophysics* **12**(1), 73–76. doi: 10.3758/bf03212846

Chang F, Dell GS and Bock K (2006) Becoming syntactic. *Psychological Review* **113**(2), 234–272. doi: 10.1037/0033-295x.113.2.234

Chang F, Dell GS, Bock K and Griffin ZM (2000) Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research* **29**(2), 217–229. doi: 10.1023/A:1005101313330

Clahsen H and Felser C (2006) Grammatical processing in language learners. *Applied Psycholinguistics* **27**(1), 3–42. doi: 10.1017/S0142716406060024

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

Clahsen H and Felser C (2017) Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition* **40**, 693–706. doi:10.1017/S0272263117000250

Delaney PF, Verkoeijen PPJL and Spirgel AS (2010) Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In Ross BH (Ed.), *The psychology of learning and motivation* (Vol. 53, pp. 63–147). Burlington, MA: Academic Press. doi: 10.1016/S0079-7421(10)53003-2

Ellis NC (1996) Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition* **18**(1), 91–126. doi: 10.1017/S0272263100014698

Ellis R (2004) The definition and measurement of L2 explicit knowledge. *Language Learning* **54**(2), 227–275. doi: 10.1111/j.1467-9922.2004.00255.x

Ellis R (2005) Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition* **27** (2), 141–172. doi: 10.1017/S0272263105050096

Field J (2004) *Psycholinguistics: The key concepts*. London, United Kingdom: Routledge.

Freunberger D & Roehm D (2016) Semantic prediction in language comprehension: Evidence from brain potentials. *Language, Cognition and Neuroscience*. **31**(9), doi: 10.1080/23273798.2016.1205202.

Garnham A (1996) The other side of mental models: Theories of language comprehension. In Oakhill JV & Garnham A (eds), *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird* (pp. 35–52). Hove, United Kingdom: Psychology Press.

Gernsbacher MA (1985) Surface information loss in comprehension. *Cognitive Psychology* **17**(3), 324–363. doi: 10.1016/0010-0285(85)90012-x

Gibbs RW (1981) Memory for requests in conversation. *Journal of Verbal Learning and Verbal Behavior* **20**(6), 630–640. doi: 10.1016/S0022-5371(81)90211-5

Glenberg AM, Meyer M and Lindem K (1987) Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language* **26**(1), 69–83. doi: 10.1016/0749-596X(87)90063-5

Goldberg AE (2006) *Constructions at work: The nature of generalization in language*. Oxford, United Kingdom: Oxford University Press. doi: 10.1093/acprof:oso/9780199268511.001.0001

Goldhahn D, Eckart T and Quasthoff U (2012) Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In Calzolari N, Choukri K, Declerck T, Doğan M, Maegaard B, Mariani J, … Piperidis S (eds), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 759–765). Istanbul, Turkey: European Language Resources Association.

Grabe W (2008) *Reading in a Second Language*. New York: Cambridge University Press.

Gurevich O, Johnson MA and Goldberg AE (2010) Incidental verbatim memory for language. *Language and Cognition* **2**(1), 45–78. doi: 10.1515/langcog.2010.003

Holtgraves T (2008) Conversation, speech acts, and memory. *Memory & Cognition* **36**(2), 361–374. doi: 10.3758/mc.36.2.361

Hyönä J & Niemi P (1990) Eye movements in repeated reading of a text. *Acta psychologica*. **73**(3), 259–80. doi: 10.1016/0001-6918(90)90026-C.

Jarvella RJ (1971) Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior* **10**(4), 409–416. doi: 10.1016/s0022-5371(71)80040-3

Jarvella RJ and Herman SJ (1972) Clause structure of sentences and speech processing. *Perception & Psychophysics* **11**(5), 381–384. doi: 10.3758/bf03206272

Johnson-Laird PN (1977) Procedural semantics. *Cognition* **5**(3), 189–214. doi: 10.1016/0010-0277(77)90001-4

Johnson-Laird PN and Stevenson R (1970) Memory for syntax. *Nature* **227** (5256), 412. doi: 10.1038/227412a0

Just MA and Carpenter PA (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* **99**(1), 122–149. doi: 10.1037//0033-295x.99.1.122

Keysar B, Hayakawa SL and Sun Gyu A (2012) The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science* **23**(6), 661–668. doi: 10.1177/0956797611432178

Kintsch W (1988) The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* **95**(2), 163–182. doi: 10.1037/0033-295X.95.2.163

Kintsch W (1998) *Comprehension: A paradigm for cognition*. Cambridge, United Kingdom: Cambridge University Press.

Kintsch W (1994) Text comprehension, memory, and learning. *American Psychologist* **49**(4), 294–303.

Kintsch W and Bates E (1977) Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory* **3**(2), 150–159. doi: 10.1037/0278-7393.3.2.150

Kroll JF and Stewart E (1990, November). *Concept mediation in bilingual translation*. Paper presented at the 31st annual meeting of the Psychonomic Society, New Orleans, LA.

Kroll JF and Stewart E (1994) Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* **33**(2), 149–174. doi: 10.1006/jmla.1994.1008

Kuznetsova A, Brockhoff PB and Christensen RHB (2017) lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* **82**(13), 1–26. doi: 10.18637/jss.v082.i13

Langacker RW (1988) A usage-based model. In Rudzka-Ostyn B (Ed.), *Topics in Cognitive Linguistics* (pp. 127–161). Amsterdam, Netherlands: John Benjamins.

Laufer B (1998) The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics* **19**(2), 255–271. doi: 10.1093/applin/19.2.255

Lloyd FJ and Reyna VF (2009) Clinical gist and medical education: Connecting the dots. *JAMA: The Journal of the American Medical Association* **302**(12), 1332–1333. 10.1001/jama.2009.1383

Lombardi L and Potter MC (1992) The regeneration of syntax in short term memory. *Journal of Memory and Language* **31**(6), 713–733. doi: 10.1016/0749-596x(92)90036-w

Mannes SM and Kintsch W (1987) Knowledge organization and text organization. *Cognition and Instruction* **4**(2), 91–115. doi: 10.1207/s1532690xci0402_2

Metzger RL, Warren AR, Shelton JT, Price J, Reed AW and Williams D (2008) Do children "DRM" like adults? False memory production in children. *Developmental Psychology* **44**(1), 169–181. http://dx.doi.org/10.1037/0012-1649.44.1.169

Pickering MJ and Branigan HP (1998) The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* **39**(4), 633–651. doi: 10.1006/jmla.1998.2592

Pickering MJ and Ferreira VS (2008) Structural priming: A critical review. *Psychological Bulletin* **134**(3), 427–459. doi: 10.1037/0033-2909.134.3.427

Pierce LJ, Klein D, Chen J-K, Delcenserie A and Genesee F (2014) Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences* **111**(48): 17314–17319. doi: 10.1073/pnas.1409411111

Potter MC and Lombardi L (1998) Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* **38**(3), 265–282. doi: 10.1006/jmla.1997.2546

Raney GE and Rayner K (1995) Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* **49**(2), 151–173. http://dx.doi.org/10.1037/1196-1961.49.2.151

R Core Team. (2018) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reyna VF (2004) How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science* **13**(2), 60–66. doi: 10.1111/j.0963-7214.2004.00275.x

Reyna VF (2012) A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. *Judgment and Decision Making* **7**(3), 332–359. Retrieved from http://journal.sjdm.org/11/111031/jdm111031.pdf

Reyna VF and Brainerd CJ (1995) Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences* **7**(1), 1–75.

Reyna VF and Lloyd FJ (2006) Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing.

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press

Journal of Experimental Psychology: Applied **12**(3), 179–195. doi: 10.1037/1076-898x.12.3.179

Reyna VF, Lloyd FJ and Brainerd CJ (2003) Memory, development, and rationality: An integrative theory of judgement and decision-making. In Schneider SL & Shanteau J (eds), *Emerging perspectives on judgement and decision research* (pp. 201–245). Cambridge, United Kingdom: Cambridge University Press.

Roelofs A (1992) A spreading-activation theory of lemma retrieval in speaking. *Cognition* **42**(1–3), 107–142. doi: 10.1016/0010-0277(92)90041-F

Roelofs A (1993) Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition* **47**(1), 59–87. doi: 10.1016/0010-0277(93)90062-Z

Rummer R and Engelkamp J (2001) Phonological information contributes to short-term recall of auditorily presented sentences. *Journal of Memory and Language* **45**(3), 451–467. doi: 10.1006/jmla.2000.2788

Rummer R, Schweppe J and Martin RC (2013) Two modality effects in verbal short-term memory: Evidence from sentence recall. *Journal of Cognitive Psychology* **25**(3), 231–247. doi: 10.1080/20445911.2013.769953

Sachs JS (1967) Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics* **2**(9), 437–442. doi: 10.3758/bf03208784

Sachs JS (1974) Memory in reading and listening to discourse. *Memory & Cognition* **2**(1), 95–100. doi: 10.3758/bf03197498

Sampaio C and Konopka AE (2013) Memory for non-native language: The role of lexical processing in the retention of surface form. *Memory* **21**(4), 537–544. doi: 10.1080/09658211.2012.746371

Singmann H, Bolker B, Westfall J, Aust F and Ben-Shachar M (2020) *afex: Analysis of Factorial Experiments*. R package version 0.26-0. https://CRAN.R-project.org/package=afex

Stanovich KE (2000) *Progress in understanding reading: Scientific foundations and new frontiers*. New York, NY: Guilford Press.

Tomasello M (2003) *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Ullman MT (2004) Contributions of memory circuits to language: The declarative/procedural model. *Cognition* **92**(1–2), 231–270. doi: 10.1016/j.cognition.2003.10.008

Ullman MT (2005) A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In Sanz C (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141–178). Washington, DC: Georgetown University Press.

Ullman MT (2016) The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In Hickock G & Small SL (eds), *Neurobiology of Language* (pp. 953–968). San Diego: Academic Press.

Ullman MT and Lovelett JT (2018) Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research* **34**(1), 39–65.

van Dijk TA and Kintsch W (1983) *Strategies of discourse comprehension*. New York, NY: Academic Press.

https://doi.org/10.1017/S1366728920000772 Published online by Cambridge University Press