Original Article

The International Journal of Bilingualism

Context effects in the L2: Evidence for compensatory mechanisms

International Journal of Bilingualism
1–37

DOI: 10.1177/13670069231153301
journals.sagepub.com/home/ijb
SAGE

Tal Norman and Tamar Degani
University of Haifa, Israel

Abstract

Aims and objectives: We examined how context is used to facilitate reading in the second language (L2) compared with the first language (L1), and how L2 availability and age modulate these context effects.

Methodology: Using self-paced reading, participants read high- and low-constraint Hebrew sentences. In Experiment 1, L1 ($n = 45$) and L2 ($n = 48$) Hebrew readers were compared, whereas in Experiment 2, only L2 readers ($n = 131$) were examined, testing modulations by L2 availability and age.

Data and analysis: Reading times of target, post target, and sentence final words were analyzed using linear-mixed-effects models.

Findings: In Experiment 1, L2 readers differed from L1 readers in contextual processing, as evident in the significant interaction between context type and language background on the final word measure. In Experiment 2, L2 readers with lower L2 availability scores differed from those with higher scores, and younger readers differed from older ones, in the way high- and low-constraining context affected their reading behavior in the target word and in the final word of the sentence. These differences were indicated by significant interactions between context type and L2 availability as well as between context type and age group. These findings are best understood under a compensatory processing account.

Originality: By complementing L1–L2 group comparisons with in-depth examination of the L2 profile, the current study reveals a continuous effect of L2 availability, such that a lower L2 availability is associated with a greater reliance on context. Furthermore, the inclusion of older and younger adults provides converging evidence to the use of contextual support as a compensatory mechanism when lexical processing is more effortful.

Keywords

Context effects, compensatory mechanism, L2 reading, predictability, older adults

Words are usually recognized within larger pieces of verbal information, and thus, their processing can be facilitated by a constraining or biasing context, especially in cases of lexical ambiguity

Corresponding author:
Tamar Degani, Department of Communication Sciences and Disorders, University of Haifa, 199 Abba Hushi, Haifa 3498838, Israel.
Email: tdegani@research.haifa.ac.il

(Morris, 2006). This facilitation may be explained either by top–down prediction processes or by bottom–up integration processes, which are often difficult to distinguish (Pickering & Gambi, 2018). Facilitation by prediction occurs when the context allows comprehenders to preactivate lexical or sublexical representations before they are activated by the verbal input itself, whereas facilitation by integration occurs when words are encountered after a semantically related context, which eases their integration within the existing representation of the text. Regardless of the mechanism involved, the ability to use previous context to facilitate the processing of words in sentences plays an important role in language comprehension. Thus, investigating this issue is of relevance to both first language (L1) and second language (L2) processing.

Evidence from studies testing L1 processing[1] suggests that word processing is faster after semantically constraining contexts, than after neutral contexts (e.g., Brothers & Kuperberg, 2021; Ehrlich & Rayner, 1981; Hess et al., 1995; Rayner & Well, 1996). For example, Brothers and Kuperberg (2021) reported that reading times in a self-paced reading (SPR) task were faster for words (e.g., glasses) embedded in a highly constraining context (e.g., Her vision is terrible and she has to wear *glasses* in class), than for words embedded in a more neutral context (e.g., Her mother was adamant that she has to wear *glasses* in class). Similarly, they found that responses in a picture naming task were faster when the pictured objects (e.g., glasses) were presented after the highly constraining context (e.g., Her vision is terrible and she has to wear ___). These findings indicate that previous context modulates lexical access both when comprehending and when producing L1 words.

Furthermore, numerous L1 studies examined how semantic context is used to resolve within-language semantic ambiguity in the case of words that consist of more than one meaning (i.e., homonyms like the word 'bank'; for a review see, Degani & Tokowicz, 2010). In particular, the influential "re-ordered access" model (Duffy et al., 1988) emphasizes the importance of previous semantic context in lexical ambiguity resolution. According to this model, context facilitates activation of the context-appropriate meaning of ambiguous words. For example, using eye-tracking methodology, Duffy et al. (1988) showed that under a biasing context, gaze durations on balanced ambiguous words that consist of two equally frequent meanings (e.g., pitcher), did not differ from their unambiguous controls (e.g., whiskey), suggesting that in such biasing contexts, previous semantic context facilitates the lexical processing of ambiguous words by boosting the activation of the context-appropriate meaning and reducing the competition between the two interpretations.

Critically, it is still unclear whether L1 and L2 comprehenders differ in the way they take advantage of contextual cues to process upcoming words (ambiguous and unambiguous ones). The few studies investigating this issue have provided conflicting evidence. Some findings suggest that L1 and L2 comprehenders similarly benefit from constraining context. For example, in an eye-movement study, Whitford and Titone (2017) examined word processing during L1 and L2 paragraph reading, as a function of word predictability (i.e., the degree to which a word can be predicted by previous context based on cloze probabilities) among younger and older French–English bilingual adults with relatively high proficiency in the L2 and varied current L2 exposure. They found that the effect of word predictability was language-invariant in both early-stage reading measures (taken to reflect lexical access) and late-stage reading measures (taken to reflect post-lexical integration), irrespective of age and current L2 exposure. Thus, a constraining context facilitated word reading to the same extent in both languages.

Similarly, Gollan et al. (2011) found that high-constraint contexts reduced the processing disadvantage of less frequent words, among both English monolinguals (L1) and English-dominant Spanish–English and Dutch-dominant Dutch–English proficient bilinguals (L2).

Along the same lines, in an event-related-potentials (ERP) study, Foucart et al. (2014) examined the processing of the article (masculine/feminine) preceding expected and unexpected target nouns,

as well as the processing of the noun itself. Specifically, Spanish monolinguals, French–Spanish-proficient late bilinguals, and Spanish–Catalan early bilinguals read Spanish (L1 or L2) sentences (e.g., The pirate had the secret map, but he never found **the** [masculine] **treasure/the** [feminine] **cave** he was looking for). The results showed the same context effects in the pattern of brain activity (i.e., an increase in the N400 amplitude during the processing of unexpected articles and nouns) in all three groups of Spanish readers, indicating that context-based lexical expectations were generated during both L1 and L2 sentence reading.

Conversely, other studies have demonstrated weaker, delayed, or different context effects in the L2, relative to the L1. For example, using a similar procedure to that of Foucart et al. (2014) with English sentences (e.g., He was very tired so he sat on **a chair/an armchair**), Martin et al. (2013) found that contrary to monolingual L1 readers, proficient L2 Spanish–English readers failed to show an increase in the N400 amplitude when processing unexpected articles and nouns during sentence reading. This finding suggests that L2 readers do not use the previous context to the same extent as L1 readers do (see also Mitsugi & MacWhinney, 2016).

In another ERP study that tested proficient Spanish–English L2 readers, Ito et al. (2017) found an increase in the N400 amplitude during the processing of sentences (e.g., The student is going to the library to borrow a ___) that ended with unpredictable and implausible target words ([TWs] e.g., sofa), relative to predictable and plausible TWs (e.g., book). However, they did not find N400 effects that depended on word predictability (i.e., the cloze values of the predictable words), for TWs that were form related (e.g., hook) or semantically related (e.g., page) to the predictable word. The authors concluded that although L2 readers are generally sensitive to sentence plausibility and can use previous context to facilitate lexical processing, they do not generate context-based predictions about the form and meaning of upcoming words, unlike monolingual L1 readers (as tested in Ito et al., 2016).

Moreover, in two response related (RT) and ERP priming studies, Elston-Güttler and Friederici (2005, 2007) examined how L1 and advanced L2 English readers process words with more than one meaning (i.e., homonyms) as a function of context. They reported that the two groups differed in the time-course of selecting the contextually appropriate meaning of these ambiguous words. Specifically, native English speakers with a minimal knowledge of German and late German–English bilinguals read English sentences that ended with a homonym. This homonym was displayed separately, after the rest of the sentence, for either 200, 500 (Elston-Güttler & Friederici, 2005), or 800 ms (Elston-Güttler & Friederici, 2007). Then, a TW that was related either to the contextually appropriate or inappropriate meaning of the homonym was displayed and participants had to decide whether or not it was a real word in English. At the 200-ms stimulus-onset asynchrony (SOA), they found an overall priming effect that was invariant to context condition and language group, in both the RT and N400 measures. Thus, at this early processing stage, both meanings of the homonym were active, in both language groups. At the 500-ms SOA, the RT measure revealed that contextually inappropriate meanings were no longer active for both groups. Yet, the N400 measure showed that activation of contextually inappropriate meanings had decayed for L1, but not for L2 readers. Finally, at the 800-ms SOA, both measures revealed that only contextually appropriate meanings were still active, in both language groups, indicating that at the final stage of disambiguation, the two groups were comparable in meaning selection and integration processes. The results of these two studies demonstrate that L2 readers employ similar processing mechanisms to those of L1 readers when dealing with lexical ambiguity resolution in context. However, lexical disambiguation using context seems to be slower in the L2 than in the L1.

Findings such as these can be interpreted within the framework of the *prediction-by-production* model (Pickering & Gambi, 2018). This model assumes that language comprehenders covertly imitate what they have comprehended so far from the utterance, and construct a representation of

the underlying communicative intention. Then, they run this intention through their production system, and thus, preactivate the upcoming utterance. Therefore, according to this model, the production system has a critical role in generating predictions during language comprehension (Pickering & Gambi, 2018). This prediction by production mechanism is considered to be an optional, non-automatic, and effortful process. With respect to L2 processing, this model postulates that L2 comprehenders are less likely to generate context-based predictions, because the reliance on the production system requires time and cognitive resources that may not be available during L2 processing (Ito & Pickering, 2021).

Yet, other findings suggest that L2 processing, especially among less proficient L2 users, may rely *more extensively* on contextual cues than L1 processing. For example, in a picture naming task, in which participants named pictures presented either in isolation, in low-constraint context, or in high-constraint context, Gollan et al. (2011) found that less proficient Dutch–English bilinguals (L2), but not highly proficient Spanish–English bilinguals (L2), benefited more from high-constraint context than did English monolinguals (L1). These findings suggest that reduced proficiency in the L2 might lead to a greater reliance on context. Similar evidence was also reported by Mor and Prior (2022). They tested Hebrew–English bilinguals with an intermediate level of English proficiency and demonstrated that the effect of word predictability, as indicated by TWs' total reading times (i.e., the sum of all fixation durations on the target word in a given trial), was more pronounced in participants' L2-English than in their L1-Hebrew.

Findings such as these are consistent with the *interactive-compensatory* model (Stanovich, 1980, 1984), which postulates that less skilled readers may rely more heavily on context to compensate for limited abilities in lexical processing or integration and to reduce overall cognitive effort (e.g., Ashby et al., 2005; Huettig & Brouwer, 2015). For instance, Ashby et al. (2005) found a significant context effect among average readers but not among highly skilled readers. Specifically, in highly constraining sentence contexts, in comparison to non-constraining contexts, only average readers were faster to process low-frequency words. Thus, the *interactive-compensatory* model may further predict that L2 readers, which are typically less skilled relative to L1 readers, are likely to generate context-based predictions in order to overcome other reading difficulties.

In sum, the existing literature is inconsistent with respect to whether L2 readers differ from L1 readers in their use of previous context to enhance online lexical processing. Dissimilarities across studies in the experimental settings (e.g., linguistic material, paradigm) may explain these distinct outcomes because they may tap different context-based processes, at different time points along the sentence, with varied sensitivity. Of relevance, these mixed results may also be explained by other modulating factors, including differences between participants in specific L2 characteristics such as proficiency, experience, L2 structure, and similarity to the L1. For instance, the study documenting larger context effects in the L2 than in the L1 (Mor & Prior, 2022) tested different-script Hebrew–English bilinguals, whereas all other studies focused on same-script Indo-European bilinguals (French–English in Whitford & Titone, 2017; Spanish–English in Gollan et al., 2011, Martin et al., 2013, and Ito et al., 2017; French–Spanish and Catalan–Spanish in Foucart et al., 2014; German–English in Elston-Güttler & Friederici, 2005, 2007; see Share, 2008 for discussion of Anglocentricities in reading research). Thus, further examination of potential L1–L2 differences in reliance on context is needed, especially with respect to additional populations.

The present study expanded this line of research in two important directions. First, it examined another population of L2 users—Arabic–Hebrew bilinguals—these two languages consist of similar orthographic principles and morphological characteristics but differ in script. These bilinguals are expected to adopt context-based reading strategies due to the specific features of the writing systems of both of their languages. Specifically, Hebrew and Arabic are usually written and read without diacritics that mark partial vowel information. In the absence of vowel

markers, the complete phonological form of words is not available from the script, such that the same orthographic form can refer to more than one phonological and semantic entry (i.e., heterophonic homographs; Abu-Rabia, 2001; Frost & Bentin, 1992). Thus, Hebrew and Arabic readers are routinely faced with the need to resolve lexical (phonological and semantic) ambiguity during reading and therefore may apply more extensively context-based processing strategies (Abu-Rabia, 1997; Bar-On et al., 2017). Furthermore, Arabic speakers may similarly employ such context-based processing strategies when reading in their L2 Hebrew, because they might transfer these reading strategies from their L1 to their L2 (Norman et al., 2016). Second, the present study examined reading behavior while treating bilingualism and L2 use as a continuum of experiences, rather than simply dichotomizing L1 from L2 readers. Thus, in order to capture large and continuous variability in L2 experience and age, which may modulate context effects in the L2, the second experiment reported here tested a relatively large and heterogeneous sample of L2 users and examined how individual differences in L2 availability might modulate reliance on context during reading.

The influence of language experience and proficiency on context effects *in* the L2

The language profiles of bilinguals are diverse and characterized by unique environments of language acquisition and use (Gullifer et al., 2021; Marian & Hayakawa, 2021), which may modulate cognitive abilities in general, and different language functions in particular. Indeed, recent studies have shown that variability in L2 performance may be explained by different aspects of L2 experience, including proficiency, use, exposure, context and age of acquisition (AoA), and interactional contexts, among other factors (e.g., Beatty-Martínez et al., 2020; Kastenbaum et al., 2019).

While the previous literature has greatly focused on the role of proficiency in shaping L2 processing, the current work has examined the role of L2 availability—the extent to which the language is available for use during online processing. The degree of L2 availability, which we define as a combination of different aspects of L2 experience and proficiency, as detailed below, may modulate the ability to use contextual information during sentence reading, in two alternative ways. First, less experienced L2 users are exposed to the target language less often and have less opportunities to use it. As such, their lexical representations are assumed to be of reduced quality (Gollan et al., 2008; Stanovich, 1980, 1984). Consequently, such individuals may rely on context more extensively to compensate for weaker lexical processing and for the reduced quality of lexical representations. Alternatively, individuals with reduced L2 experience and limited L2 proficiency may depend on context to a lesser extent than those with higher experience and proficiency, because they may have less available cognitive resources to devote to the demanding process of context-based prediction (Ito & Pickering, 2021).

Previous studies examining L2 lexical processing in context have reported that the effect of word predictability was not associated with L2 proficiency, as was measured by a vocabulary knowledge test and a single-word-reading fluency test (Mor & Prior, 2022), or by the percentage of L2 exposure time (Whitford & Titone, 2017). Yet, it was found that these two measures were related to the effect of word frequency (i.e., faster processing for more frequent words), such that greater L2 proficiency and exposure led to smaller frequency effects (Mor & Prior, 2022; Whitford & Titone, 2012, 2017 see also Gollan et al., 2011). These findings suggest that proficiency (i.e., lexical knowledge) and exposure may modulate lexical processing but not context-based processing during L2 reading. Mor and Prior (2022) suggested that the two proficiency measures used in their study (i.e., vocabulary knowledge and reading fluency) did not explain the effect of word predictability because these measures capture lexical knowledge that is not sufficient to support top–down context-based processes, such as prediction or integration, during reading.

Thus, in order to find a more suitable measure that may better explain context effects in the L2, here we created an L2-Availability Factor that accounted for participants' variability in verbal fluency, use, exposure, and AoA. This measure captures different aspects of the L2 experience, and as such, may better represent the degree to which linguistic knowledge and cognitive resources are available for top–down context-based processing during online L2 reading.

The influence of age on context effects in the L2

In addition to the possible influence of L2 experience and proficiency, age may also be an important factor that could modulate context effects in the L2. On one hand, older adults have accumulated larger L2 vocabulary knowledge and greater reading experience than younger adults (e.g., Stanovich et al., 1995), and thus, may be more efficient in using previous context during L2 sentence reading. On the other hand, a wide range of sensory abilities and cognitive skills necessary for dealing with complex tasks decline with age (e.g., Bopp & Verhaeghen, 2005; Fozard & Gordon-Salant, 2001; Park & Reuter-Lorenz, 2009; Salthouse, 2010). Thus, according to the *interactive-compensatory* model discussed above, such declines may necessitate the operation of context-based processing mechanisms in order to compensate for slower processing abilities.

Studies investigating the influence of age on L1 sentence reading show that, in general, reading is slower for older (65+ years) than younger (18–30 years) adults. Specifically, older adults tend to have more and longer fixations, higher skipping rates, longer saccades, and more regressions back to earlier regions (e.g., Kemper et al., 2004; Kemper & Liu, 2007; Kliegl et al., 2004; Rayner et al., 2006; Whitford & Titone, 2016 for a review see Gordon et al., 2016). Furthermore, older adults tend to show reduced lexical quality and accessibility (indexed by a greater word frequency effect) and more cross-language activation (indexed by a greater cross-language neighborhood density effect) in both their L1 and L2 (Whitford & Titone, 2017). Importantly, age was also found to influence context effects during sentence reading, however, previous studies have reported inconsistent findings as to the direction of this age-related effect in the L1, and this issue has rarely been studied in the L2.

Findings from eye-movement studies suggest that older, relative to younger, adults rely more heavily on previous context to facilitate lexical identification and integration (Choi et al., 2017; Rayner et al., 2006; Zhao et al., 2019, 2021). For example, Choi et al. (2017) compared eye-movement measures of younger (19–25 years) and older (67–80 years) adults during L1 sentence reading. Sentences consisted of either a predictable or an unpredictable target word (e.g., The doctor told Fred that his drinking would damage his **liver/heart** very quickly). They found stronger context effect for older, than younger adults, indexed by shorter reading times for predictable than unpredictable TWs. Similarly, Zhao et al. (2019, 2021) found that reading time measures, sensitive to both lexical identification (i.e., gaze duration) and contextual integration (i.e., regression-path reading times), were faster for predictable, relative to unpredictable TWs, and that this context effect in the L1 was larger among older adults than among younger adults.

In contrast to these behavioral findings, evidence from ERP studies, which may reflect neural efficiency and resource allocation, suggests that older adults show weaker and/or delayed context effects in the L1 (Federmeier et al., 2003; Federmeier & Kutas, 2005; Payne & Federmeier, 2018; Wlotko et al., 2012). For example, Federmeier and Kutas (2005) found effects of contextual constraint on brain responses to sentence-final words (FWs). Specifically, N400 amplitudes were reduced when the same words (e.g., beard) were presented at the end of high-constraint sentences (e.g., No one at the reunion recognized Dan because he had grown a **beard**) as compared with low-constraint sentences (e.g., At the children's park next to the beach she saw a man with a **beard**).

Critically, for older adults, this context effect was smaller and later than that observed for younger adults, suggesting that older adults can use constraining sentence-level information, but not as quickly and effectively as younger adults do.

One way in which these behavioral and neural findings can be reconciled is by dissociating lexical prediction per se from other contextual facilitation that is independent from prediction processes, such as integration. In the ERP study of Dave et al. (2018), participants read two-sentence passages and were instructed to use the context of both sentences to predict the passage FW. After reading the passage, they had to indicate whether the passage FW matched the word they had predicted. The authors reported no age-difference both in the proportion of accurately predicted passage-final-words and in the neural benefits for correct, relative to incorrect predictions. However, age-related reductions were observed for the N400 effect of contextual support (i.e., constraining vs. un-constraining sentences) that were independent of prediction accuracy. These findings suggest that even though aging may not result in a specific decline in the predictive ability, it still has influence on the ability to use constraining context to facilitate lexical processing.

Most relevant to the current study, in the one study that examined the effect of age on both L1 and L2 paragraph reading, Whitford and Titone (2017) reported that the effect of word predictability, in both languages, was age-invariant across both early- and late-stage eye-movement reading measures. This pattern suggests that older and younger adults comparably use contextual cues to predict upcoming words during reading. Yet, to the best of our knowledge, this is the only study investigating the influence of age on context-based lexical processing in the L2, and it was conducted with same-script (French–English) bilinguals. Thus, to reach stronger conclusions, further examination of this issue in other bilingual populations is needed.

The present study

As discussed above, little is known about context-based reading processes in the L2 and about the role of L2 experience and age in shaping these processes. Therefore, the two main goals of the current study were (1) to compare the way L1 and L2 readers use previous semantic context to facilitate lexical processing of ambiguous and unambiguous words during sentence comprehension; and (2) to further examine the influence of L2 availability and age on L2 readers' ability to benefit from constraining context.

To this end, we conducted two experiments using Hebrew sentences in conjunction with the SPR task (Jegerski, 2014; Marsden et al., 2018). In the task, participants read high- and low-constraint sentences that included either ambiguous (i.e., homonyms) or unambiguous TWs, which were never positioned at the end of the sentence. To achieve the first goal, in Experiment 1, we compared the reading times of L1 and L2 Hebrew readers. To achieve the second goal, in Experiment 2, we focused on L2-Hebrew readers and calculated a composite score of five measures that capture different aspects of participants' variability in the L2.

Predictions were as follows. If L1/L2 readers use previous semantic context to facilitate lexical processing, then word reading should be faster in high- than in low-constraint contexts. Moreover, in both language groups, context effects are expected to be modulated by the lexical ambiguity of TWs, such that a constraining context should facilitate processing to a greater extent in the case of ambiguous homonyms, than in the case of unambiguous controls, because of its contribution to disambiguation processes (Duffy et al., 1988). Furthermore, based on the *interactive-compensatory* model (Stanovich, 1980, 1984) described above, we hypothesized that context effects would be more pronounced in the L2 relative to the L1, in less experienced L2 readers than in more experienced ones, and in older adults relative to younger adults.

Table 1. Participants' characteristics as a function of L1 group in Experiment 1; *M* (*SD*).

Measure	L1-Arabic	L1-Hebrew
N	48	45
Males/females	16/32	21/24
Age (in years)*	23.3 (3.7)	25.7 (3.6)
Education (in years)	14.5 (1.7)	14.5 (2.3)
Maternal education (in years)*	12.1 (3.6)	15.2 (3.0)
Hebrew age of acquisition (in years)*	7.6 (1.0)	0.0 (0.0)
Hebrew current exposure (%)*	31.3 (16.9)	85.3 (10.3)
Hebrew current use (0–10)*	6.8 (1.5)	8.1 (1.3)
Hebrew subjective proficiency (0–10)*	8.6 (0.9)	9.6 (0.6)
Hebrew semantic fluency*	21.8 (7.9)	36.2 (8.7)
Arabic current exposure (%)*	48.5 (21.7)	0.4 (1.7)
Average Arabic current use (0–10)	6.3 (1.9)	–
Overall Arabic current use (0–10)	–	0.7 (1.3)
Arabic subjective proficiency (0–10)	9.6 (0.6)	–
Overall Arabic subjective proficiency (0–10)	–	1.0 (1.3)
Arabic semantic fluency	29.3 (6.7)	–

Note. An asterisk marks a significant difference between the two L1 groups at the .05 level based on an independent-sample *t* test. Current exposure is a self-estimate of the percentage of time, out of 100%, of current exposure to each language. Current use is the mean score of the self-rated level of current use in speaking, writing, reading, internet, listening to music/radio, and watching TV/movies, on a scale of 0—*the lowest level of use*—to 10—*the highest level of use*—in each language. Subjective proficiency is the mean score of the self-rated proficiency in speaking, writing, reading, and spoken language comprehension, on a scale of 0—*the lowest level of ability*—to 10—*the highest level of ability*—in each language. In the L1-Hebrew group, Overall Arabic current use and Overall Arabic subjective proficiency are reported, because in this group only global ratings were collected to verify that participants do not have significant knowledge in Arabic. Semantic fluency is the score on a semantic fluency test (Kavé, 2005).

Experiment 1: L1–L2 differences

Method

Participants. A total of 93 students (ages 18–35; 37 males) with normal or corrected to normal vision and no learning or hearing disabilities participated in Experiment 1. Of these, 48 were L1-Arabic speakers who have learned Hebrew as their L2 starting in the third grade. They were relatively proficient in Hebrew and at least partially immersed in a Hebrew-speaking environment at the time of testing, as they were all learning at a Hebrew-speaking university. The remaining 45 participants were L1-Hebrew speakers with minimal knowledge of Arabic. Five additional participants were excluded due to learning ($n=2$) or hearing ($n=1$) disabilities, because of a technical error during task administration ($n=1$), or because of exposure to another language at home ($n=1$). All participants were also relatively proficient in English as this is a language that is formally learned in Israeli schools. They all signed an informed consent approving their participation in the current study. Participants' characteristics as a function of L1 group (L1-Hebrew/L1-Arabic) are summarized in Table 1 based on their self-report ratings on a language history questionnaire (a modified version of the LeapQ, Marian et al., 2007), and on their performance on an objective proficiency measure (semantic fluency test, Kavé, 2005). As seen in Table 1, the two groups differed on age and socio-economic status (SES; as indexed by maternal education), which were therefore included as covariates in the analyses.

Stimuli

TWs. The critical TWs were 15 Hebrew homonyms (selected from Peleg et al., 2012), which are words corresponding to two meanings (e.g., the word "מפה" /mapa/ which means either

Table 2. Target word characteristics as a function of target type in Experiment 1; *M* (*SD*).

Measure	Homonyms	Controls
Number of items	15	15
Word length (in letters)	3.47 (.74)	3.47 (.99)
Word length (in syllables)	2.00 (.85)	2.00 (.76)
Word frequency	36.40 (28.38)	25.47 (29.87)

Note. The two target types did not significantly differ ($p < .05$) in all four measures based on independent-sample *t* tests.

Table 3. Examples of critical sentences as a function of target and context type in Experiment 1.

Target	Context	Word		Sentence
Homonym	Low constraint	Hebrew	/mapa/	מצאתי בארון מַפָּה שקניתי בטיול בספרד לפני חמש שנים.
		English[a] translation	Map/ tablecloth	I found in the closet a map that I bought on a trip to Spain five years ago.
	High constraint	Hebrew	/mapa/	לפני שהמציאו את ה'וויז' אנשים ניווטו את הדרך על ידי מצפן ומַפָּה של האזור.
		English translation	map	Before Waze was invented people navigated the way using a compass and a map of the area.
Control	Low constraint	Hebrew	/pri/	בכל פעם שאני חוזר הביתה אני אוהב לאכול פֶּרִי טרי מהסלסלה.
		English translation	fruit	Every time I come home I like to eat a fresh[b] fruit from the basket.
	High constraint	Hebrew	/pri/	כמו התפוח, גם התפוז הוא פֶּרִי מזין שניתן לקטוף מהעצים.
		English translation	fruit	Like the apple, the orange is also a[b] nutritious fruit that can be picked from trees.

[a]English translations were never presented during the experiment.
[b]In Hebrew, the adjective follows the noun, and thus does not create a biasing context.

a map or a tablecloth). Furthermore, for each homonym, an unambiguous control word was selected (e.g., the word "פרי" /pri/ which means a fruit). Control words were matched to homonyms in length (i.e., number of letters and syllables) and frequency (based on HebWaC corpus via SketchEngine; Kilgariff et al., 2010, 2014). See Table 2 for TW characteristics as a function of target type. In addition, 15 false cognate words (i.e., Hebrew words that are phonologically similar to Arabic words but differ in meaning) and their 15 matching control words served as fillers for current purposes.

Sentences. Two sentences were constructed for each TW (for the homonyms and their matching controls, as well as for the false cognates and their matching controls that served as fillers), one creating a highly constraining semantic context and the other creating a neutral semantic context. High constraint sentences were always biased toward the dominant meaning of the homonyms (determined based on Peleg et al., 2012). This was initially done to allow comparison with targets in the false cognate condition. However, these comparisons were eventually considered beyond the scope of the current study, in which the false cognate items were treated as fillers. In all sentences, at least one content word appeared before and after the TW. In addition, 20 filler sentences with no ambiguous words were created to be followed by a corresponding yes/no comprehension question. See Table 3

Table 4. Sentence characteristics as a function of target and context type in Experiment 1; *M* (*SD*).

Target	Context	
	High constraint	Low constraint
Homonyms		
Target predictability*	0.79 (0.21)	0.00 (0.00)
Target location in sentence*	8.73 (1.91)	4.93 (2.25)
Sentence length (number of words)*	10.93 (1.98)	8.47 (1.92)
Sentence length (number of characters)*	57.33 (11.23)	45.13 (9.72)
Controls		
Target predictability*	0.75 (0.19)	0.00 (0.00)
Target location in sentence*	9.33 (2.77)	5.47 (2.26)
Sentence length (number of words)*	11.93 (2.58)	8.00 (2.07)
Sentence length (number of characters)*	64.2 (11.52)	43.33 (11.43)

Note. An asterisk marks a significant difference between the two context types at the .05 level based on a repeated measures analysis of variance with the Bonferroni corrections for multiple comparisons. Importantly, in all measures, in both the high- and low-constraint contexts, the two target types did not significantly differ based on independent-sample *t* tests.

for examples of critical sentences as a function of target and context type and Appendix 1 for the full set of critical stimuli in Experiment 1.

Semantic context in the high- and low-constraint sentences was determined based on cloze probability norms collected from a group of 20 native Hebrew speakers, who did not participate in the main experiment. Two versions of an online questionnaire were created, such that each participant saw only one version with either the high- or low-constraint sentence for a given TW. Each sentence was truncated before the TW and was completed by 10 participants. Based on these norms, word predictability for TWs was significantly higher in the high- than in the low-constraint context, for both homonyms and controls (see Table 4 for sentence characteristics). Note that high- and low-constraint sentences, for each target type, differed not only in the predictability of the TW, but also in length (number of words/characters) as well as in the location of the TW in the sentence. These factors were therefore included as covariates in the analyses.[2]

Two versions of the stimuli set were created, such that each included a total of 80 sentences, 30 presenting a high-constraint semantic context, 30 presenting a low-constraint semantic context, and 20 filler sentences with no ambiguous words, which were followed by a yes/no comprehension question. Each version included 30 critical sentences with either the homonyms (15) or their matching controls (15), and 30 sentences with either false cognates (15) or their matching controls (15) that served as fillers for current purposes (as well as 20 sentences followed by comprehension questions). These two versions were counterbalanced across participants.

Procedure. L1-Hebrew and L1-Arabic participants were recruited and tested by native Hebrew- or Arabic-speaking experimenters, respectively, but task instructions were given in Hebrew to all participants. Participants first completed the SPR task that was followed by a semantic fluency test in Hebrew and a post-test verifying their familiarity with the two Hebrew meanings of the ambiguous homonyms. L1-Arabic participants then completed in addition a semantic fluency test in Arabic. Finally, all participants completed the language history questionnaire (adapted from Marian et al., 2007).

SPR task. Following the typical protocol for the SPR task (e.g., Prior et al., 2017; Tokowicz & Warren, 2010), participants were instructed to silently read Hebrew sentences, presented word by

word on the computer screen, and advance through the words at their own pace by pressing a button. Reading times per word were thus measured by the computer program (E-prime, Psychology Software Tools, Pittsburgh, PA). On each trial, a fixation cross appeared at the center of the screen until participants pressed a button, at which point the sentence was presented centrally, one word at a time, through which participants advanced by button press. The last word of each sentence appeared with a period to mark the end of the sentence. Following 25% of the sentences (20 filler sentences), a yes/no comprehension question was presented to verify reading for comprehension. Four practice sentences preceded the experimental sentences. Participants had the opportunity to take a short break after 40 sentences, or whenever the fixation cross appeared on the screen.

Semantic fluency test. In this test, participants were asked to produce out loud as many words as they could within 1 minute, for each of two different semantic categories, within a given language (Gollan et al., 2002; Kavé, 2005). Categories in Hebrew, administered to both L1-Hebrew and L1-Arabic participants, included (1) fruits and vegetables and (2) vehicles, whereas Arabic categories, administered only to the L1-Arabic participants, included two different categories—(1) animals and (2) clothing. Following the category name, a 1-minute hourglass was presented on the screen to mark the time left for production, and responses were recorded for later coding of accuracy.

Post-test. To verify participants' familiarity with the two Hebrew meanings of the homonyms, all participants were presented with a list of words, in which each of the 15 homonyms was presented twice, each time with only one meaning, written in English for the L1-Hebrew participants or in Arabic for the L1-Arabic participants. Participants were to mark unfamiliar Hebrew words or meanings. In addition, for the L1-Arabic participants, this post-test also included another list of Hebrew false cognate words not analyzed here.

Analysis approach. To evaluate the influence of previous context on word reading at different points along the sentence, we analyzed the reading times of the TW, reflecting the immediate influence of context on lexical processing, the post target word (PTW), reflecting spillover effects on the processing of the next word, and the sentence-FW, reflecting wrap-up effects of comprehension processes (e.g., semantic integration) that cannot be executed immediately, and thus, are postponed until the end of the sentence (Just & Carpenter, 1980; Just et al., 1982; Tokowicz & Warren, 2010). Thus, three separate analyses were performed for these three different reading time measures.

For each analysis, a maximal linear-mixed-effect (LME) model was submitted to the *buildmer* function in the "buildmer" package (v. 2.2, Voeten, 2021) in R (v. 4.0.3, R Core Team, 2020), which uses the *lmer* function from the "lme4" package (v. 1.1.-21, Bates et al., 2014). In these models, in addition to random intercepts, random slopes justified by the design were included to account for the possible variability of participants and/or items in their sensitivity to the experimental manipulations (Barr et al., 2013). Starting from the maximal model, and using backward-fitting model selection procedure, the *buildmer* function systematically simplifies the random slopes until convergence, in addition to using likelihood ratio tests, to examine the contribution of random slopes to the fit of the model (one of the common methods to test model fit; Matuschek et al., 2017, p. 308). Note that this systematic selection procedure is not based on decisions made by the researcher and is fully replicable from the data, resulting in a model that is not too complex to be supported by the data (Bates et al., 2015). In addition, the *buildmer* function tests the contribution of each fixed effect to the model fit via a chi-square test on the residual sum of squares of each model. To obtain these estimations, we employed the *include=as.formula* argument to always keep the critical fixed effects and interactions in the model.

The maximal LME model in each analysis included (1) the fixed effects of the control variables: Age, Maternal Education, Target Location (i.e., the TW number in the sentence), Sentence Length (i.e., the number of words in the sentence), Word Length (i.e., the number of letters in either the TW, PTW, or FW), and Word Frequency (of either the TW, PTW, or FW), which were all continuous and normalized; (2) the fixed effects of the variables of interest (dummy coded): Context (Low/High, with High as the reference), Target Type (Homonym/Control, with Control as the reference), Group (L1-Hebrew/L1-Arabic, with L1-Arabic as the reference), and the interactions among them; and (3) the random effects of Participants and Items with by-participant and by-item intercepts, by-participant slopes for Context and Target Type, and by-item slopes for Context and Group (see Barr et al., 2013). The selected models for the TW, PTW, and FW analyses are presented in Table 5.

The selected model in each analysis was refitted using the lmer function, and *p*-values for all fixed effects and interactions were determined using the *anova* function from the "stats" package (v. 4.0.3, R Core Team, 2020), which calculate a Type III ANOVA table with Satterthwaite's method. In addition, interactions and pairwise comparisons were tested using the testInteraction function from the "phia" package (v. 0.2-1, De Rosario-Martinez, 2015), which computes Chi-square test with Bonferroni adjustments for multiple comparisons—these are presented in the text. Model summaries (obtained from the summary function) are presented in Table 5. Note that the fixed effects presented in this table reflect simple effects (e.g., the coefficient for Context reflects its effect at the reference level of the other factors, namely, the L1-Arabic group and the Control targets) rather than main effects collapsing across all levels.

Results

RTs were log transformed, since examination of the RT distribution revealed substantial deviation from normality, and log transformation was reported to be the best remedy when using the SPR task, as it generally makes the distribution acceptable for statistical analyses without eliminating or alternating potentially legitimate data points (Nicklin & Plonsky, 2020). Indeed, log transformation improved the QQ plot, skew (raw RT = 37.03; log RT = 0.625), and kurtosis (raw RT = 3346.71; log RT = 2.33) of the distribution.

Across all three measures, there was a main effect of Group: TW: $F(1) = 31.94$, $p < .001$; PTW: $F(1) = 35.74$, $p < .001$; FW: $F(1) = 36.20$, $p < .001$. Namely, the L1-Hebrew participants responded faster than the L1-Arabic participants, irrespective of the other variables. Critically, within the FW selected model, the two-way interaction between Group and Context was significant, $F(1) = 5.09$, $p = .02$, indicating that there was a significant difference between L1-Arabic and L1-Herbew participants in the pattern of response to high- and low-constraining context (Figure 1). However, note that the simple effects of Context, namely, the RT-difference between High- and Low- constraint sentences within each level of the L1 Group, did not reach significance either for L1-Herbew participants, $\chi^2(1) = 0.37$, $p = 1.00$, or for L1-Arabic participants, $\chi^2(1) = 3.32$, $p = .14$.

In sum, in Experiment 1, although the difference between the two context conditions did not reach significance in either group, the significant interaction between Context and Language Group may suggest differential contextual processing as a function of language background. We return to this issue in Experiment 2. In addition, in all three models the effect of Target Type was not significant, and this factor did not significantly interact with Context, suggesting that lexical ambiguity had no influence on reading times and did not modulate the context effect, in contrast to our prediction (see Table 5).

Table 5. The three LME models in Experiment 1 predicting RTs for the target word, post target word, and final word. Effect size (*b*), standard errors (*SE*), and *t* value (*t*).

Fixed effects	Target word			Post target word			Final word		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	6.30	.04	139.51***	6.33	0.05	128.20***	6.42	0.05	129.67***
Context (low)	0.03	0.03	0.78	0.01	0.04	0.34	0.04	0.03	1.09
Group (L1-Hebrew)	–0.34	0.06	–5.63***	–0.32	0.06	–5.18***	–0.33	0.07	–4.89***
Target Type (Hom)	0.04	0.04	1.08	0.06	0.04	1.44	–0.03	0.04	–0.79
Group (L1-Hebrew): Target Type (Hom)	–0.02	0.03	–0.58	–0.03	0.03	–0.86	–0.02	0.05	–0.43
Context (low): Target Type (Hom)	–0.05	0.05	–1.04	0.01	0.05	0.28	0.02	0.05	0.37
Context (low): Group (L1-Hebrew)	–0.03	0.04	–0.89	–0.02	0.03	–0.64	–0.11	0.04	–2.97**
Context (low): Group (L1-Hebrew): Target Type (Hom)	0.06	0.05	1.22	–0.01	0.05	–0.22	0.09	0.06	1.69~
Control variables	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
Word length	0.02	0.01	3.15**	0.05	0.01	4.76***	0.05	0.01	4.70***
Random effects	*Variance*	*SD*		*Variance*		*SD*	*Variance*		*SD*
Participant (intercept)	0.07	0.26		0.08		0.28	0.08		0.28
Context (low)	0.01	0.09		–		–	–		–
Item (intercept)	0.00	0.07		0.01		0.09	0.01		0.08
Context (low)	0.01	0.08		0.01		0.10	0.01		0.09
Group (L1-Hebrew)	–	–		–		–	0.01		0.08
Residual	0.10	0.31		0.10		0.31	0.11		0.34

Note. Fixed effects reflect simple effects relative to the reference level, when other factors are at their reference level, without correction for multiple comparisons. For main effects see *F* values in the text.
Sig. codes. 0 "***" 0.001 "**" 0.01 "*" 0.05 "~" 0.1.

Experiment 2: the influence of L2-availability and age on context effect in the L2

Experiment 1 compared L1 and L2 processing and revealed potential differences in contextual processing between the two language groups. However, as the context effect did not reach significance in either group, it is possible that within-group variability obscured this effect, especially for L2 readers. Indeed, as is evident in Table 1, L1-Arabic participants varied greatly in their patterns of L2 proficiency and use. As alluded to above, readers who are exposed to and use the L2 less often than others, may have reduced access to L2 lexical representations. As such, they may rely on contextual support to a greater extent than those with higher L2 accessibility. Alternatively, individuals with higher L2 experience may have more cognitive resources available to allocate to the demanding process of reliance on context.

Therefore, Experiment 2 adopted a complementary approach of testing L2 variability as a continuous factor. To this end, we tested a considerably larger sample of L2 readers, which increased the variability among participants relative to Experiment 1. This sampling approach also increased variability in participants' age, which may also modulate reliance on contextual support. Therefore, Experiment 2 allowed examination of whether L2 availability and age could explain contextual sensitivity in L2 reading.

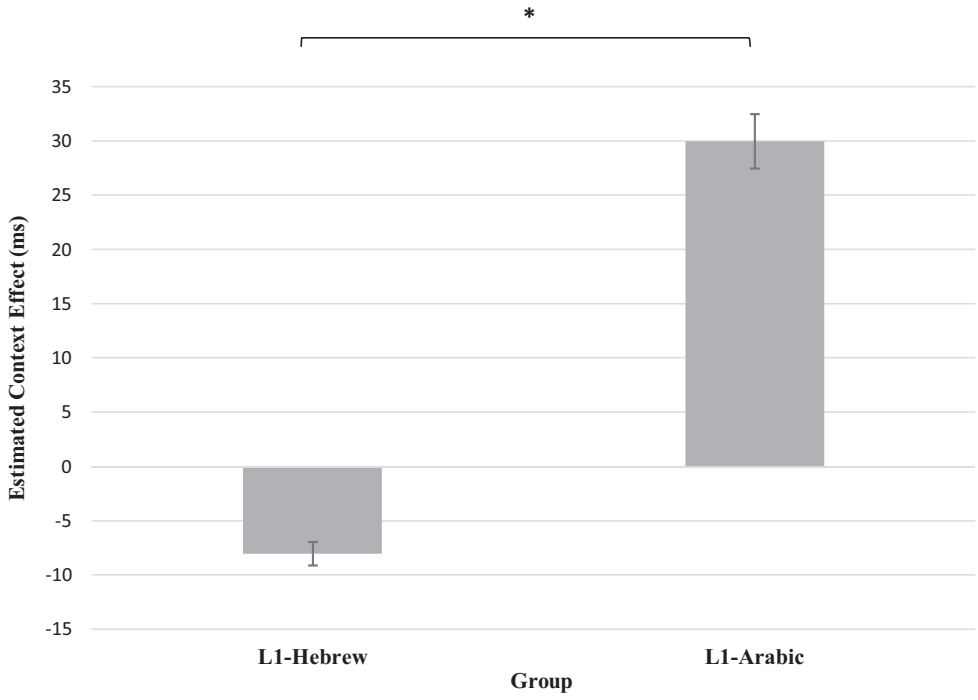

Figure 1. Estimated context effect by group in the final word analysis of Experiment 1.
Note. Estimated Context Effect is the difference in the estimated marginal mean RTs (ms) between Low- and High-constraint sentences. Error bars mark *SE*.
*$p < .05$.

Method

Participants. A total of 131 participants (ages 16–72; 0 males) with normal or corrected to normal vision and no learning or hearing disabilities took part in Experiment 2. They were all native Arabic speakers who have learned Hebrew as their L2 but varied in age and in their Hebrew experience. They were recruited by native Arabic speakers from Arabic-speaking communities, with no restriction on the gender of the recruited participants. This rather heterogeneous sample was selected as such to increase variability in L2 availability and age, as well as to increase ecological validity and avoid reliance on WEIRD samples only (Henrich et al., 2010). Seven additional participants were excluded due to learning disabilities ($n=5$) or due to missing data (i.e., semantic fluency scores in Arabic and Hebrew; $n=2$). All participants signed an informed consent approving their participation in the study. Participants' characteristics are summarized in Table 6 based on their self-report ratings on a language history questionnaire (a modified version of the LeapQ, Marian et al., 2007) and on their performance in the L1-Arabic and L2-Hebrew objective proficiency measures (i.e., semantic fluency test, Kavé, 2005). To control for potential differences in L1 verbal abilities within this sample of Arabic-Hebrew bilinguals, the semantic fluency score in the L1-Arabic was included as a covariate in the analyses.

Stimuli

TWs. The critical TWs were similar to those used in Experiment 1, except that 2 homonyms and 1 control word were replaced because they partially overlapped phonologically across languages

Table 6. Participants' characteristics in Experiment 2; *M* (*SD*).

Measures		
N	131	
Males/females	0/131	
Age (in years)	28.8 (15.8)	
Education (in years)	13.9 (2.7)	
Maternal education (in years)	11.9 (4.6)	
Language	Arabic (L1)	Hebrew (L2)
Age of acquisition (in years)*	0 (0)	7.2 (2.0)
Current exposure (%)*	60.1 (17.5)	36.9 (21.0)
Current use (0–10)*	6.1 (1.9)	5.4 (2.2)
Subjective proficiency (0–10)*	9.3 (1.1)	8.3 (1.6)
Semantic fluency*	31.5 (7.5)	20.4 (6.7)

Note. An asterisk marks a significant difference between the two languages at the .05 level based on a paired-sample *t* test. See the note in Table 1 for details on what each measure represents.

Table 7. Target word characteristics as a function of target type in Experiment 2; *M* (*SD*).

Measure	Homonyms	Controls
Number of items	18	36
Word length (in letters)	3.72 (1.02)	3.58 (0.91)
Word length (in syllables)	2.22 (0.81)	2.08 (0.77)
Word frequency	26.94 (15.40)	27.95 (34.06)

Note. The two target types did not significantly differ in all four measures based on independent-sample *t* tests. As in Experiment 1, frequency estimates were calculated based on HebWaC corpus via SketchEngine (Kilgarriff et al., 2010, 2014).

(i.e., could be considered Hebrew–Arabic cognates), and an additional 3 homonyms and 3 control words were added, resulting in 18 homonyms and 18 unambiguous controls that were matched in length and frequency to the set of homonyms, as in Experiment 1. In addition, 18 false cognate words served as fillers for current purposes, and their 18 control words served as additional control words, since they were also matched in length and frequency to the current set of homonyms. See Table 7 for TW characteristics as a function of target type.

Sentences. As in Experiment 1, two sentences were constructed for each TW, one creating a highly constraining semantic context and the other creating a neutral semantic context. Some of the sentences were identical to those used in Experiment 1, some were slightly changed, and some were new. These were constructed similarly to the sentences in Experiment 1. In all sentences, at least two content words appeared before and after the TW. In addition, 24 filler sentences with no ambiguous words were used to be followed by a corresponding yes/no comprehension question. See Table 3 for examples of critical sentences as a function of target and context type and Appendix 2 for the full set of critical stimuli in Experiment 2.

As in Experiment 1, semantic context in the high- and low-constraint sentences was determined based on cloze probability norms, collected from a new group of 20 native Hebrew speakers, who did not participate in the main experiment. Based on these norms, word predictability for TWs was

Table 8. Sentence characteristics as a function of target and context type in Experiment 2; *M* (*SD*).

Target	Context	
	High constraint	Low constraint
Homonyms		
Target predictability*	0.89 (0.32)	0.00 (0.00)
Target location in sentence*	7.61 (1.94)	6.06 (2.13)
Sentence length (number of words)	10.78 (1.77)	10.17 (1.95)
Sentence length (number of characters)	56.11 (10.31)	54.78 (9.77)
Controls		
Target predictability*	0.80 (0.41)	0.00 (0.00)
Target location in sentence*	8.36 (2.61)	5.36 (1.81)
Sentence length (number of words)	10.81 (2.36)	10.03 (1.46)
Sentence length (number of characters)*	59.53 (12.17)	54.72 (6.98)

Note. An asterisk marks a significant difference between the two context types at the .05 level based on a repeated measures ANOVA with the Bonferroni corrections for multiple comparisons. Importantly, in all measures, in both the high- and low-constraint contexts, the two target types did not significantly differ based on independent-sample *t* tests.

significantly higher in the high- than in the low-constraint context for both homonyms and controls (see Table 8 for sentence characteristics). Note that high- and low-constraint sentences, for each target type, differed not only in the predictability of TWs, but also in length (number of words/characters), as well as in the location of the TW in the sentence. Therefore, these factors were included as covariates in the analyses.[3]

Two versions of the stimuli set were created, such that each included a total of 96 sentences, 36 presenting a high-constraint semantic context, 36 presenting a low-constraint semantic context, and 24 filler sentences with no ambiguous words, which were followed by a yes/no comprehension question. Each version included 54 critical sentences with either homonyms (18) or their matching controls (36) and 18 sentences with false cognates that served as fillers for current purposes. These two versions were counterbalanced across participants.

Procedure. The procedure was identical to the one employed in Experiment 1.

Analysis approach

L2 profile and age. Experiment 2 focused on the influence of participants' L2 profile and age on the effects of context in the L2. Based on the distribution of age, we first created a categorical variable of Age Group (Younger = 16–35; Older = 58–72), which resulted in 108 Younger participants and 23 Older participants. Then, we examined the Pearson correlations among the 5 different L2-Hebrew measures that were collected. These included AoA, Current exposure (self-estimate of the percentage of time, out of 100%, of current exposure to each language); Current use (mean score of the self-rated level of current use in speaking, writing, reading, internet, listening to music/radio, and watching TV/movies, on a scale of 0—*the lowest level of use*—to 10—*the highest level of use*—in each language); Subjective proficiency (mean score of the self-rated proficiency in speaking, writing, reading, and spoken language comprehension, on a scale of 0—*the lowest level of ability*—to 10—*the highest level of ability*—in each language); and Semantic Fluency (the score in a semantic fluency test; Kavé, 2005). As seen in Table 9, there were substantial correlations across the five measures.

Table 9. Pearson correlations among the L2-Hebrew measures.

	L2 measures	1	2	3	4	5
1	Hebrew AoA	–				
2	Hebrew current exposure	–.07	–			
3	Hebrew current use	–.16	.35**	–		
4	Hebrew subjective proficiency	–.21*	.39**	.49**	–	
5	Hebrew semantic fluency	–.17	.18*	.24**	.35**	–

Sig. codes. 0 "***" 0.001 "**" 0.01 "*" 0.05.

Table 10. Loadings of the collected L2 measures on the L2-Availability Factor based on a PCA.

L2-Hebrew measures	Loadings
AoA	–.39
Current exposure	.64
Current use	.74
Subjective proficiency	.81
Semantic fluency	.58
Unique variance explained (%)	42

Note. PCA: principal component analysis; AoA: age of acquisition.

Thus, to capture shared variance among the predictors in order to better represent bilingual experience, we applied a principal component analysis (PCA) to the data. Indeed, the Kaiser–Meyer–Olkin measure of sampling adequacy of 0.72 and Bartlett's test of sphericity ($p < .001$) indicated that the predictors were highly correlated, suggesting that a PCA was warranted for this dataset. We extracted factors with eigenvalues over 1, resulting in one factor, cumulatively capturing 42% of the variance in the original predictors. Thus, all five L2 measures loaded on a single factor, which was termed the L2-Availability Factor. Table 10 presents the results of the PCA, including the L2-Availability Factor loadings and the percentage of unique variance explained by this factor.

Analysis protocol. As in Experiment 1, three separate analyses were performed for the TW, PTW, and FW reading measures. For each analysis, a maximal model of fixed and random effects was submitted to the *buildmer* function in the *buildmer* package (v. 2.2, Voeten, 2021) in R (version 4.0.3, R Core Team, 2020), which uses the *lmer* function from the *lme4* package (v. 1.1.-21, Bates et al., 2014). The maximal LME model in each analysis included (1) the fixed effects of the control variables: L1-Arabic Semantic Fluency (to control for potential differences in participants' L1 verbal abilities), Target Location (i.e., the TW number in the sentence), Sentence Length (i.e., the number of words in the sentence), Word Length (i.e., the number of letters in either the TW, PTW, or FW), and Word Frequency (of either the TW, PTW, or FW), which were all continuous and normalized; (2) the fixed effects of the variables of interest[4] (dummy coded): Context (Low/High, with High as the reference), Age Group (Younger/Older, with Older as the reference), L2-Availability Factor (continuous, centered with a mean of 0 as the reference), as well as the interactions among Context and the two predictors: Age Group and L2-Availability Factor (all of which were included in the final models using the *include=as.function* argument to allow

estimation of their contribution); (3) the random effects of Participants and Items with by-participant and by-item intercepts, by-participant slopes for Context, and by-item slopes for Context, Age Group, L2-Availability Factor, as well as for the interactions between Context and the predictors: Age Group and L2-Availability Factor. The selected models for the TW, PTW, and FW analyses are presented in Table 11.

Finally, as in Experiment 1, the selected model in each analysis was refitted using the *lmer* function, and *p*-values for all fixed effects and interactions were determined using the *anova* function from the *stats* package (v. 4.0.3, R Core Team, 2020). Furthermore, interactions and pairwise comparisons were tested using the *testInteraction* function from the *phia* package (v. 0.2-1, De Rosario-Martinez, 2015) and the *emmeans* function from the *emmeans* package (v. 1.7.1-1, Russell, 2021). These are presented in the text. In addition, model summaries (obtained from the *summary* function) are presented in Table 11. Note that the fixed effects presented in Table 11 reflect simple effects (e.g., the coefficient for Context reflects its effect at the reference level of the other factors, namely, the Older group, and the mean score of the L2-Availability Factor) rather than the main effects collapsing across all levels.

Results

As in Experiment 1, RTs were log transformed since examination of the RT distribution revealed substantial deviation from normality and log transformation of the raw RT improved the QQ plot, skew (raw RT = 42.23; log RT = 0.73), and kurtosis (raw RT = 3680.15; log RT = 3.93) of the distribution.

Across all three measures, there was a main effect of the L2-Availability Factor: TW: $F(1) = 54.99$, $p < .001$; PTW: $F(1) = 57.35$, $p < .001$; FW: $F(1) = 57.58$, $p < .001$. Namely, participants with higher L2-Availabilty scores responded faster than participants with lower scores, irrespective of the other variables. In addition, there was a main effect of Age Group, TW: $F(1) = 9.82$, $p = .002$; PTW: $F(1) = 12.44$, $p < .001$; FW: $F(1) = 16.85$, $p < .001$, such that younger participants responded faster than older ones, irrespective of the other variables.

Importantly, in the TW analysis, the two-way interaction between the L2-Availability Factor and Context was significant, $F(1) = 4.34$, $p = .04$. Thus, the difference between High and Low constraint contexts was modulated by individual differences in L2-availability score. As seen in Figure 2, having lower L2-availability scores resulted in faster responses for High than for Low constraint sentences, whereas having higher L2-availability scores led to the opposite pattern. However, the simple effects of Context, namely, the difference between High and Low constraint sentences in each level of L2-availability, was not significant for both low scores (e.g., 2 *SD*s below the mean score: $Z = -1.10$, $p = .27$) and high scores (e.g., 2 *SD*s above the mean score: $Z = 1.71$, $p = .09$).

Furthermore, in the FW analysis, the two-way interaction between the L2-Availability Factor and Context was also significant, $F(1) = 4.44$, $p = .03$. Thus, similar to the TW analysis, the difference between High and Low constraint contexts was modulated by individual differences in L2-availability score. As seen in Figure 3, having higher L2-availability scores resulted in faster responses for High- than for Low-constraint sentences, yet having lower L2-availability scores led to the opposite pattern. However, the simple effects of Context, namely, the difference between High- and Low-constraint sentences in each level of L2-availability, was not significant for both low scores (e.g., 2 *SD*s below the *M* score: $Z = 1.47$, $p = .14$) and high scores (e.g., 2 *SD*s above the *M* score: $Z = -1.49$, $p = .14$).

In addition, the two-way interaction between Age Group and Context was significant, $F(1) = 4.24$, $p = .04$, indicating that there was a significant difference between Older and Younger participants in the pattern of response to the two Context conditions (see Figure 4). However, the simple effects

Table 11. The three LME models in Experiment 2 predicting RTs for the target word, post target word, and final word. Effect size (*b*), standard errors (*SE*), and *t* value (*t*).

Fixed effects	Target word			Post target word			Final word		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	6.55	.06	114.78***	6.58	0.06	111.52***	6.59	0.06	111.06***
L2-Availability Factor	–0.17	0.02	–7.41***	–0.18	0.02	–7.14***	–0.02	0.02	–7.85***
Age Group (Younger)	–0.21	0.06	–3.39***	–0.23	0.06	–3.64***	–0.02	0.06	–3.65***
Context (Low)	–0.01	0.03	–0.38	–0.03	0.03	–3.64	0.02	0.03	0.91
Context (Low): L2-Availability Factor	–0.02	0.01	–2.10*	–0.01	0.02	–0.63	0.02	0.01	2.11*
Context (Low): Age Group (Younger)	0.01	0.03	0.24	0.03	0.02	1.28	–0.05	0.02	–2.06*
Control variables	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*z*
ArSemanticFluency	–0.05	0.02	–2.33*	–	–	–	–	–	–
TargetLocation	–0.06	0.01	–3.66***	–0.03	0.01	–2.71**	–	–	–
SentenceLength	0.03	0.01	2.38*	–	–	–	–0.02	0.01	–2.29*
WordLength	0.03	0.01	2.38*	0.06	0.01	6.41***	0.06	0.01	7.17***
WordFrequency	–0.02	0.01	–2.38*	–	–	–	–	–	–
Random effects	*Variance*		*SD*	*Variance*		*SD*	*Variance*		*SD*
Participant (intercept)	0.06		0.25	0.07		0.26	0.07		0.26
Context (low)	0.00		0.06	–		–	–		–
Item (intercept)	0.00		0.09	0.01		0.11	0.01		0.10
Context (low)	0.01		0.10	0.02		0.14	0.01		0.10
L2-Availability Factor	0.00		0.04	0.00		0.06	–		–
Context (low): L2-Availability Factor	–		–	0.01		0.09	–		–
Residual	0.12		0.35	0.14		0.37	0.14		0.37

Note. Fixed effects reflect simple effects relative to the reference level, when other factors are at their reference level, without correction for multiple comparisons. For main effects see *F* values in the text. ArSemanticFluency is participants' mean scores in a semantic fluency test in the L1-Arabic that was included to control for potential differences in L1 verbal abilities within this sample of Arabic–Hebrew bilinguals (see Procedure in the "Method" section). LME: linear-mixed-effect.

Sig. codes. 0 "***" 0.001 "**" 0.01 "*" 0.05.

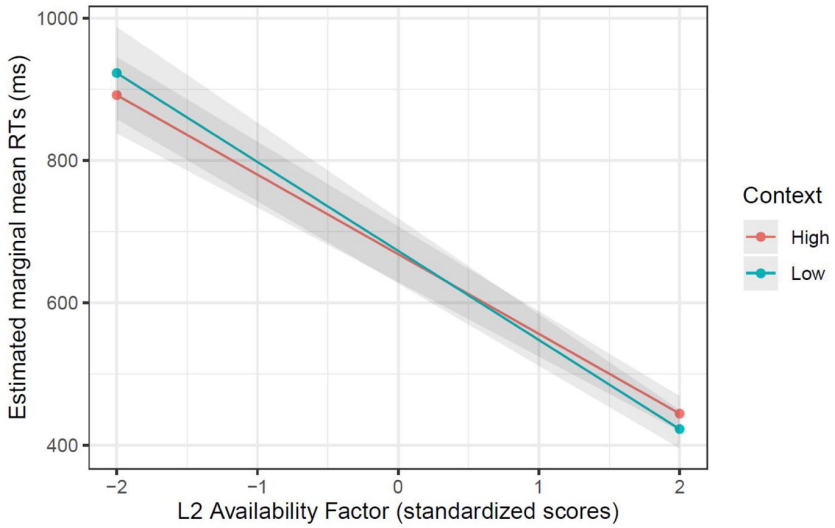

Figure 2. Estimated marginal mean RTs by L2-availability factor and context in the target word analysis of Experiment 2.
Note. Shaded areas mark *SE*.

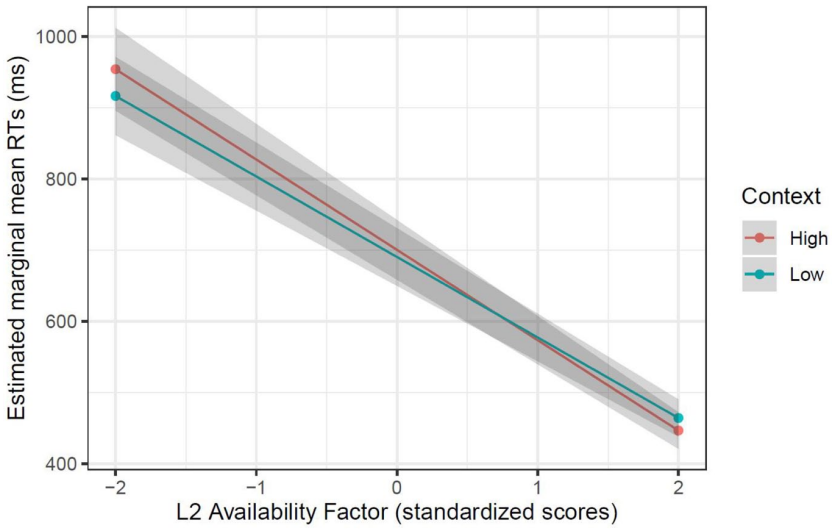

Figure 3. Estimated marginal mean RTs by L2-availability factor and context in the final word analysis of Experiment 2.
Note. Shaded areas mark *SE*.

of Context, namely, the difference between High- and Low-constraint sentences in each level of Age Group did not reach significance for either Younger participants, $\chi^2(1) = 1.99$, $p = .32$, or Older participants, $\chi^2(1) = 0.83$, $p = .72$.

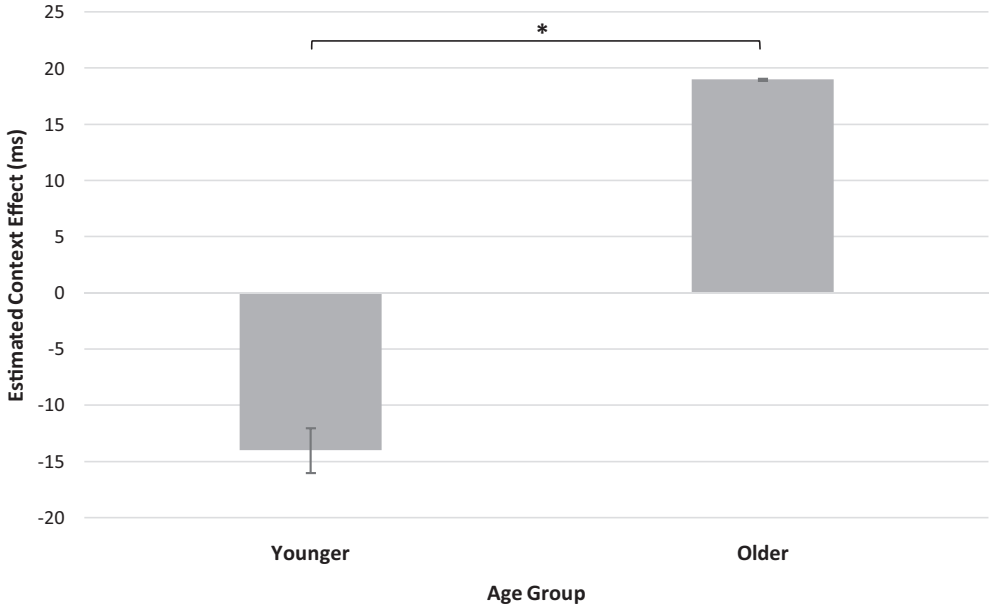

Figure 4. Estimated context effect by age group in the final word analysis of Experiment 2.
Note. Estimated Context Effect is the difference in the estimated marginal mean RTs (ms) between Low- and High-constraint sentences. Error bars mark *SE*.
*$p < .05$.

In sum, in Experiment 2, the effect of context again did not reach significance in any of the reading measures (TW, PTW, and FW). Nonetheless, in both the TW and FW, the sensitivity to high- and low-constraint contexts was modulated by the L2-Availability Factor, however, in opposite directions (as indexed by the opposing slope signs in the two regions, see Table 11). In the TW, semantic context appears to have benefited participants with lower scores of L2-Availability Factor. However, in the FW, semantic context seems to have benefited participants with higher scores of L2-Availability Factor. In addition, in the FW, contextual sensitivity was also modulated by Age Group, such that the facilitative effect of high constraint sentences was numerically evident only among Older participants. Note that none of the simple effects tests within each level of the Age Group or the ± 2 *SD* L2 Availability Factor reached significance.

General discussion

In two experiments we examined whether Hebrew readers use contextual semantic information to facilitate the processing of ambiguous (i.e., homonyms) and unambiguous TWs embedded in sentences. Using the SPR task, participants read Hebrew sentences, in which an ambiguous or unambiguous TW was preceded by either a highly constraining context or a neutral context. Reading times at three different points along the sentence (i.e., TW, PTW, and FW) were analyzed in order to reveal the influence of previous semantic context on sentence reading across time. In Experiment 1, we compared L1 and L2 Hebrew readers, whereas in Experiment 2 we focused on L2 Hebrew readers and examined the influence of L2 availability and age on L2 reading behavior.

Experiment 1 revealed that L2 readers were slower than L1 readers across all three reading time measures, indicating that L2 reading was more effortful than L1 reading. More importantly, L1 and L2 readers demonstrated significantly different reading behavior, as indicated by the significant Context by Group interaction. Examination of the pattern of reading times may suggest that L2 readers used contextual support more than L1 readers. As this difference emerged only on the FW of the sentence, it may have resulted from the involvement of integration differences, rather than from differences on purely predictive processes. Critically, the difference between high- and low-constraint sentences was not significant in either group. Two possible explanations may underlie the lack of this context effect—stimulus variability and participant variability.

With respect to stimulus variability, because the study design in both experiments was such that different sentences were used in the low- and high-constraint contexts, possible variations among the sentences in these two context conditions may have influenced sentence processing, especially by modulating wrap-up effects in the FW. To control for these potential modulations, target frequency and location within the sentence as well as the sentence length and syntactic complexity were taken into account in the statistical analyses. However, other factors such as the mean frequency and predictability of words in each sentence, or in the TWP, were not accounted for (see also Mor & Prior, 2022). This was the case since the limited availability of normative data in Hebrew have reduced our ability to adequately account for such sentence variance (for discussion, see Tokowicz & Degani, 2021). With such understudied languages, a complementary approach by which the same sentence is used with different TWs (e.g., Frisson et al., 2017) may provide a better remedy for potentially uncontrolled sentence differences in future work.

With respect to participant variability, although Experiment 1 contrasted two groups which differ in their language profile, heterogeneity within the L2 group may have increased within-group variance. We reasoned that L1-Arabic participants who are exposed to Hebrew on a regular basis and use it more often may resemble the L1-Heberw group in reading behavior (Norman et al., 2016), because of greater accessibility to the L2-Hebrew. In contrast, those L2 readers who are less proficient in Hebrew and use it less often, may rely more strongly on contextual cues as a compensatory cognitive mechanism that reduces reading effort (Stanovich, 1980, 1984). As such, the critical difference may not lie between L1 and L2 readers per se but may be linked to language accessibility as a continuous dimension. Indeed, in a picture naming task, Gollan et al. (2011) showed that less proficient Dutch–English bilinguals (L2), but not highly proficient Spanish–English bilinguals (L2), benefited more from high-constraint context than did English monolinguals (L1), suggesting that it is the reduced proficiency in the L2 that might lead to greater reliance on context.

Experiment 2 was specifically designed to examine this possibility by focusing on L2 readers and testing a wider and more variable group of participants, in order to examine the influence of a continuous L2 availability score and age on context effects in the L2. This experiment revealed that across all three reading time measures, L2 readers with low L2 availability exhibited slower reading times than readers with high L2 availability. Furthermore, older adults exhibited slower reading times than younger adults, suggesting that L2 reading is more effortful when the L2 is less available and when readers are older. More importantly, these two factors significantly modulated participants' L2 reading behavior in the two context conditions.

First, with respect to the modulation of L2 availability, participants with lower L2 availability score seem to benefit from constraining context earlier along the sentence (i.e., when processing the TW). Participants with higher L2 availability, however, were influenced by previous context later along the sentence (i.e., when processing the FW of the sentence; see Figures 2 and 3). In both cases, the sensitivity to previous context was clearly modulated by L2 availability, as indicated by the significant interaction between the L2 Availability Factor and Context, but the difference between the two context conditions was not significant for both low (2 *SD*s below the mean score)

and high (2 *SD*s above the mean score) scores of L2 availability. This lack of a significant context effect is assumingly due to our attempt to sample a heterogeneous group of Arabic–Hebrew bilinguals that resulted in less statistical power. Thus, the current pattern of result can serve only as suggestive evidence that as the availability of the L2 decreases, previous context may influence word reading earlier along the sentence.

Moreover, as opposed to measures capturing a specific aspect of variability in the L2 (i.e., word knowledge, single word reading fluency, exposure percentage) as considered in previous studies (Mor & Prior, 2021; Whitford & Titone, 2017), it seems that the L2-Availability Factor used in the current study, which captured a wider range of L2 characteristics (i.e., AoA, current exposure, current use, subjective proficiency, semantic fluency), is more suitable to explain variability among L2 readers in the tendency to use contextual information to facilitate online sentence reading. Indeed, as Mor and Prior (2022) have previously suggested, it appears that this tendency may relate to several aspects of language experience and abilities, and thus, cannot be explained by lexical proficiency measures alone.

Examination of the results across the two experiments revealed that in Experiment 1 the sensitivity to previous context was modulated by language background (i.e., L1 vs. L2) only at the end of the sentence, whereas in Experiment 2, context sensitivity was modulated by language background (i.e., L2 availability measure), either earlier along the sentence for participants with lower L2 availability, or at the end of the sentence for participants with higher L2 availability, similar to the results of Experiment 1. Thus, the group of Arabic–Hebrew bilinguals in Experiment 1 may be more similar to the group of the Arabic–Hebrew bilinguals with the higher L2 availability in Experiment 2, in terms of their reading behavior, because both groups exhibited sensitivity to previous context on sentence's FW. Conversely, Arabic–Hebrew bilinguals with lower L2 availability in Experiment 2 were more sensitive to contextual information earlier along the sentence, on the TW.

To further investigate these differences in reading behavior, we performed a median split on the L2 availability scores of participants in Experiment 2, creating two groups of participants with low- and high- L2 availability, and compared them with the participants in the L1-Arabic (L2-Hebrew) group of Experiment 1, in terms of their collected L2-Hebrew measures (see Appendix 4). Participants in the high L2 availability group (Experiment 2) had a comparable mean rating of current Hebrew use and Hebrew semantic fluency to that of participants in the L2-Hebrew group (Experiment 1). However, participants in the high L2 availability group (Experiment 2) had started to acquire Hebrew earlier and had greater current exposure to Hebrew than participants in the L2-Hebrew group (Experiment 1). Thus, the fact that participants in the L2-Hebrew group (Experiment 1) and in the high L2 availability group (Experiment 2) exhibited the same pattern of context effects and were comparable only in their mean rating of current Hebrew use and in the semantic fluency score, but not in AoA and current exposure, suggests that perhaps current use and fluency may be the more critical predictors in explaining context effects in the L2. Future studies will explore to what extent variability in L2 fluency and current use indeed consistently affect the balance between top–down (contextual) and bottom–up (lexical and sublexical) reading processes in the L2.

The second finding in Experiment 2 was that age significantly modulated the sensitivity to context, as indicated by the significant interaction between Context and Age Group. This finding suggests that older and younger adults significantly differed in the way they processed high- and low-constraint sentences. However, when examined within each group separately, the difference between high and low constraining context did not reach significance. Nevertheless, the direction of the effects observed in the current findings raise the possibility that older adults tend to benefit more from constraining context than younger adults. While this possibility needs to be verified in future studies, it is in line with previous L1 studies that have demonstrated stronger context effects in older, relative to younger adults (Choi et al., 2017; Rayner et al., 2006; Zhao et al., 2019, 2021),

as well as with a compensatory processing account, according to which older adults tend to compensate for slower lexical processing by making greater use of contextual cues.

Of note, in both studies, individuals who were characterized by slower reading times were those who were more likely to exhibit context sensitivity. Thus, it could be argued that slower processing per se allows for context-dependent effects to emerge, and that there is no need to postulate the involvement of compensatory mechanisms. However, if this was the case, then one would expect a similar direction of the context effect across individuals with slower and faster processing. Examination of the pattern exhibited across Figures 1 and 4 in the current study indicate a qualitative difference, in that those individuals who processed the sentences more quickly patterned in the opposite direction than those who processed the sentences more slowly, with no indication toward context-based facilitation. Thus, it is more likely that individuals with slower reading engage in compensatory mechanisms that allow them to tip the balance toward reliance on context to compensate for the slower and more effortful bottom–up (lexical) process. Future studies in which effort measures are taken into account may be revealing on this issue.

Finally, as opposed to previous L1 and L2 studies (e.g., Duffy et al., 1988, L1; Elston-Güttler & Friederici, 2005, 2007), in the current study we did not find an effect of lexical ambiguity, in both languages. Namely, both L1 and L2 sentence reading were not influenced by whether TWs consisted of one or two possible meanings. In addition, although we expected that processing of ambiguous words would benefit more from constraining context than processing of unambiguous words, lexical ambiguity did not modulate the sensitivity to context. One possible reason could be that both Hebrew and Arabic consist of a relatively high rate of ambiguous heterophonic homographs (i.e., words that are written the same but pronounced differently) due to the unique characteristics of their writing systems, which do not convey the complete phonological form of words (Abu-Rabia, 2001; Frost & Bentin, 1992). Therefore, both L1-Herbew readers and L2-Hebrew readers whose L1 is Arabic (which is similar to Hebrew in this respect), are highly experienced in dealing with lexical ambiguity, and thus, do not show the expected sensitivity to lexical ambiguity. At the same time, it is also possible that frequency dominance of the selected homonyms affected the observed pattern. Specifically, the high-constraint sentences in the current study were biased toward the dominant meaning of the homonyms that already had a frequency advantage in activation, regardless of context. Thus, under these conditions the effect of lexical ambiguity was eliminated. Therefore, the ambiguity manipulation employed in the current study did not strengthen the effects of context, but L2 availability and age still appear to modulate context sensitivity.

To conclude, the results of the current study suggest that when the language is less available due to lower use, exposure, and proficiency (i.e., L2 users), or when reading performance decline as a result of aging processes, readers are more likely to make use of previous context to enhance sentence processing. Furthermore, they suggest that as the availability of a language decreases, readers may use contextual information earlier along the sentence. However, previous studies have shown that L1 readers, which presumably have the highest levels of language availability, use context earlier along the sentence, in comparison to L2 readers (Elston-Güttler & Friederici, 2005, 2007). Thus, the relation between language availability and the use of context may be nonlinear, and this issue requires additional work. Finally, consistent with the *interactive-compensatory* model the evidence presented in this paper points toward the role of compensatory reliance on context, as evident in both L2 readers and older adults. Notably, however, as the effects of context were not significant when examined within each group separately, more research is needed for stronger conclusions to be made.

Furthermore, as natural language processing typically unfolds in context, the way in which different individuals can capitalize on such contextual support may help explain individual differences in reading and language comprehension. Moreover, the current study fits with the growing understanding that language background characteristics may affect individuals' reading (Nisbet et al., 2021), and suggest that these influences may include the way in which different individuals utilize

context. Future studies in which additional modulating factors are examined would facilitate a fuller understanding of the factors that determine individual performance in linguistic tasks.

Acknowledgements

The authors thank Anat Prior for fruitful discussions of this project.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tamar Degani https://orcid.org/0000-0003-2604-8523

Notes

1. Note that these L1 studies often did not specify participants' language profile. Therefore, it is unclear whether participants in these studies were monolinguals using their native language or whether they had knowledge in other languages, deeming this the L1 of multilingual speakers, which may have resulted in distinct processing.
2. Further examination of the current set of sentences revealed that the high- and low-constraint sentences significantly differ ($p < .05$) also in syntactic complexity as reflected by the number of clauses per sentence—commonly used measure for evaluating syntactic complexity (Jagaiah et al., 2020), but only for control sentences (High: $M = 2.33$, $SD = 0.18$; Low: $M = 1.47$, $SD = 0.13$). In addition, control and homonym sentences significantly differ based on this measure only in the high-constraint context. Nevertheless, including this syntactic complexity measure in the statistical analyses did not result in a significant effect, and as such this variable was not selected by the *buildmer* function (see the analysis approach in the "Result" section). Thus, the pattern of results reflected in the selected model did not change.
3. Examination of the sentences in terms of syntactic complexity (number of clauses per sentence, Jagaiah et al., 2020) revealed that high- and low-constraint sentences significantly differ ($p < .05$) also in this measure, but only for control sentences (High: $M = 2.06$, $SD = 0.12$; Low: $M = 1.70$, $SD = 0.13$). Nevertheless, including this syntactic complexity measure in the statistical analyses as a control variable did not result in a significant effect, and as such it was not retained by the *buildmer* function (see the analysis approach in the "Result" section). Accordingly, the pattern reflected by the selected final model did not change.
4. Note that we decided to exclude the Target Type factor from the maximal model submitted to the *buildmer* function, based on the results of Experiment 1 (see Table 5) and on preliminary analyses of the data (see Appendix 3), in which the Target Type factor was not significant and did not significantly interact with any of the other variables of interest. This was done to preserve statistical power.

References

Abu-Rabia, S. (1997). Reading in Arabic orthography: The effect of vowels and context on reading accuracy of poor and skilled native Arabic readers. *Reading and Writing*, *9*(1), 65–78. https://doi.org.ezproxy.haifa.ac.il/10.1023/A:1007962408827

Abu-Rabia, S. (2001). The role of vowels in reading Semitic scripts: Data from Arabic and Hebrew. *Reading and Writing*, *14*(1), 39–59. https://doi.org.ezproxy.haifa.ac.il/10.1023/A:1008147606320

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086. https://doi.org.ezproxy.haifa.ac.il/10.1080/02724980443000476

Bar-On, A., Dattner, E., & Ravid, D. (2017). Context effects on heterophonic-homography resolution in learning to read Hebrew. *Reading and Writing*, *30*(3), 463–487. https://doi.org.ezproxy.haifa.ac.il/10.1007/s11145-016-9685-1

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv preprint arXiv:*1506*.04967.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. arXiv preprint arXiv:1406.5823.

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(6), 1022–1047. https://doi.org.ezproxy.haifa.ac.il/10.1037/xlm0000770

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*(5), 223–233. https://doi.org.ezproxy.haifa.ac.il/10.1093/geronb/60.5.P223

Brothers, T., & Kuperberg, G. R. (2021). Word predictability effects are linear, not logarithmic: Implications for probabilistic models of sentence comprehension. *Journal of Memory and Language*, *116*, 104–174. https://doi.org/10.1016/j.jml.2020.104174

Choi, W., Lowder, M. W., Ferreira, F., Swaab, T. Y., & Henderson, J. M. (2017). Effects of word predictability and preview lexicality on eye movements during reading: A comparison between young and older adults. *Psychology and Aging*, *32*(3), 232–242. https://doi.org/10.1037/pag0000160

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, *117*, 135–147. https://doi.org/10.1016/j.neuropsychologia.2018.05.023

Degani, T., & Tokowicz, N. (2010). Semantic ambiguity within and across languages: An integrative review. *Quarterly Journal of Experimental Psychology*, *63*(7), 1266–1303. https://doi.org.ezproxy.haifa.ac.il/10.1080/17470210903377372

De Rosario-Martinez, H. (2015). *phia: Post-Hoc interaction analysis* (R package version 0.2-1). https://CRAN.R-project.org/package=phia

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Elston-Güttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, *52*(2), 256–283. https://doi.org/10.1016/j.jml.2004.11.002

Elston-Güttler, K. E., & Friederici, A. D. (2007). Ambiguous words in sentences: Brain indices for native and non-native disambiguation. *Neuroscience Letters*, *414*(1), 85–89. https://doi.org/10.1016/j.neulet.2006.12.002

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141. https://doi.org/10.1111/j.1469-8986.2005.00274.x

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: Age-related changes across levels of language processing. *Psychology and Aging*, *18*(4), 858–872. https://doi.org.ezproxy.haifa.ac.il/10.1037/0882-7974.18.4.858

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461–1469. https://doi.org/10.1037/a0036756

Fozard, J. L., & Gordon-Salant, S. (2001). Changes in vision and hearing with aging. In J. E. Birren, & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed., pp. 241–266). Academic Press.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200–214. https://doi.org/10.1016/j.jml.2017.04.007

Frost, R., & Bentin, S. (1992). Processing phonological and semantic ambiguity: Evidence from semantic priming at different SOAs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 58–68. https://doi.org.ezproxy.haifa.ac.il/10.1037/0278-7393.18.1.58

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. http://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–576. https://doi.org/10.1037/0894-4105.16.4.562

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209. https://doi.org/10.1037/a0022256

Gordon, P. C., Lowder, M. W., & Hoedemaker, R. S. (2016). Reading in normally aging adults. *Cognition, Language and Aging*, 165–191. https://doi.org/10.1075/z.200.07gor

Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., & Titone, D. (2021). Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. *Applied Psycholinguistics*, *42*(2), 245–278. https://doi.org.ezproxy.haifa.ac.il/10.1017/S0142716420000521

Henrich, J., Heine, S., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466, 29. https://doi.org/10.1038/466029a

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*(1), 62–82. https://doi.org.ezproxy.haifa.ac.il/10.1037/0096-3445.124.1.62

Huettig, F., & Brouwer, S. (2015). Delayed anticipatory spoken language processing in adults with dyslexia—Evidence from eye-tracking. *Dyslexia*, *21*(2), 97–122. https://doi.org.ezproxy.haifa.ac.il/10.1002/dys.1497

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, *86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652. https://doi.org/10.1037/xlm0000315

Ito, A., & Pickering, M. J. (2021). Automaticity and prediction in non-native language comprehension. In E. Kaan, & T. Grüter (Eds.), *Prediction in second-language processing and learning*. John Benjamins, pp. 25–46.

Jagaiah, T., Olinghouse, N. G., & Kearns, D. M. (2020). Syntactic complexity measures: Variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing*, *33*(10), 2577–2638.

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org.ezproxy.haifa.ac.il/10.1037/0096-3445.111.2.228

Kastenbaum, J. G., Bedore, L. M., Peña, E. D., Sheng, L., Mavis, I., Sebastian-Vaytadden, R., . . .Kiran, S. (2019). The influence of proficiency and language combination on bilingual lexical access. *Bilingualism, Language and Cognition*, *22*(2), 300–330. https://doi.org/10.1017/s1366728918000366

Kavé, G. (2005). Phonemic fluency, semantic fluency, and difference scores: Normative data for adult Hebrew speakers. *Journal of Clinical and Experimental Neuropsychology*, *27*(6), 690–699. https://doi.org.ezproxy.haifa.ac.il/10.1080/13803390490918499

Kemper, S., Crow, A., & Kemtes, K. (2004). Eye-fixation patterns of high-and low-span young and older adults: Down the garden path and back again. *Psychology and Aging*, *19*(1), 157–170. https://doi.org/10.1037/0882-7974.19.1.157

Kemper, S., & Liu, C. J. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, *22*(1), 84–93. https://doi.org/10.1037/0882-7974.22.1.84

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., . . .Suchomel, V. (2014). The Sketch engine: Ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kilgarriff, A., Reddy, S., Pomikálek, J., & Avinesh, P. V. S. (2010, May). A corpus factory for many languages. In Proceedings of the seventh conference on International Language Resources and Evaluation (LREC 2010) (pp. 904–910). European Language Resources Association.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284 . https://doi.org/10.1080/09541440340000213

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., & Hayakawa, S. (2021). Measuring bilingualism: The quest for a "bilingualism quotient." *Applied Psycholinguistics*, *42*(2), 527–548. https://doi.org.ezproxy.haifa.ac.il/10.1017/S0142716420000533

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Martinez, H. D. R. (2015). *Analysing interactions of fitted models*. http://cran.wustl.edu/web/packages/phia/vignettes/phia.pdf

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mitsugi, S., & Macwhinney, B. (2016). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition*, *19*(1), 19–35. https://doi.org/10.1017/S1366728914000881

Mor, B., & Prior, A. (2022). Frequency and predictability effects in first and second language of different script bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*, 1363–1383. https://doi.org/10.1037/xlm0000927

Morris, R. K. (2006). Lexical processing and sentence context effects. In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.). Academic Press, pp. 377–401. https://doi.org/10.1016/B978-012369374-7/50011-0

Nicklin, C., & Plonsky, L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data reanalysis. *Annual Review of Applied Linguistics*, *40*, 26–55. https://doi.org/10.1017/S0267190520000057

Nisbet, K., Bertram, R., Erlinghagen, C., Pieczykolan, A., & Kuperman, V. (2022). Quantifying the difference in reading fluency between L1 and L2 readers of English. *Studies in Second Language Acquisition*, *44*, 407–434. https://doi.org/10.1017/S0272263121000279

Norman, T., Degani, T., & Peleg, O. (2016). Transfer of L1 visual word recognition strategies during early stages of L2 learning: Evidence from Hebrew learners whose first language is either Semitic or Indo-European. *Second Language Research*, *32*(1), 109–122. https://doi.org.ezproxy.haifa.ac.il/10.1177/0267658315608913

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology*, *60*, 173–196. https://doi.org.ezproxy.haifa.ac.il/10.1146/annurev.psych.59.103006.093656

Payne, B. R., & Federmeier, K. D. (2018). Contextual constraints on Lexico-semantic processing in aging: Evidence from single-word event-related brain potentials. *Brain Research*, 1687, 117–128. https://doi.org/10.1016/j.brainres.2018.02.021

Peleg, O., Markus, A., & Eviatar, Z. (2012). Hemispheric asymmetries in meaning selection: Evidence from the disambiguation of homophonic vs. heterophonic homographs. *Brain and Cognition*, *80*(3), 328–337. https://doi.org/10.1016/j.bandc.2012.08.005

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

Prior, A., Degani, T., Awawdy, S., Yassin, R., & Korem, N. (2017). Is susceptibility to cross-language interference domain specific? *Cognition*, *165*, 10–25. https://doi.org/10.1016/j.cognition.2017.04.006

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. https://doi.org.ezproxy.haifa.ac.il/10.3758/BF03214555

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Russell, L. (2021). *emmeans: Estimated marginal means, aka least-squares means* (R Package Version, 1.7.1-1). https://CRAN.R-project.org/package=emmeans

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*(5), 754–760. https://doi.org/10.1017/S1355617710000706

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, *16*(1), 32–71. https://doi.org/10.2307/747348

Stanovich, K. E. (1984). The interactive-compensatory model of reading: A confluence of developmental, experimental, and educational psychology. *Remedial and Special Education*, *5*(3), 11–19. https://doi.org.ezproxy.haifa.ac.il/10.1177/074193258400500306

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, *31*(5), 811–826. https://doi.org/10.1037/0012-1649.31.5.811

Tokowicz, N., & Degani, T. (2021). Design challenges: Development of bilingual materials and making appropriate comparisons. In W. S. Francis (Ed.), *Bilingualism across the lifespan* (pp. 106–122). Routledge.

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, *22*(7), 1092–1106. https://doi.org/10.1080/09541440903325178

Voeten, C. (2021). *buildmer: Stepwise elimination and term reordering for mixed-effects regression* (R package version 2.2). https://CRAN.R-project.org/package=buildmer

Whitford, V., & Titone, D. (2012). Second-language experience modulates first-and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*(1), 73–80. https://doi.org.ezproxy.haifa.ac.il/10.3758/s13423-011-0179-5

Whitford, V., & Titone, D. (2016). Eye movements and the perceptual span during first-and second-language sentence reading in bilingual older adults. *Psychology and Aging*, *31*(1), 58–70 . https://doi.org.ezproxy.haifa.ac.il/10.1037/a0039971

Whitford, V., & Titone, D. (2017). The effects of word frequency and word predictability during first-and second-language paragraph reading in bilingual older and younger adults. *Psychology and Aging*, *32*(2), 158–177. https://doi.org/10.1037/pag0000151

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. https://doi.org/0.1037/a0029206

Zhao, S., Li, L., Chang, M., Wang, J., & Paterson, K. B. (2021). A further look at ageing and word predictability effects in Chinese reading: Evidence from one-character words. *Quarterly Journal of Experimental Psychology*, *74*(1), 68–76. https://doi.org/10.1177/1747021820951131

Zhao, S., Li, L., Chang, M., Xu, Q., Zhang, K., Wang, J., & Paterson, K. B. (2019). Older adults make greater use of word predictability in Chinese reading. *Psychology and Aging*, *34*(6), 780–790. https://doi.org/10.1037/pag0000382

Author biographies

Tal Norman is a post-doctoral researcher studying L1-L2 processing differences at the morphological, phonological, orthographic and semantic levels.

Tamar Degani is a cognitive psychologist, studying multilingualism and language learning. Her research examines the interplay among the different languages of multilingual speakers, and between languageand cognition more broadly.

Appendix 1

List of critical stimuli used in Experiment 1

Ambiguous homonyms.

Target word	Context type	Sentence
אות	Low	בשלט רואים **אות** גדולה ובולטת ובמיוחד.
	High	לוקח לילדים ללמוד שאחרי מם ונון מגיעה **האות** סמך.
גיל	Low	כל אחד מאיתנו יגיע **לגיל** הזקנה בשלב מאוחר בחייו.
	High	אמא אמרה שאוכל להוציא רישיון כשאגיע **לגיל** המתאים לנהיגה.
רוק	Low	המורה הגדירה **רוק** כתמיסה שמורכבת בעיקר ממים.
	High	בתוך הפה יש נוזל שקוף שנקרא **רוק** והוא חשוב לבריאות.
מתח	Low	אני מרגישה הרבה **מתח** בין משתתפי הקבוצה.
	High	בדרך כלל בסיום של פרק בסדרה אני נשארת **במתח** לקראת הפרק הבא.
הנחה	Low	בתקופה האחרונה אני מקבלת **הנחה** בכל ביקור בחנות.
	High	מחיר החולצה שרציתי לקנות היה יקר ולכן חיכיתי שתהיה עליה **הנחה** בסוף העונה.
רווח	Low	בסופו של דבר **הרווח** שקיבלתי בעסקה היה קטן.
	High	לפני שאדם פותח עסק הוא בודק מה יהיה ההפסד הכספי מול **הרווח** שלו מכך.
פרקים	Low	אחד הדברים שאני מסתכל עליהם בזמן קניית ספר הוא כמה **פרקים** יש בו.
	High	העונה השנייה של הסדרה האהובה עליי צילמה הפעם עשרה **פרקים** מותחים ומרגשים.
אח	Low	בתחילת החודש קראתי כתבה על **אח** של זמרת מפורסמת.
	High	לאמא שלי יש שתי אחיות **ואח** קטן.
עצב	Low	היו לי רגשות של **עצב** מהולים בשמחה בעקבות האירוע.
	High	לכל אדם יש רגעים של שמחה ורגעים של **עצב** במהלך חייו.
שיח	Low	זה סוג של **שיח** שניתן לגדל בתוך הבית.
	High	הפירות של עגבניה לא גדלים על עץ אלא על **שיח** לא גדול.
מפה	Low	מצאתי בארון **מפה** שקניתי בטיול בספרד לפני חמש שנים.
	High	לפני שהמציאו את הווייז אנשים ניווטו את הדרך על ידי מצפן **ומפה** של האזור.
הדחה	Low	אחד הדברים שאני פחות אוהבת בלצפות בחידונים הוא **הדחה** של משתתפים חלשים.
	High	בתוכניות ריאליטי בסוף כל פרק מתקיימת **הדחה** גורלית.
מנצח	Low	אני תמיד מרגיש **מנצח** כשאני מקבל ציון טוב.
	High	הראשון שמגיע לקו הסיום בתחרות ריצה הוא **המנצח** הגדול של התחרות.
סרטן	Low	בגלל שלסבתא יש **סרטן** נשימתה מאומצת.
	High	עישון מעלה ב90% את הסיכוי לחלות **בסרטן** הריאות ומחלות אחרות.
סרט	Low	באתר האינטרנט הופיע **הסרט** שכל כך חיכיתי לו.
	High	גל גדות ניצחה בפרס האוסקר על השתתפותה **בסרט** ''וונדר וומן'' המצליח.

(Continued)

Unambiguous controls.

Target word	Context type	Sentence
פרי	Low	בכל פעם שאני חוזר הביתה אני אוהב לאכול **פרי** טרי מהסלסלה.
	High	כמו תפוח, גם תפוז הוא **פרי** מזין שניתן לקטוף מעצים.
דלק	Low	בארוחת שישי כל המשפחה דנה בנושא ה**דלק** האוזל במדינת ישראל.
	High	נכנסתי לאוטו אחריי שאחותי נהגה עליו ומיד נדלקה נורית ה**דלק** כי המיכל היה ריק.
מגף	Low	עם התקרבות סוף העונה הלכתי לקניון עם החברה שלי וקניתי **מגף** שחור ומבריק.
	High	בחורף, כשיורד גשם, לא כדאי לנעול סנדל או נעל פתוחה אלא **מגף** אטום ומתאים.
בננה	Low	אני אוהבת לשלב **בננה** בתוך מילקשייק.
	High	הפרי האהוב על קופים הוא **בננה** צהובה ועסיסית.
זוג	Low	הקיץ התקרב, ולכן החלטתי לקנות לי **זוג** סנדלים יפים.
	High	כשרואים את גיל ודנה מיד חושבים שהם כזה **זוג** נחמד ומאושר.
מפלגה	Low	אתמול דיברו ברדיו על ה**מפלגה** החדשה שהוקמה.
	High	בבחירות הקרובות אני אצביע לשלמה כהן שעומד בראשות ה**מפלגה** החדשה שקמה.
קצף	Low	אפשר בבקשה להזמין קפה עם **קצף** בצד וקצת קינמון.
	High	יש למזוג בירה בכוס מוטה הצידה על מנת שלא יצטבר הרבה **קצף** בכוס ומעט נוזל.
בגד	Low	.אני לא צריכה **בגד** חדש כל חודש
	High	כשהולכים לים לובשים **בגד** ים ונועלים כפכפים.
מגבת	Low	בטיול השנתי איבדתי את ה**מגבת** הירוקה שהבאתי.
	High	אין לי סבלנות לתת לכלים להתייבש לבד ואני מעדיפה להשתמש ב**מגבת** ולהחזיר למקום.
דואר	Low	דנה לא מאמינה ב**דואר** ישראל בעקבות התנסויות שליליות.
	High	.המכתבים לא הגיעו אליי בזמן בגלל תקלות רבות ב**דואר** ישראל לאחרונה
שחף	Low	אני יודעת ש**שחף** הוא ממשפחת העופות.
	High	אבא סיפר שהציפור הלבנה הקולנית בים נקראת **שחף** והיא ציפור חכמה ומתוחכמת.
שלב	Low	.שמעתי שהם הגיעו ל**שלב** מתקדם בחקירה
	High	אחי הקטן שבר שיא במשחק מריו והצליח לעלות ל**שלב** האחרון במשחק.
נר	Low	אחותי קנתה לי **נר** כמתנה ליום הולדת.
	High	ביום הולדתו של איציק בן השנה הגישה אמא עוגה ואיציק היה צריך לכבות **נר** אחד בלבד.
חול	Low	אפשר למצוא סוגים שונים של **חול** במדבר הסהרה.
	High	בני בן השנה התפרץ בבכי על חוף הים כאשר מעד ופיו התמלא ב**חול** מים ולכלוך.
ספרייה	Low	לפי דעתי, ה**ספרייה** היא מקום נפלא.
	High	המקום השקט ביותר ללמוד בו באוניברסיטה הוא ה**ספרייה** שמשמשת אותי הרבה.

Appendix 2

List of critical stimuli used in Experiment 2

Ambiguous homonyms.

Target word	Context type	Sentence
אות	Low	בשלט רואים **אות** ממש גדולה ובולטת.
	High	לוקח לילדים ללמוד שאחרי מם ונון מגיעה ה**אות** סמך ורק אחריה עין ופה.
אח	Low	אמא שלי הזכירה לי שלפני בערך עשר שנים **אח** שלי טס לאנגליה.
	High	לאמא שלי יש שתי אחיות ו**אח** אחד שצעיר ממנה בעשר שנים.
הדחה	Low	בשל התנהלות לא נאותה המנהל דן באפשרות ה**הדחה** של הסגן הותיק.
	High	בפרק של הישרדות נפרדו משני מתמודדים בטקס ה**הדחה** ולא רק מאחד.
הנחה	Low	סימה התאכזבה מאד לגלות שה**הנחה** על הבשמים במשביר הסתיימה.
	High	בחנות קסטרו עשו 10% **הנחה** על הקולקציה החדשה.
השלמה	Low	לפעמים חשוב לעשות **השלמה** לפני מבחן כדי להצליח יותר.
	High	במבחן יש להשתמש במחסן המילים ולבצע למשפטים החלקיים **השלמה** כדי להצליח.
חלל	Low	הספה החדשה שקנינו תפסה מקום ב**חלל** הסלון והוסיפה אווירה חמה.
	High	המרחב הקיים בין גופים שמיים נקרא **חלל**, והוא כולל את כדור הארץ.
מנה	Low	אמרתי לחברה שקשה לי עם ה**מנה** הזאת, היא גורמת לי לכאבי בטן.
	High	במסעדת השף הגישו לי **מנה** מיוחדת של פסטה.
מנצח	Low	אתמול בערב, לאחר מהדורת החדשות, פרסמו את שם ה**מנצח** הגרלת הלוטו.
	High	הראשון שמגיע לקו הסיום בתחרות ריצה הוא ה**מנצח** הגדול של התחרות.
מפה	Low	קיפלתי והכנסתי לתיק את ה**מפה** לקראת הטיול מחר ברמות מנשה.
	High	לפני שהמציאו את הוויז אנשים ניווטו את הדרך על ידי מצפן ו**מפה** של האזור.
מתח	Low	אני מרגישה הרבה **מתח** בין משתתפי הקבוצה.
	High	הוא עצר את הסרט בשיא ה**מתח** ולפני שהסתיים.
סרט	Low	יוחנן ניגש לחנות כדי לקנות **סרט** עבור בתו הקטנה לכבוד חגיגת יום ההולדת.
	High	גל גדות ניצחה בפרס האוסקר על השתתפותה ב**סרט** "וונדר וומן" המצליח.
עמודים	Low	החומר החדש ממנו עשויים ה**עמודים** חזק יותר מהחומרים בהם השתמשו בעבר.
	High	לספר הארוך יש חמש-מאות **עמודים** בלי אף תמונה.
עצב	Low	היו לי רגשות של **עצב** מהולים בשמחה בעקבות האירוע.
	High	לכל אדם יש רגעים של שמחה ורגעים של **עצב** במהלך חייו.
פרקים	Low	אחד הדברים שאני מסתכל עליהם בזמן קניית ספר הוא כמה **פרקים** יש בו.
	High	העונה השנייה של הסדרה האהובה עליי צילמה הפעם עשרה **פרקים** מותחים ומרגשים.
רווח	Low	בסופו של דבר ה**רווח** שקיבלתי בעסקה היה קטן.
	High	בכל חברה יש דו"ח הפסד ו**רווח** שמרכז את כל ההכנסות וההוצאות.
רוק	Low	בתוכנית המדע סיפרו על חשיבות ה**רוק** לעיבוד יעיל של מזון.
	High	בתוך הפה יש נוזל שקוף שנקרא **רוק** והוא חשוב לבריאות.
שאיפה	Low	לפני שנכנסתי לראיון העבודה לקחתי **שאיפה** עמוקה ונכנסתי בראש מורם.
	High	תהליך הנשימה כולל את שלב הנשיפה ושלב ה**שאיפה** בו מוכנס אויר לריאות.
שיח	Low	נשארנו הרבה זמן לאחר המפגש כדי לשמוע על ה**שיח** שהיה בסיור אתמול.
	High	הפירות של עגבניה לא גדלים על עץ אלא על **שיח** לא גדול.

(Continued)

Unambiguous controls.

Target word	Context type	Sentence
אור	Low	בשיעור ביולוגיה למדנו על **אור** והשפעתו על הצמחיה.
	High	יאיר פחד לישון בחדר חשוך וביקש שאדליק את **האור** כדי שיוכל להירדם.
בגד	Low	אני לא צריכה **בגד** חדש כל חודש ואפילו לא כל חודשיים.
	High	כשהולכים לים לובשים **בגד** ים ונועלים כפכפים.
בננה	Low	ישנם סיבים תזונתיים חשובים ב**בננה** המאפשרים פעולת מעיים תקינה.
	High	הפרי האהוב על קופים הוא **בננה** צהובה ועסיסית.
ברז	Low	כולם חשבו שזה **ברז** איכותי אבל למעשה הוא חיקוי זול.
	High	בלילה התעוררתי מקולות טפטוף מים וגיליתי ששכחתי לסגור את ה**ברז** במקלחת כראוי.
דג	Low	לאחר חיפוש די ארוך מצאתי סוף סוף **דג** שאני מסוגלת לאכול.
	High	פורל הוא סוג של **דג** הנוטה לשחות נגד הזרם.
דואר	Low	דנה לא מאמינה ב**דואר** ישראל בעקבות התנסויות שליליות.
	High	המכתבים לא הגיעו אליי בזמן בגלל תקלות רבות ב**דואר** ישראל לאחרונה.
דלק	Low	בארוחת שישי כל המשפחה דנה בנושא ה**דלק** האוזל במדינת ישראל.
	High	נכנסתי לאוטו אחריי שאחותי נהגה עליו ומיד נדלקה נורית הדלק כי המיכל התרוקן.
הצגה	Low	חשבתי ללכת עם אבא שלי ל**הצגה** שהתקיימה בשבוע שעבר.
	High	השבוע תעלה בתאטרון **הצגה** חדשה ומסעירה.
זמר	Low	הילד הקטן נחשב ל**זמר** מצליח לגילו בזכות ביטחונו העצמי.
	High	להקת הבנים שהופיעה במועדון הזאפה כללה נגן תופים, נגן גיטרה ו**זמר** מוכשר במיוחד.
חול	Low	אפשר למצוא סוגים שונים של **חול** במדבריות הסהרה.
	High	בני בן השנה התפרץ בבכי על חוף הים כאשר מעד ופיו התמלא בחול מים ולכלוך.
חולצה	Low	הייתי צריכה לקחת **חולצה** יותר ארוכה כדי שלא יהיה לי כל כך קר.
	High	אתמול כשהתלבשתי לבית הספר, התאמתי לג'ינס את ה**חולצה** המשובצת הצבעונית.
חוף	Low	אחד הדברים שאני אוהבת במיוחד זה לאכול במסעדה ליד **חוף** הים התיכון בחיפה.
	High	בקיץ כיף ללכת לים ולאכול במסעדה על ה**חוף** מול השקיעה.
חור	Low	אני, אמא ואבא ראינו **חור** ענק בדרך לגן החיות העירוני.
	High	בטעות התיישבתי על מסמר בולט מן הכיסא אשר יצר במכנסיי **חור** גדול ומעצבן.
כעס	Low	הבחור הצעיר כעס מאוד בעקבות החוק החדש שמגביל את תנועתו.
	High	משה לקח ליוסי את החטיף האהוב עליו ממשלוח המנות ולכן יוסי **כעס** על משה.
מאמר	Low	שוק ההון מעניין בעיניי ולכן אני מחפש כל **מאמר** שיוצא בנושא.
	High	החוקרת כותבת ומפרסמת כל שנה **מאמר** חדש ומעניין.
מגבת	Low	בטיול השנתי איבדתי את ה**מגבת** הירוקה שקיבלתי מסבתא שלי.
	High	אין לי סבלנות לתת לכלים להתייבש לבד ואני מעדיפה להשתמש ב**מגבת** ולהחזיר למקום.
מגף	Low	במאה ה-17 הפך ה**מגף** לראשונה לחלק מאופנת הנשים מהמעמד הגבוה.
	High	בחורף, כשיורד גשם, לא כדאי לנעול סנדל או נעל פתוחה אלא **מגף** אטום ומתאים.
מדבר	Low	משפחתי מטיילת ב**מדבר** באופן קבוע לפחות פעם בשנה למשך שלושה ימים כל פעם.
	High	בישראל קיימים אזורי אקלים שונים, למשל אזור יבש וצחיח הנקרא **מדבר** ואזור הררי.
מחסה	Low	לרוב בעלי החיים בטבע יש **מחסה** בחורף מפני הגשם והקור.
	High	בזמן ההפגזה המפקד צעק לכולנו לתפוס **מחסה** ולהישכב מיד.
מפלגה	Low	התעניינתי לשמוע על ההיסטוריה של ה**מפלגה** הקומוניסטית בברית המועצות.
	High	התעניינתי לשמוע על עקרונות ומצע ה**מפלגה** הקומוניסטית בברית המועצות.
מקהלה	Low	בעקבות השיחה, חשבתי להצטרף ל**מקהלה** של בית הספר השנה.
	High	קבוצת זמרים השרים בתיאום נקראת **מקהלה** או חבורת זמר.
משטרה	Low	באופן מפתיע רק היום גיליתי ש**משטרה** לא עוסקת במקרים כאלה.
	High	לאחרונה חלה החמרה בדו"חות מהירות שחילקה ה**משטרה** לנהגים שנלכדים.
משרתת	Low	מרי ביקשה בצורה מנומסת מה**משרתת** בארמון להביא לה כוס מיץ.
	High	בעבר היה נהוג להעסיק באחוזה **משרתת** שתדאג לבית.

(Continued)

Target word	Context type	Sentence
נר	Low	אחותי קנתה לי **נר** כמתנה ליום הולדתי העשירי.
	High	ביום הולדתו של איציק בן השנה הגישה אמא עוגה ואיציק היה צריך לכבות **נר** אחד בלבד.
נשק	Low	שמים את ה**נשק** במקום מוגן כמו כספת או ארון נעול.
	High	אפילו בפורים אסור להביא לבית הספר כלי **נשק** כחלק מהתחפושת.
סירה	Low	כשהכנתי דגם במדעים על הסביבה הימית, הוספתי גם **סירה** קטנה שיכולה לשוט.
	High	על מנת לשוט בנהר הירקון, יש לשכור **סירה** מהחברה שעובדת במקום.
ספרייה	Low	לפי דעתי, ה**ספרייה** היא מקום נפלא לבלות בו אחר צהריים גשום.
	High	המקום השקט ביותר ללמוד בו באוניברסיטה הוא ה**ספרייה** שמשמשת אותי הרבה.
ענף	Low	דני יצא לריצה ובדרך נתקל ב**ענף** שהיה מוסתר מתחת לעלים על הקרקע.
	High	טיפסתי על העץ כי ראיתי קן ציפורים על ה**ענף** הגבוה ביותר.
פרח	Low	כשהייתי במוזיאון ראיתי תמונה של **פרח** צהוב ומיוחד במינו.
	High	כמחווה רומנטית קטפתי לנורית **פרח** קטן, יפה וכחול.
פרי	Low	לארוחת הבוקר היום אכלתי **פרי** מתוק וטעים במיוחד.
	High	כמו תפוח, גם תפוז הוא **פרי** מזין ובריא.
צוות	Low	אמא שלי קיבלה מתנה מאוד יפה מה**צוות** בעבודה לרגל יום ההולדת.
	High	בכל שבוע המחלקה שלי בחברה מתכנסת לישיבת **צוות** בנושא בטיחות.
צמח	Low	ראיתי בחנות שבשכונה שלי **צמח** מיוחד שיובא מדרום אמריקה.
	High	עץ האורן הוא סוג של **צמח** הגדל בעיקר באזורים קרים.
קצף	Low	אפשר בבקשה להזמין קפה עם **קצף** בצד וקצת קינמון.
	High	יש למזוג בירה בכוס מוטה הצידה על מנת שלא יצטבר הרבה **קצף** רב בכוס.
שחף	Low	אני יודעת ש**שחף** הוא ממשפחת עופות המים הניזונים מדגים.
	High	אבא סיפר שהציפור הלבנה הקולנית בים נקראת **שחף** והיא חכמה ומתוחכמת.
שלב	Low	שמעתי שהם הגיעו ל**שלב** מתקדם בחקירת אירוע הגניבה הגדולה שהייתה בבנק.
	High	אחי הקטן שבר שיא במשחק מריו והצליח לעלות ל**שלב** האחרון במשחק.
שריר	Low	לא נמצא קשר ישיר בין מידת הנזק של ה**שריר** לבין תחושת הכאבים.
	High	בזמן אימון בחדר כושר כדאי להישמר שלא ייתפס ה**שריר** בצורה חזקה מדי.

Appendix 3

*The three LME models of Experiment 2 including target type, predicting RTs for the target word, post target word, and final word. Effect size (*b*), standard errors (*SE*), and* t *value (*t*)*

Fixed effects	Target word			Post target word			Final word		
	b	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
(Intercept)	6.54	0.06	111.64***	6.58	0.06	108.00***	6.60	0.06	108.41***
Context (low)	–0.00	0.03	–0.15	–0.02	0.04	–0.54	0.02	0.03	0.80
L2-Availability Factor	–0.17	0.02	–7.18***	–0.18	0.03	–6.98***	–0.19	0.02	–7.89***
AgeGroup (Younger)	–0.2	0.06	–3.23**	–0.22	0.06	–3.39***	–0.22	–0.06	-3.47***
Target Type (Hom)	0.02	0.04	0.58	0.02	0.05	0.46	–0.02	–0.04	–0.54
Context (Low): TargetType (Hom)	–0.01	0.05	–0.25	–0.04	0.06	–0.66	–0.01	0.05	–0.11
L2-Availability Factor: TargetType (Hom)	–0.01	0.02	–0.56	0.01	0.02	0.37	0.02	0.02	0.90
AgeGroup (Younger): TargetType (Hom)	–0.02	0.03	–0,62	–0.04	0.04	–1.14	–0.03	0.04	–0.80
Context (Low): L2-Availability Factor	–0.02	0.01	–2.20	–0.01	0.02	–0.60	0.02	0.01	2.16*
Context (Low): AgeGroup (Younger)	0.01	0.03	0.45	–0.02	0.03	0.67	–0.06	0.03	–2.05*
Context (Low): TargetType (Hom): L2-Availability Factor	0.02	0.02	0.86	0.00	0.03	0.10	–0.01	0.02	–0.66
Context (Low): TargetType (Hom): AgeGroup (Younger)	–0.03	0.05	–0.55	0.04	0.05	0.73	0.03	0.05	0.60
Control variables	*b*	*SE*	*t*	*b*	*SE*	*t*	*b*	*SE*	*t*
WordLength	0.02	0.01	2.59*	0.05	0.01	6.26***	0.06	0.01	7.10***
ArSemanticFluency	–0.05	0.02	–2.34*	–	–	–	–	–	–
TargetLocation	–0.05	0.02	–3.41**	0.03	0.01	–2.64*	–	–	–
SentenceLength	0.03	0.01	2.23*	–	–	–	–0.02	0.01	–2.35*
WordFrequency	–0.02	0.01	–2.35*	–	–	–	–	–	–

(Continued)

Appendix 3. (Continued)

Random effects	*Variance*	*SD*	*Variance*	*SD*	*Variance*	*SD*
Participant (intercept)	0.06	0.25	0.07	0.26	0.07	0.26
Context (low)	0.00	0.06	–	–	–	–
Item (intercept)	0.01	0.09	0.01	0.11	0.01	0.10
Context (low)	0.01	0.10	0.02	0.14	0.01	0.10
L2-Availability Factor	0.00	0.04	0.00	0.06	0.00	0.02
Context (low): L2-Availability Factor	–	–	0.01	0.09	–	–
Residual	0.12	0.35	0.14	0.37	0.14	0.37

Note. Fixed effects reflect simple effects relative to the reference level, when other factors are at their reference level, without correction for multiple comparisons. For main effects, see *F* values in the text.
Sig. codes. 0 "***" 0.001 "**" 0.01 "*" 0.05 "~" 0.1 " " 1.

Appendix 4

L2-Hebrew characteristics of participants in the High and Low L2-availability groups (Experiment 2) and in the L2-Hebrew group (Experiment 1); M (SD)

L2-Hebrew measures	L2 group (Exp. 1)	High L2 (Exp. 2)	Low L2 (Exp. 2)
N	48	65	66
Age of acquisition (in years)	7.6 (1.0)$_a$	6.6 (2.1)$_b$	7.7 (1.8)$_a$
Current exposure (%)	31.3 (16.9)$_a$	48.1 (18.0)$_b$	25.8 (17.7)$_a$
Current use (0–10)	6.8 (1.5)$_a$	6.7 (1.6)$_a$	4.0 (1.9)$_b$
Subjective proficiency (0–10)	8.6 (0.9)$_a$	9.2 (0.9)$_a$	7.4 (1.6)$_a$
Semantic fluency	21.8 (7.9)$_a$	23.7 (6.1)$_a$	17.2 (5.6)$_b$

Note. Means in the same row that do not share subscripts differ at the .05 level based on a one-way analysis of variance test with the Bonferroni corrections for multiple comparisons. See note in Table 1 for details on each measure.