



Fernandez, L. B., Scheepers, C. and Allen, S. E. M. (2023) Cross-language semantic and orthographic parafoveal processing by bilingual L1 German-L2 English readers. *Bilingualism: Language and Cognition*, (doi: [10.1017/S1366728923000536](https://doi.org/10.1017/S1366728923000536))

This is the author version of the work deposited here under a Creative Commons license: <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Copyright © 2023 The Authors

There may be differences between this version and the published version. You are advised to consult the published version if you wish to cite from it: <https://doi.org/10.1017/S1366728923000536>

<https://eprints.gla.ac.uk/301658/>

Deposited on 26 June 2023

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Cross-language parafoveal processing

Cross-language semantic and orthographic parafoveal processing by bilingual L1
German/L2 English readers

Leigh B. Fernandez
University of Kaiserslautern-Landau (RPTU)

Christoph Scheepers
University of Glasgow
Shanley E. M. Allen
University of Kaiserslautern-Landau (RPTU)

Address for correspondence:

Leigh B. Fernandez
University of Kaiserslautern-Landau (RPTU), Faculty of Social Sciences

Erwin-Schrödinger-Straße, Building 57, 67663 Kaiserslautern, Germany
+49(0)631-205-4138
leigh.fernandez@rptu.de

In a recent study, Fernandez et al. (2021) investigated parafoveal processing in L1 English and L1 German/L2 English readers using the gaze contingent boundary paradigm (Rayner, 1975). Unexpectedly, L2 readers derived an interference from a non-cognate translation parafoveal mask (arrow vs. pfeil), but derived a benefit from a German orthographic parafoveal mask (arrow vs. pfexk) when reading in English. The authors argued that bilingual readers incurred a switching cost from the complete German word, and derived a benefit by keeping both lexicons active from the partial German word. In this registered report, we further test this finding with L1 German/L2 English participants using improved items, but with the sentences presented in German. We were able to replicate the non-cognate translation interference but not the orthographic facilitation. Follow up comparisons showed that all parafoveal masks evoked similar inhibition, suggesting that bilingual readers do not process non-cognate semantic or orthographic information parafoveally.

Keywords: parafoveal processing; cross-linguistic influence, eye-tracking; word recognition

Introduction

During reading our eyes do not move smoothly through the text. Rather, we make a series of fixations (in which our eye is relatively still, and we take in information) and saccades (in which we move our eyes to the next fixation location, and do not take in information). When the eye fixates on a word during reading, it is primarily ascertaining information about the word in the fovea - the center of the retina which has the highest visual acuity, approximately 2 degrees of visual angle from the fixation. Importantly, however, it is also ascertaining information about words in the parafovea - the area outside the fovea extending to approximately 5 degrees of visual angle (e.g., Schotter, Angele & Rayner, 2012). The parafoveal area is important as it allows the readers to preprocess upcoming information that is not being directly fixated upon, thereby allowing readers to plan upcoming eye movements.

Evidence for parafoveal processing comes from, for example, the gaze-contingent boundary paradigm (GCB; Rayner, 1975). This paradigm involves masking a critical word of interest and embedding an invisible boundary prior to that region; when the reader makes a saccade across the invisible boundary, the masked region is unmasked. Typically, the participant is unaware of this change because of saccadic suppression (Matin, 1974). For example, in Figure 1 the critical word (lawn) is masked with a series of Xs (therefore blocking the word from being parafoveally processed), and upon making a saccade across the invisible boundary (the dashed line) the Xs are permanently replaced with the critical word lawn (as in B).

<Insert Figure 1 about here>

Using this paradigm, evidence for parafoveal processing has been found in the

form of the parafoveal N+1 preview effect¹. An N+1 preview effect occurs when there are measurable reading differences on a target word (N+1) when that word is available in the parafoveal view of N (the word prior to the critical word, as in B) compared to the same word when it is blocked in the parafoveal view of N (as in A). Traditionally this type of effect has been referred to as a preview benefit but, in line with Vasilev and Angele (2017), we refer to it as the N+1 preview effect.

Parafoveal masks that do not share information with the critical word (e.g., the Xs in Figure 1) are *uninformative* and lead to an increased reading time on the critical word, or N+1 inhibition (e.g., Fernandez, Scheepers, & Allen, 2020; Hutzler, Schuster, Marx, & Hawelka, 2019; Marx, Hawelka, Schuster & Hutzler, 2015; Vasilev & Angele, 2017). In their meta-analysis Vasilev & Angele found a graded effect from uninformative masks, with N+1 inhibition increasing as a parafoveal mask becomes less word-like (i.e., identical word < unrelated word < pseudo-word < string of random letters < string of X's). However, parafoveal masks need not be uninformative; they can also be informative about certain aspects of the word. For example, the parafoveal mask can share orthographic information with the critical word (e.g., *lawc*; in which the first three letters are shared between the mask and critical word) or semantic information with the critical word (e.g., *yard*). Research has shown that when the parafoveal mask and the critical word share the first few letters, there is a large N+1 benefit (see Schotter et al., 2012, for a review).

However, the benefit from having a *semantic* relationship between the parafoveal mask and critical word is less consistent. Both Rayner, Balota, & Pollastek (1986) and Rayner, Schotter, & Drieghe (2014) found no evidence of an N+1 semantic benefit in

English when a critical word (e.g., *song*) was parafoveally masked with a semantically related word (e.g., *tune*). More recent research has found evidence for a semantic N+1 benefit in English, in certain situations. For example, a benefit is found when parafoveal masks are synonyms as opposed to semantic associates (Schotter, 2013; Schotter, Lee, Reiderman, & Rayner; 2015), when some individual differences are taken into account (e.g., reading skill; Veldre & Andrews, 2016a), and when the N+1 word is capitalized thus increasing visual attention to the word (Rayner & Schotter, 2014). Additionally, a preview benefit has been found in English for parafoveal masks that are plausible within the sentence context, even if the mask and the critical word do not share information (Schotter & Fenell, 2019; Schotter & Jia, 2016; Veldre & Andrews, 2016b; Wakeford & Murray, 2013). This suggests that, at least in early reading measures, semantic preview benefits may arise from the fit between the context of the sentence and the plausibility of the mask.

Semantic N+1 effects have been more readily found in languages such as Chinese (e.g., Tsai, Kliegl, & Yan, 2012; Yan, Zhou, Shu, & Kliegl, 2012; Yang, Wang, Tong, & Rayner, 2012), Korean (Yan, Wang, Song, & Kliegl, 2019), and German (e.g., Hohenstein & Kliegl, 2014). Yan et al. (2009) has proposed that properties of the written system dictate which type of information is parafoveally processed. In contrast to alphabetic scripts, meaning in the Chinese writing system is mapped more closely to orthography than to phonology. Additionally, most Chinese words consist of two characters, meaning that the N+1 word is typically viewed with higher visual acuity than in alphabetic scripts where words have more characters. Indeed, Tsai et al. (2012), Yan et al. (2019), and Yang et al. (2012) have found evidence for semantic N+1 benefit in

Chinese. Yan et al. (2019) also found evidence for a semantic N+1 benefit in Korean. Interestingly, characters in the Hangul script used for Korean represent a syllable, and words consist of only a few characters, thus providing readers with higher N+1 visual acuity. Additionally, the spelling to sound correspondence is a relatively clear in Korean, making lexical access more straightforward (unlike English).

In German, Hohenstein & Kliegl (2014) found a semantic N+1 benefit across three GCB studies. In neutral sentence contexts they found an N+1 preview benefit when a critical word (e.g., *Wolle* (wool)) was masked with a related word (e.g., *Seide* (silk)), but not with an unrelated word (e.g., *Seife* (soap))². They also found a preview benefit for related words compared to unrelated words regardless of capitalization (the first letter of German nouns is capitalized), suggesting that this benefit was not just the result of increased parafoveal visual salience of German nouns due to capitalization. The clear semantic N+1 benefit seen in German may result from the relatively clear spelling-to-sound correspondence in German (see for example, Laubrock & Hohenstein, 2012), which may make lexical information easier to extract relative to languages like English (with an opaque spelling-to-sound correspondence).

From the L1 evidence previously outlined, the semantic N+1 benefit seems to arise under certain conditions, and with particular languages. However, most of our understanding of reading and language processing is based on monolingual speakers, despite the reality that more than half of the world speaks more than one language (Marian & Shook, 2012). While research investigating bilingual language processing has become more prominent, there are still relatively few studies that systematically investigate classic oculomotor behavior in L2 speakers (e.g., skipping behavior and

parafoveal processing). We seek to form a more comprehensive picture of L2 reading behavior by investigating L2 parafoveal processing. In the current study, we expand on a previous “seemingly English” parafoveal processing study with L1 German/late L2 English speakers (Fernandez, Scheepers, Allen, 2021) by presenting similar but translated items in a “seemingly German” parafoveal processing study with L1 German/L2 English.

To the knowledge of the authors, there are only a handful of published studies investigating L2 parafoveal processing. Three of these studies have focused on effects that are not semantic. First, Jouravlev and Jared (2018) found that English/Russian bilinguals reading in English can make use of orthographic and phonological information from Russian parafoveal masks presented in Cyrillic. Second, Vaughan-Evans, Liversedge, Fitzsimmons, and Jones (2020) found that Welsh/English bilinguals reading in English are sensitive to English parafoveal previews that adhere to Welsh soft mutation morpho-syntactic rules (relative to those with an aberrant soft mutation). Finally, Fernandez, Scheepers, & Allen (2020) found that L1 English and L1 German/L2 English readers showed a graded N+1 interference from uninformative masks, with the interference increasing the less “word-like” the masks became.

A number of other L2 studies focused on semantic N+1 effects, but their results are somewhat mixed. Three studies show no semantic preview effect. Altarriba, Kambe, Pollatsek, and Rayner (2001) investigated parafoveal processing in a group of Spanish/English bilinguals while reading in both Spanish and English. They found that there was an N+1 benefit with cognate translation masks (*cream/crema* (cream)) and pseudo-cognate masks (*grass/grasa* (grease)) relative to non-cognate translation masks (*strong/fuerte* (strong)). This suggests that the bilingual readers only derived a benefit

when the mask shared phonological and orthographic features with the target word, and not when the mask shared only semantic information. More recently, this lack of N+1 benefit from non-cognate translations in Spanish/English bilinguals was replicated with different materials (Hoversten & Traxler, 2020), again suggesting that bilingual readers of alphabetic languages do not access translation/semantic information parafoveally. A similar result was found with late L2 Chinese speakers (L1 Korean) reading in Chinese; Wang, Zhou, Shu, and Yan (2014) found an N+1 preview benefit when the mask and critical word shared orthographic features but not when they shared semantic features. Interestingly, Wang et al. included a measure of L2 proficiency and found that more proficient participants were more efficient at extracting parafoveal information. They argued that less skilled readers had to devote more of their cognitive resources to the foveal word, which left less cognitive resources to devote to the processing of the parafoveal word. The role of proficiency was not tested in Altarriba et al., nor it is clear whether their participants were early bilinguals, late bilinguals, or heritage speakers. Proficiency was measured in the Hoversten and Traxler paper, with all participants showing similar proficiency in both English and Spanish respectively, but proficiency level was not included in their statistical models.

Three additional studies investigating semantic L2 parafoveal processing, however, did find evidence for an L2 N+1 semantic benefit. The first investigated late L2 Chinese speakers (with a Korean L1) reading in a seemingly Chinese task where, however, the parafoveal masks were presented in Korean (Wang et al., 2016). They found an N+1 semantic benefit during reading for both cognate and non-cognate semantic masks, suggesting that the benefit was not purely orthographic in nature, but rather

semantic. They argue that this benefit could be the result of the relatively transparent spelling-to-sound correspondence of Korean (which has a similar orthographic depth to German). The second study found evidence for an L2 N+1 semantic benefit for Basque/Spanish bilinguals, although with some methodological differences from previous studies (Antúnez et al., 2021). This study used Fixation-Related Potentials to measure brain activity during reading. In addition, participants did not read sentences, but rather viewed two words during a Spanish reading task; word N+1 was masked when the participant was fixating on word N, and was replaced with the critical word when the participant made a saccade to word N+1. Antúnez et al. found evidence for a facilitation from a non-cognate Basque translation parafoveal mask relative to an unrelated Basque parafoveal mask. This suggests that, even in alphabetic languages, word recognition is not language-selective. The third study to find some evidence for an N+1 semantic benefit with L2 speakers investigated L1 speakers of English and L2 speakers of English (with an L1 of German) reading in an English reading task (Fernandez, Scheepers, & Allen, 2021). Given that the current study aims to extend the Fernandez et al. study, we devote more attention to this study.

Fernandez et al. (2021) tested orthographic and semantic parafoveal processing for both L1 speakers of English (with no German knowledge) and late L2 speakers of English with German L1 (mean age of English acquisition was 10.2 years old), using the GCB. They presented neutral context sentences in English (The man found the arrow in a tree) with the critical word (arrow) masked with six different masks types: an identical word mask (arrow), an English orthographic mask (arrzm), an English string mask (xezne) in which ascending and descending letters were matched, a

German translation (pfeil), a German orthographic mask (pfeck), or a German string mask (ylctb) in which ascending and descending letters were matched. Critically, the German translation mask was carefully matched in length, and the English meaning could not be guessed based on the German word (as guessed by 250 native English speakers who had never learned German or any Germanic language other than English; materials taken from Friel & Kennison, 2001). Therefore, L1 speakers of English were not able to extract any useful information from the German masks, while L2 speakers of English with German L1 would potentially be able to extract semantic and/or orthographic information from the German masks.

In their study, they made several comparisons, each with a specific hypothesis; see Table 1 for all comparisons, hypotheses, and findings. First, they hypothesized similar reading times between the identical (arrow) and the English orthographic mask (arrzm). They found no difference between the two mask types, suggesting that both groups showed orthographic N+1 facilitation. Second, they hypothesized greater reading durations for the English string mask (xezne) compared to the English orthographic mask (arrzm). They found no such inhibition, which they argued may be the result of the masks being very similar (i.e., they matched in ascending and descending letter shapes), making the display change harder to detect and thus limiting inhibition. Third, they hypothesized an N+1 semantic facilitation (with similar reading times) between the German translation mask (pfeil) compared to the identical mask (arrow) for the L2 English speakers with German L1, and an inhibition for the L1 English speakers. They found that both language groups showed an inhibition in the German translation mask condition (in terms of increased reading durations in the German mask condition), which

they argued may reflect a switching cost for the L2 speakers and may reflect that the mask was less word like for the L1 group. Fourth, they hypothesized an N+1 orthographic benefit for L2 speakers when comparing the German mask (pfeil) to the German orthographic mask (pfeck) and no difference for L1 speakers since both should be non-words. They found an N+1 orthographic benefit from the German orthographic mask, with L2 speakers showing shorter reading times after the pfeck relative to pfeil conditions, while L1 readers showed no such difference³. Fifth, they hypothesized an N+1 inhibition for the German string mask (ylctb) compared to the German orthographic mask (pfeck). They found an N+1 inhibition for both groups (with greater reading durations for the German string mask), suggesting that both groups showed an inhibition due to the German string mask being less word like. In terms of individual differences, they hypothesized that participants with higher quality of lexical representation (tested via spelling skill) and with higher proficiency (tested via English morpho-syntactic knowledge) would be more efficient at extracting parafoveal information and would have shorter reading times. They found this was the case for quality of lexical representation but not for proficiency level.

<Insert Table 1 about here>

According to the Bilingual Interactive Activation Plus (BIA+) model of word recognition (Dijkstra & Van Heuven, 2002), the bilingual lexicon is non-selective and integrated between languages. Within this framework, Fernandez et al. (2021) argued that when a late L2 English speaker (with an L1 of German) reads in English, but views a complete German word in the parafovea (e.g., pfeil), the German language node is

activated and the English language node is inhibited. This is further compounded by the fact that the word ultimately fixated on (and the continuation of the sentence) is in English, which could incur an additional language switching cost leading to N+1 inhibition. However, when viewing an orthographic mask (but not a real word in German or English, e.g., p f e x k) in the parafovea, both German and English lexicons remain active, and lexical and semantic candidates are narrowed down in both English and German, thus leading to quicker access to `arrow` upon foveal viewing. This leads to N+1 facilitation and, given that both English and German remain active, few if any language switching costs are incurred. It is important to note, however, that L1 speakers showed a somewhat similar pattern (a decrease in reading time in the German orthographic condition (p f e x k) relative to the German translation condition (p f e i l)) and, while this effect did not reach significance, there is no reason this pattern should emerge.

The unexpected finding that the German translation mask was inhibitory while the German orthographic mask was facilitatory suggests that L2 readers are engaging in complex processing and are able to use parafoveal information to narrow down lexical candidates in a language non-specific way (much like they do for foveal processing). There are only a handful of studies investigating parafoveal processing in L2 readers and, to our knowledge, theories of eye-movements during reading do not typically consider L2 reading behavior (despite the fact that these theories account for reading behavior in other groups such as children and older adults). However, previous L2 research does not seem to support the idea that L2 readers are able to engage in such complex reading behaviors. In a recent study, Cop, Drieghe, & Duyck (2015) report a comprehensive description of

both bilingual and monolingual reading behavior during the reading of a novel. Dutch L1 (late English L2) speakers and monolingual L1 English speakers read the same novel in its entirety while having their eyes tracked; the bilingual group read half in Dutch and half in English, while the monolingual group read the novel entirely in English. Analysis within the same bilingual participant (the same participant was compared reading in their L1 and their L2) showed several effects at the sentence level when reading in an L2 relative to reading in an L1: average sentence reading times were longer, there were more fixations per sentence, average fixation durations were longer, average saccade lengths were shorter, and the likelihood of skipping a word was lower. Cop et al. argue that this pattern of behavior is similar to that of 7- to 11-year-old L1 readers, and that there is an overall slowing of lexical access that drives differences between L1 and L2 reading behavior. This would suggest that L2 readers should take longer to read the foveal word and as a consequence have fewer resources to invest in parafoveal processing.

In terms of the Fernandez et al. (2021) study, it is possible that the unexpected results (inhibition from the German translation mask but facilitation from the German orthographic mask) from Fernandez et al. (2021) are in part due to the relatively low power in their study (51 L1-speakers / 51 L2-speakers, each reading 4 items per condition). Therefore, in the current registered report we increased the power with 100 L1 speakers of German (who were late English L2 speakers) reading 18 items per condition. In addition, we expanded on the Fernandez et al. (2021) findings by using similar materials but translated into German. This allowed us to see whether the same pattern of results holds true regardless of the language frame in which the masks were embedded. Finally, we made several changes to the design⁴ that we believe allowed us to

more directly test the extent to which readers are able to use L2 parafoveal information during reading (see Table 2).

<Insert Table 2 about here>

First, we removed the two string mask conditions (e.g., `xezne` and `ylctb`) as we did not have any additional theoretical questions about these conditions. We also replaced the German orthography condition (`pfeck`) with an English pseudo-word mask (`Clain`) with no orthographic overlap. It is possible that the facilitation from the orthographic-word mask (`pfeck`) was due to the fact that this mask happens to be “word-like”, thus evoking less interference (see Fernandez et al., 2020; Veldre & Andrews, 2017) rather than due to orthographic and semantic overlap. This allowed us to directly test whether the facilitation from the orthographic mask was due to orthographic overlap or merely to the fact that this mask is “word-like”.

Second, we added a comparison of the English translation to the two non-word masks (the English pseudo-word mask `Clain` and the English orthography mask `Arrzm`). As mentioned previously, we can expect an N+1 inhibition effect from uninformative masks in the parafovea, and that this inhibition becomes greater the less “word-like” the mask becomes for L1 speakers. Interestingly, Fernandez et al. (2021) reported the opposite for viewing words in another language parafoveally: they found a greater N+1 inhibition effect when a complete, real-word translation equivalent of the target word (e.g., the mask `pfeil` for the target word `arrow`) was shown parafoveally to L1 German/L2 English speakers when reading in English (L2), relative to a non-word like `pfeck` (which only exhibits orthographic overlap with the translation equivalent).

This pattern suggests that bilingual readers benefit from a non-word relative to an L1 word in the parafovea, which suggests that there is a switching cost from the translation word, but not from the (orthographically overlapping) non-word. This finding would be in direct contrast to what is found with L1 speakers (inhibition from non-words) and would further support the idea that L2 speakers are able to extract semantic information from the parafovea.

Third, it is possible that the differences in Fernandez et al. (2021) when comparing the critical word to the German translation (*arrow* to *pfeil*) were not due to the semantic activation of *pfeil*, but rather to the fact that *arrow* was the target word (while *pfeil* was not). In order to test this possibility, in the current study, a semantically unrelated but orthographically related English word condition is added, that is, an English word that shares the first few letters with the English translation condition, but is semantically unrelated to the English translation word; e.g., *Array*. By comparing *Arrow* to *Array* in the current study we test whether differences were due to the fact that there was an English word in the parafovea that shares semantic information with the target word, or whether differences were merely due to the fact that there was an English word in the parafovea. While we deviated from the original design of Fernandez et al., these changes still allowed us to replicate with greater power, as well as expand on whether bilingual readers are indeed able to derive a semantic benefit from an orthographic match in their L2.

Before outlining the current study, we briefly discuss the asymmetry associated with tasks that involve single word language switches and research with inter-sentential

language switches. The asymmetry associated with switch costs and translation priming is likely to have important implications for the current study. Research has found asymmetric switch costs; that is, the cost of switching from L2 to L1 is greater than switching from L1 to L2 (for reviews see Bobb & Wodniecka, 2013; Declerck & Philipp, 2015). Additionally, research with non-cognate translation masked priming has shown that priming effects occur from both L1 to L2 and L2 to L1. However, the effect is smaller from L2 to L1 than L1 to L2 (e.g., Schoonbaert, Duyck, Bryspaert, & Hartsuiker, 2009). In the Fernandez et al. (2021) study, L2 speakers read in their L2, parafoveally viewed translation masks in their L1, but ultimately viewed the critical word in their L2, which in theory should be symmetric (L1 to L2, at least in terms of the critical word Pfeil to arrow). In contrast, in the current study L1 speakers read in their L1, parafoveally view translation masks in their L2, but ultimately view the critical word in their L1, which would theoretically be asymmetric (L2 to L1, at least in terms of the viewing of the critical word Arrow to Pfeil). Therefore, it stands to reason that in the current study there might be a greater likelihood for inhibition between the English translation mask and the German identical mask relative to the Fernandez et al. study. However, we believe both studies will incur similar costs, given that both studies have overall switches from L1 to L2 and from L2 to L1 (Fernandez et al. (2021): English (L2) to German (L1) to English (L2); current study: German (L1) to English (L2) to German (L1)).

Research investigating bilingual sentence processing in which there is a language change within a sentence has shown that the language of the sentence preceding the switch is not enough to suppress activation of the non-target language (see Van Assche,

Duyck, & Hartsuiker, 2012; Lauro & Schwartz, 2017). However, whether contextual information can constrain lexical candidates is somewhat less clear. In the face of sentences with low contextual constraint, it seems that bilinguals access both languages (Duyck et al., 2007; Libben & Titone, 2009; Pivneva, Mercier, & Titone, 2014; Schwartz & Kroll, 2006; van Assche et al., 2011; van Hell & de Groot, 2008). However, in the face of high contextual constraints, language-specific lexical access may be constrained (Lauro & Schwartz, 2017). Given that our current study investigates low contextually constraining sentences (see the methods section), we will not discuss this further.

Previous research shows that sentence level processing and language switches impact bilingual lexical access. However, we believe that drawing too many parallels between word recognition research (which primarily employs masked priming and lexical decision tasks) and the current study (which employs parafoveal processing during reading) is not particularly useful since the paradigms are vastly different. The former typically involves single word presentation and a lexical decision task, while the latter involves naturalistic sentence reading within one language (and participants are not aware of any cross-language changes throughout the whole study). We also believe it is not fruitful to draw too many parallels with inter-sentential language switch research which typically involves the participant reading a sentence in one language, an obvious language switch, and sentence continuation. In the current study a language switch also occurs, but the switched word is not perceived with foveal awareness because it is never directly fixated upon. Additionally, most previous studies have investigated the role of cognates and interlingual homographs in which the critical word shares at least some orthographic characteristics across both languages (e.g., Duyck, Van Assche, Drieghe, &

Hartsuiker, 2007; Elston-Güttler, Gunter, & Kotz, 2005; Schwartz & Kroll, 2006; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). However, in the current study the critical word shares only semantic information across languages and the participant is not aware that a language switch has occurred, meaning that any N+1 benefit would arise solely from semantic information. Therefore, we believe this study has the potential to provide new insight into the process of bilingual lexical access.

Current study

As outlined above, in the current registered report we aimed to overcome some of the potential confounds of the Fernandez et al. (2021) study, and expand on their findings by using more theoretically motivated masks (while at the same time increasing power) to test whether the pattern of inhibition from a translation mask and facilitation from a purely orthographic mask remains. In order to test the generalizability of response patterns in L1 versus L2 (further corroborating or indeed disconfirming the Fernandez et al. findings) we presented the sentence frames in German. In the Fernandez et al. (2021) study, L1 German speakers read sentences in a “seemingly English” (L2) task, using both English and German-based masks, to test whether readers are able to make use of translations from their L1 while reading in their L2. In the current study, we tested L1 speakers of German reading sentences in a “seemingly German” task, i.e., they read sentences in their L1. However, we employed both German-based and English-based masks to test whether L1 speakers make use of translations from their L2 while reading in their L1. Additionally, the original Fernandez et al. (2021) study tested the role of English proficiency and quality of lexical representation, given previous findings which suggested that individuals are more efficient at extracting parafoveal information if they

have a higher quality of lexical representation (QLR) (Veldre & Andrews, 2014; L2 speakers – Whitford & Titone, 2015) or higher proficiency (Wang et al., 2014). While Fernandez et al. (2021) found no impact of English proficiency on any of their measures, they did find that reading durations decreased as spelling scores increased. In the current study, we also assessed the role of QLR and English proficiency. Consistent with Fernandez et al., we tested English proficiency using a test of morphosyntactic ability (Oxford Placement Test, Part A). Given the relatively clear spelling-to-sound correspondence in German, we opted to use the German LexTale test as a measure of QLR (Lemhöfer & Broersma, 2012).

Given the potential confounds outlined previously, and the rather low power in Fernandez et al. (2021), the current study used a modified design and materials. Firstly, we removed the German orthography condition, the English string, and German string conditions, as we did not have additional theoretical questions about these conditions. Secondly, we added two new conditions: the first is an English pseudo-word condition, which allowed us to directly test whether it is the orthographic overlap between the orthographic mask and the translation that drives the facilitation, or whether it is the fact that the orthographic mask is merely more “word-like” that drives the facilitation. The second is an English word condition (the word shares orthographic overlap but no semantic overlap with the English translation), which allowed us to directly test whether any inhibition for the English translation mask (relative to the Identical mask) arises merely because there is an English word in the parafovea, or whether it is due to the semantic overlap between the Identical word and the English translation. Additionally, we compared the English translation mask to the two non-word masks (the English

orthography and the English pseudo-word masks) to test whether the translation preview lead to greater N+1 inhibition relative to the two non-word previews (this would also provide additional evidence of a parafoveal switch cost).

We summarize the hypotheses for the current study in Table 3.

<Insert Table 3 about here>

Methods

Participants

A total of 107 L1 speakers of German, who were late L2 learners of English, were recruited from the RPTU-Kaiserslautern, Kaiserslautern Germany. Based on our registered exclusion criterion, we removed two participants who were raised bilingually and four participants who scored 50% or below on the English proficiency test. An additional participant was removed due to a technical issue. The remaining 100 participants met our registered requirements: they were native speakers of German (between 18 and 45 years old) who grew up in Germany with only German spoken at home, and were not exposed to a second language before the age of 6 years; see Table 4. Additionally, no participants reported a language disorder and all reported having normal or corrected-to-normal vision. Their mean English acquisition age was 9.30 years of age ($SD = 1.79$; range: 19-44 years of age).

<Insert Table 4 about here>

Apparatus

Eye movements were recorded using an EyeLink 1000 sampling at 1000 Hz. Stimuli were presented on a Dell P1130 19" flat screen cathode ray tube (1024 X 768 resolution; 120 Hz refresh rate) with approximately 2.65 characters subtended 1° of

visual angle. The refresh rate yielded a mean display change of 8.32 msec ($SD = 2.42$; range: 0 – 13.08 msec). Participants sat approximately 65 cm away from the monitor with their head stabilized using a chin rest. Viewing was binocular but only the right eye was recorded.

Materials

The experiment consisted of 4 practice trials plus 90 critical items presented in German. An example of the critical stimuli can be seen in Table 2 (placed here again for convenience). Overall, the design was one-way with five levels (mask type: Identical, English translation, English orthography, English word, English pseudo-word) with mask type manipulated within subject and item.

<Insert Table 2 about here>

The stimuli were taken from Fernandez et al. (2021) and translated into German. Sixty-six additional items were created to match the original Fernandez et al items. The critical stimuli start with a proper noun (Sebastian) or an article and a noun denoting a human or group of humans (e.g., der Mann/die Damen), followed by a verb in the past tense (fand), an article (den) with an embedded invisible boundary (l), the critical target word (Pfeil) and parafoveal mask, and a spill over region (in einem Baum).

The critical target word and English translation were non-cognates. The German target word and its English translation did not share the same first letter (to avoid orthographic overlap) and were matched for length within item (3-11 characters). Given that the current study was run in German, the first letter of each mask was capitalized (unlike in Fernandez et al., 2021). However, if the German word contained a letter with an umlaut or an Eszett (ß) it is respectively presented with “e” following the vowel (e.g.,

Jäger → Jaeger (hunter)) or presented with “ss” (e.g. Fluß → Fluss (river)). The German target words were relatively common (mean log₁₀ frequency of 1.15 per million) according to the TenTen German Web Corpus (Jakubíček, Kilgarriff, Kovář, Rychlý, & Suchomel, 2013). The English translations and English orthographically (but not semantically) matched words were also reasonably frequent according to the TenTen English Web Corpus (Jakubíček, et al., 2013) (mean log₁₀ frequency of 1.35 per million and 0.85 per million respectively). Both the English translation and English orthography masks shared the first 2-4 characters; the remainder of the letters in the English orthography mask were replaced with letters that matched in ascension and descension which overall formed a non-word (e.g., Arrow → Arrzm). The Wuggy software was used to create pronounceable length-matched English pseudo-word masks (Keuleers & Brysbaert, 2010)⁵. See OSF (https://osf.io/vezfn/?view_only=c77210941cf94bd99c36326613b0efc2) for a complete list of critical items and masks (including the amount of letter overlap).

To ensure there was no strong semantic overlap between the English translation and English orthography words (e.g., Arrow and Array), we took the semantic similarity cosine scores from HAL (Hyperspace Analogue to Language; Lund & Burgess, 1996) and LSA (Latent Semantic Analysis; Landauer, Foltz, & Laham, 1998) for each pair of English words per item⁶. While HAL and LSA use somewhat different methods, their scores were correlated ($r = 0.8$) across the word pairs, so they were combined into an average score $((HAL+LSA)/2)$ for greater reliability. Cosine similarities range from -1 (semantically opposite) to 1 (semantically identical); 0 means no semantic relationship whatsoever. The mean and median of on the combined cosine metric (0.26 and 0.25

respectively) suggest low semantic relatedness between the two English words per item on average; we include this metric as a covariate in our model comparing these two conditions. See OSF for a complete list of the items and masks (including the cosine semantic similarity information).

An offline sentence completion task was used to test whether the target word was expected given the sentence context. Participants (20 native German speakers) were provided with the beginning of the sentences (“*Der Mann fand d*”) and were asked to complete the sentence with a grammatically correct article and noun (the *d* was provided to ensure the participants always used an article). For 88 items, the predictability of the target word was very low (0.23%). For the remaining two items, the items were completed with the critical word more than 20% of the time, and were replaced. An additional offline task of the same format was created with two additional items, and was completed by 11 L1 German speakers. For the 2 items, no participant completed the item with the target word. Therefore, it is safe to assume that participants have no prior expectations when encountering the critical word in the sentence for the first time.

Procedure

To ensure that participants believed that they are taking part in a German reading task, the first portion of the testing was conducted in German. After providing informed consent (the current study has been approved by the local Ethics Committee), participants completed the computerized German LexTale test, and then the eye-tracking main task. Following this, there was a mandatory break, and to ensure their English proficiency, the remainder of the testing took place in English: participants completed the Language Social Background Questionnaire (Anderson et al., 2018) and took an English

proficiency test assessing morpho-syntactic knowledge (Oxford Placement Test, Part A). Participants were paid 10 Euro or given course credit for their participation.

The eye-tracking main task began with a standard EyeLink 9-point calibration. The eye-tracking study was self-paced, such that participants could take a break as needed (with subsequent recalibration), and were recalibrated following an obligatory break halfway through the experiment. The study instructions were presented on the screen and verbally explained. Participants were asked to read silently for comprehension and instructed that after some of the sentences there will be a yes/no question about the sentence they just read, which they can answer by pressing “x” for yes and “m” for no on a standard German keyboard. One third of the sentences were followed by a comprehension question; half of the questions had a correct answer of “yes” and half had a correct answer of “no”. See OSF for a copy of the data source including comprehension questions (https://osf.io/vezfn/?view_only=c77210941cf94bd99c36326613b0efc2). Participants were also given a secondary ‘glitch detection’ task combined with the following cover story: *“Because we have to use a less than optimal monitor for this study, you may very occasionally experience a ‘glitch’ during text presentation. Such glitches are indeed not intended. Hence, in the unlikely event that you have noticed something weird happening on screen, please let the experimenter know after the affected trial, so that we’ll be able to remove that trial from analysis.”* Using this secondary task gave us some idea of conscious display-change detection per trial, while the cover story ensures that there was not too much emphasis on detecting display changes – indeed, the primary task of the study remained reading for comprehension.

Trials began with a drift correct (x-coordinate: 54, y-coordinate: 280) which corresponded to the first letter in the critical sentence. The trial proceeded after the participant fixated for 150 ms within a half-degree region surrounding the drift correct; if this criteria was not met, the participant was recalibrated. All sentences were presented in a monospaced font (Courier New) on one line in a font size of 15. When the participant finished reading the sentence, they pressed the space bar to advance to the yes/no question. The yes/no comprehension question (along with instructions to press ‘x’ for yes and ‘m’ for no below the question) was then displayed at the same coordinates as the critical sentence and in Courier New with a font size of 15. After pressing the x or m key, the next trial began with the drift correct. Presentation order was random for each participant.

Analysis

The following pre-processing steps, exclusion criteria, and analysis were registered (see https://osf.io/vezfn/?view_only=c77210941cf94bd99c36326613b0efc2 for the accepted registered steps). Prior to analysis, trials were excluded when (1) a saccade within the critical region of interest could not be calculated because it ended outside of any interest area (18.74%), (2) a saccade crossed the invisible boundary but landed on a word prior to the critical word (i.e., j-hook; 2.52%), (3) a fixation occurred on the critical word before the boundary change was triggered (0%), (4) the display change time was greater than 15ms (0%), and (5) the participant noticed something strange during text presentation (0.98%).

Comprehension accuracy was 91.10%; we did not expect any cross-condition differences and therefore do not discuss accuracy further. Two duration-based measures

(first fixation duration and gaze duration) and a probability-based measure (skipping rate) were analysed. First fixation duration (FFD) is the duration of the first fixation within the critical region, while gaze duration (GD) is the sum of the duration of fixations on the critical word before the eye moves out of the region. Prior to statistical analysis, FFD and GD data were trimmed, such that all fixations under 80ms or over 1000ms (for GD) or 1200ms (for GD) were removed for FFD (see Appendix A for percent of data removed and the remaining data points per condition). While our primary hypotheses focus on the duration measures, we are also interested in skipping rate to see whether these readers are able to use second language bottom-up visual information from the parafoveal area to inform oculomotor planning. We calculated skipping rate as the percentage of trials in which the critical target word was skipped during the first pass (see Appendix A). This measure is binary, with the critical word being skipped (1) or fixated on (0) during the first pass.

Duration measures were analysed using generalized linear mixed-effects models (GLMM) in R (R Core Team, 2019) using the *lme4* package (Bates, Maechler & Bolker, 2019), and p-value estimates were derived from the *lmerTest* package (Kuznetsova, Brockhoff & Bojesen, 2019). Given the skew in duration measures, the identity link function with a Gamma distribution was used (see Lo & Andrews, 2015) which takes into account the positive skew of duration data. The alpha threshold for the duration measures was set to $0.05/2=0.025$ given the increased potential for Type I error with multiple comparisons (see von der Malsburg & Angele, 2017). Skipping rate was analysed using a binary logistic GLMM. The skipping rate model was fit in the same way as the duration models, but the additional centred covariate of mask length (in number of characters) was

added to the model as a main effect. Given that the skipping rate measure is exploratory, the alpha threshold was set to $0.05/3=0.016$.

A custom a-priori contrast matrix (Schad, Vasishth, Hohenstein, & Kliegl, 2020) compared the following across two models: 1) Identical vs. English translation, 2) English translation vs. English orthography, 3) English translation vs. English pseudo-word, and 4) English translation vs. [English orthography + English pseudo-word], and 5) English translation vs. English word (see https://osf.io/vezfn/?view_only=c77210941cf94bd99c36326613b0efc2).

The first model consisted of the fixed effects of mask type (5 levels: Identical, English Translation, English Orthography, English Word, and English Pseudo-word), LexTale score, proficiency score (with the two latter being scaled to reduce collinearity), and their interaction, and included the first 4 contrasts. The second model was identical to the first model, but included the additional scaled cosine semantic similarity value, and included the 5th contrast comparing the English translation and English word mask (given that these are the two relevant conditions for the semantic similarity value). Across both models, the random effects structure was maximally specified (Barr, Levy, Scheepers, & Tily, 2014) with crossed random effect of participant and items. For all models, the maximal model did not converge. Therefore, as outlined in our registered statistical approach, the random slopes for the individual difference measures (LexTale Score and proficiency score) were removed (Barr et al., 2013, found that it may not be essential to include random slopes for control predictors). In all cases this model converged.⁷

Below t/z and p values are reported, see Table 5 for mean values and standard deviations, see Appendix B for additional model information, and see

(https://osf.io/vezfn/?view_only=c77210941cf94bd99c36326613b0efc2) for collected data and statistical code.

<Insert Table 5 about here>

Results

First fixation duration

For a visualization of FFD see, Figure 2. The first model revealed a main effect of the Identical condition vs. the English translation condition (Pfeil vs Arrow; $t=4.71$, $p<.001$), with FFD increasing in the English translation condition relative to the Identical condition. Both the first ($t=3.28$, $p<.01$) and second ($t=3.04$, $p<.01$) model revealed an interaction of German QLR (measured via the German LexTALE) and English proficiency (measured via the OPT). To explore this interaction, we split the LexTALE score into two groups (low LexTALE: $n=52$ (range=55.00-87.50)/ high LexTALE: $n=48$ (range=88.75-97.50)), see Figure 3. We found that individuals with lower German QLR scores showed decreased FFD as English proficiency increased, while individuals with high German QLR score showed increased FFD as English proficiency increased. No further effects reached significance (see Appendix B).

<Insert Figure 2 about here>

<Insert Figure 3 about here>

Gaze duration

For a visualization of GD see, Figure 4. The first model revealed a main effect of the Identical condition vs. the English translation condition (Pfeil vs Arrow; $t=3.22$, $p<.01$) with GD increasing in the English translation condition relative to the Identical

condition. Both the first ($t=2.62, p<.01$) and second ($t=2.34, p<.025$) model revealed an interaction between German QLR and English proficiency. As was done for FFD, we split the LexTALE to explore the interaction, see Figure 5. Again, we found the same pattern: individuals with lower German QLR scores showed decreased GD as English proficiency increased, while individuals with high German QLR score showed increased GD as English proficiency increased. No further effects reached significance (see Appendix B).

<Insert Figure 4 about here>

<Insert Figure 5 about here>

Skipping rate

For a visualization of skipping rate see Figure 6. No significant effects were registered in this variable (see Appendix B).

<Insert Figure 6 about here>

Discussion

Currently, the semantic N+1 benefit is somewhat elusive in research on both monolingual and bilingual speakers, with some languages and language combinations showing a benefit (e.g., Chinese: Tsai, Kliegl, & Yan, 2012; Yan, Zhou, Shu, & Kliegl, 2012; Yang, Wang, Tong, & Rayner, 2012; English: Schotter, 2013; Schotter et al., 2015; Schotter & Fenell, 2019; Schotter & Jia, 2016; Wakeford & Murray, 2013; Korean: Yan, Wang, Song, & Kliegl, 2019; German: Hohenstein & Kliegl, 2014; Spanish/Basque bilinguals when using Fixation-Related Potentials and using word presentation: Antúnez et al., 2021; Chinese/Korean bilinguals when L1 Korean speakers read in their L2

Chinese but are presented with a mask in their L1: Wang et al., 2016) and others showing no benefit (e.g., Finnish: Hyönä & Häikiö, 2005; English: Rayner, McConkie, & Zola, 1980; Chinese/Korean bilinguals: Wang, Zhou, Shu, & Yan, 2014; Spanish/English bilinguals: Altarriba, Kambe, Pollatsek, & Rayner, 2001; Hoversten & Traxler, 2020). Given this discrepancy and the general dearth of research on L2 parafoveal processing, in this registered report, we tested L1 German/late L2 English speakers in a “seemingly German” task with English translation masks. This study provides unique data and insight into cross-language N+1 semantic preview effects given that L1 German speakers show an N+1 semantic benefit (Hohenstein & Kliegl, 2014) and that bilingual readers are able to derive an N+1 semantic benefit from an L1 translation mask when reading in their L2 (L1 Korean/late L2 Chinese speakers; Wang et al, 2016). In addition, this study provides evidence to whether late L2 speakers are able to make use of semantic parafoveal information during reading in a seemingly monolingual task.

In the current registered report, we built upon and expanded the original Fernandez et al. (2021) study by increasing power, conducting the study in German, and adding additional conditions that allowed us to more directly test whether L1 speakers are able to make use of L2 words parafoveally during reading. Particularly, we were interested in replicating Fernandez et al.’s surprising finding that bilingual readers showed an inhibition from a translation mask (Pfeil vs Arrow) but a facilitation from a translation mask that shared only partial orthographic information (Arrow vs. Arrzm). In addition, we developed several other mask types and comparisons to identify the theoretical reasons why this pattern may occur.

If we replicated what was found in the Fernandez et al. (2021) paper, there would

be an N+1 inhibition effect (in terms of greater reading times following the English translation condition) between the German mask and English translation (Pfeil vs. Arrow), presumably arising because of costs incurred from switching from German to English and back to German. And, there would be an N+1 facilitation effect (in terms of shorter reading times following the [semantically unrelated] English orthography condition) between the English translation and English orthography mask (Arrow vs. Arrzm), presumably because both German and English lexicons remain active, and lexical and semantic candidates are narrowed down leading to quicker access to Pfeil upon foveal viewing. When looking at the data in the current study, we replicated the N+1 inhibition when comparing Pfeil vs. Arrow in FFD and GD. This suggests that readers fixate on the target word longer after parafoveally viewing the English translation. However, we were not able to replicate the N+1 facilitation when comparing Arrow vs. Arrzm in any measure. Together this indicates that bilingual readers show an immediate effect when a translation mask is viewed parafoveally but they do not show a benefit from an orthographic mask. Thus, we found no evidence that bilingual readers use partial translation information to narrow down lexical candidates parafoveally.

To investigate why this pattern emerged, we first compare the English orthography mask and the English pseudo-word mask (Arrzm vs. Clain). We hypothesized two possibilities. First, a finding that there are no differences between the two conditions would be consistent with the idea that the reader is treating both the English orthography mask and the English pseudo-word mask as non-word masks that provide no helpful information. Second, a finding that there is an N+1 inhibition effect in terms of greater reading times for Clain relative to Arrzm would indicate that the

shared orthography of *Arrzm* and the English translation leads to a semantic facilitation when ultimately reading *Pfeil*. In the current study, we found no differences when comparing *Arrzm* vs. *Clain*. This suggests that bilingual readers are treating both types of masks as non-words, despite *Arrzm* sharing orthographic information with the English translation word.

Second, we compare the English translation and the English word (*Arrow* vs. *Array*). We considered several possible outcomes. First, an N+1 inhibition effect (in terms of longer reading times from the English translation relative to the English word mask) would indicate that any N+1 inhibition seen when comparing the German word and its English translation (*Pfeil* vs. *Arrow*) must be due to the semantic overlap between *Arrow* and *Pfeil* (which led to a switch cost), and not merely due to the fact that the identical word was not in the parafovea (that led to increased reading times). Second, no differences between the conditions would suggest that the German vs. English translation masks may be driven merely by the fact that there is any word in English in the parafovea. Finally, an N+1 facilitation (in terms of shorter reading times from the English translation relative to the English word) would indicate that a language switch is occurring, and reading times would be further compounded by the fact that the English word mask shares no semantic information with the German word. In the current experiment, we found no differences between *Arrow* vs. *Array*. This lack of difference suggests that the N+1 inhibition we found between *Pfeil* and *Arrow* stemmed merely from the fact that there was an English word in the parafovea. This would also mean that the N+1 inhibition found by Fernandez et al. was not due to the semantic overlap between the target word and the translation, but was also due to the fact that there was an English

word in the parafovea.

Lastly, when comparing the English translation and the two English non-words (Arrow vs. Arrzm + Clain), three possible outcomes seemed likely. First, an N+1 facilitation (in terms of shorter reading times for the English translation relative the two non-word conditions) would indicate semantic facilitation from the English translation word relative to the other uninformative masks. Alternatively, no difference between the conditions would suggest that all three mask types were treated as uninformative masks. Finally, an N+1 inhibition (with greater reading times for the English translation relative to the two non-words) would indicate that readers incurred a switching cost from the translation but not from the non-words. Here, we found no differences between Arrow vs. Arrzm + Clain. This suggests that bilingual readers are treating all three masks as uninformative, and translation words do not incur a measurable switch cost relative to non-words. We therefore found no support for Fernandez et al.'s (2021) hypothesis that non-words evoke less interference than real words. Instead, our results support the numerous studies that have found that the less word-like the parafoveal mask, the greater the N+1 inhibition (e.g., Fernandez, Scheepers, & Allen, 2020; Hutzler, Schuster, Marx, & Hawelka, 2019; Marx, Hawelka, Schuster & Hutzler, 2015; Vasilev & Angele, 2017).

We also hypothesized that individuals with higher QLR and proficiency scores would be more efficient at extracting parafoveal information, leading to shorter first fixation and gaze durations as well as to increased skipping rate. We did not find any main effects of QLR or proficiency, nor did we find interactions with the mask comparisons for any of our dependent variables. This is somewhat surprising given the importance of QLR and proficiency during reading and parafoveal processing (Veldre &

Andrews, 2014; Wang et al., 2014; Whitford & Titone, 2015). Given that we found very little evidence of semantic N+1 processing, we speculate that the role of individual differences was just not apparent. Perhaps individual differences are more apparent when looking at orthographic N+1 processing (with the mask and the target sharing orthography), given that orthographic parafoveal processing is more readily found than semantic (e.g., Schotter et al., 2012). It is also possible that L2 QLR is important for parafoveal processing, but the L2 proficiency measure we used here is not a good proxy for L2 QLR. Though intuitively QLR and proficiency seem like they should go hand in hand, it may be the case that in opaque orthography languages like English, one could have high proficiency with lower QLR. So, while a bilingual reader may have high L2 comprehension, they may have poor orthographic representation, which may limit parafoveal processing. We believe that teasing apart these individual differences in bilingual readers is a compelling line for future research.

However, we did find an interaction between QLR and proficiency for FFD and GD. To interpret these interactions, we split the participants into high and low scoring QLR and visualized the dependent variables across proficiency score. We found that individuals with lower German QLR scores showed decreased FFD and GD as English proficiency increased, while individuals with high German QLR score showed increased FFD and GD as English proficiency increased. This suggests that there might be a tradeoff between L1 QLR and L2 proficiency such that if you have weaker L1 representation, an increase L2 proficiency may facilitate the decoding of parafoveal information leading to shorter reading times on the target word (at least in the case where parafoveal masks are more L2-like than L1-like). If you have stronger L1 representation,

then an increase in L2 proficiency may inhibit the decoding parafoveal of information, leading to longer reading times on the target word (at least in the case where parafoveal masks are more L2-like than L1-like). Again, we believe teasing apart these differences would be a compelling line for future research.

Overall, our data do not show evidence for either of our outlined hypotheses. While we were able to replicate the N+1 inhibition between the identical target and the translation mask, we were not able to replicate the N+1 facilitation between the translation mask and the orthographic mask. In terms of the replicated N+1 inhibition, our follow-up comparisons revealed that this difference most likely stems from the fact that the parafoveal mask does not match the target word, rather than from semantic overlap between the mask and target. This suggests that bilingual readers are treating all L2 mask types similarly, and do not process any L2 semantic, L2 orthographic, or L2 pseudo-word information parafoveally. This data therefore supports previous research that has found no cross-language semantic N+1 benefit with bilingual readers (Altarriba et al., 2001; Hoversten & Traxler, 2020; and Wang et al., 2014). Similarly, it fails to support previous research that has found L2 N+1 semantic benefit from bilingual readers (Antúnez et al., 2021; Fernandez et al., 2021; Wang et al., 2016).

In terms of the Fernandez et al. (2021) study, their data may indeed have been spurious. There could be several reasons for this discrepancy. It is possible that their pattern of results was due to a lack of power and may have led to false positives (see for example Brysbaert, 2019) that could not be replicated in the current study, highlighting the importance of replication in psycholinguistic research. In their study, Fernandez et al. compared 51 L1 speakers and 51 L2 speakers in a 2x6 design with 4 items per condition.

In the current study, we overcame some of the power issues by testing 100 participants in a 1x5 design (with 18 items per condition). Another possibility is that the spelling-to-sound correspondence plays a role in the ability of the participants to make use of the parafoveal information. German has a relatively clear spelling to sound correspondence, unlike English, which is relatively opaque. In the Fernandez et al. study, participants may have been able to access the semantic information of the orthographic German parafoveal mask more efficiently than when the mask was presented parafoveally in English in the current study.

Fernandez et al. (2021) attempted to explain their findings within BIA+ model of word recognition, but did not provide definitive theoretical implications of parafoveal processing. In the current study we aimed to provide further insight into these potential implications. However, given that we were not able to replicate the main Fernandez et al. findings, we do not expand on this further. The findings in the current study may still easily fit within the BIA+ if, for example, L2 N+1 semantic information does not activate the corresponding language node. Within the current study we cannot definitively say why this node is not activated, but it seems reasonable that it may be due to the opaque orthography of the English masks (making semantic access more difficult or time-consuming), a general slowing of foveal lexical access by L2 speakers (thus limiting the amount of cognitive resources available for parafoveal processing), or a combination of both.

While the main aim of the current study was not to test models of eye-movement behavior, we believe that a brief discussion is warranted in light of the findings. There is very limited research exploring L2 patterns of eye movements within the framework of

eye-movement models based on L1 reading. One such study that has discussed this issue is Cop et al. (2015); they argued that their findings fit easily within the E-Z Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle & Sheridan, 2015) model of eye-movement behavior. In the current study, in contrast to Fernandez et al., but in line with Cop et al., we did not find increased skipping rates or evidence of facilitated N+1 processing. Rather, when L1 German/L2 English speakers read a sentence in their L1, they showed little to no L2 N+1 processing. This would fit within the E-Z reader framework, if we assume that even though there are theoretically enough resources during the *familiarity check* to shift attention to the N+1 word (given the L1 sentence frame), N+1 processing will take longer (given that the N+1 word is in an L2) and require more resources (given the potential language switch and the slowed L2 access). As a result, bilingual readers will be less likely to process L2 N+1 semantic information.

Conclusions

While research investigating bilingual language processing has become more prominent, there are still very few studies that systematically investigate classic oculomotor behavior patterns in L2 speakers, at least not to the extent that is seen in L1 speakers (e.g., skipping behavior and parafoveal processing). In this registered report we attempted to replicate Fernandez et al. (2021) with greater power and new theoretically motivated types of masks. Interestingly, we were able to replicate the semantic N+1 inhibition between the identical mask (Pfeil) compared to the English translation (Arrow), but we were not able to replicate the orthographic N+1 facilitation between the English translation (Arrow) and the English orthography mask (Arrzm). We further tested the potential reasons behind this semantic N+1 inhibition and found no evidence

that the inhibition stemmed from semantic processing (or incurred a switch cost). Rather, the inhibition stemmed from the mere fact that the parafoveal word did not match the target word. Unlike the findings of Fernandez et al., the current findings provide little evidence that bilingual readers are able to make use of L2 semantic and orthographic N+1 information. Additionally, the current findings seem to fit straightforwardly within the BIA+ model of word recognition and the E-Z reader model. Overall, we were not able to replicate the findings of Fernandez et al., thus finding little evidence of L2 semantic N+1 processing, and highlighting the growing need for high powered registered replication studies within the field of bilingualism.

Competing interests: The author(s) declare none.

References

Altarriba, J., Kambe, G., Pollatsek, A., & Rayner, K. (2001). Semantic codes are not used in integrating information across eye fixations in reading: Evidence from fluent Spanish-English bilinguals. *Perception & Psychophysics*, *63*, 875-890. doi:

10.3758/BF03194444.

Antúnez, M., Mancini, S., Hernández-Cabrera, J.A., Hoversten, L.J., Barber, H., & Carreiras, M. (2021). Cross-linguistic semantic preview benefit in Basque-Spanish bilingual readers: Evidence from fixation-related potentials. *Brain and Language*, *214*, 104905. doi:10.1016/j.bandl.2020.104905.

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

Bates, D. M., Maechler, M., & Bolker, B. (2019). *lme4: Linear mixed-effects models using S4 classes*, R package version 1.1-21.

Bobb, S. C., & Wodniecka, Z. (2013), Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning, *Journal of Cognitive Psychology*, *25*, 568-585. doi:10.1080/20445911.2013.792822.

Brysbaert, M., 2019. How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*(1), p.16. doi: 10.5334/joc.72

- Declerck, M., & Philipp, A. M. (2015). A review of control processes and their locus in language switching. *Psychonomic Bulletin & Review*, *22*, 1630–1645. doi: 10.3758/s13423-015-0836-1.
- Dijkstra, T., & Van Heuven, W. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197. doi: 10.1017/S1366728902003012.
- Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R.J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. *Experimental Psychology: Learning, Memory, and Cognition*, *33*, 663–679. doi: 10.1037/0278-7393.33.4.663.
- Elston-Güttler, K. E., Gunter, T. C., & Kotz, S. A. (2005). Zooming into L2: global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research*, *25*, 57–70. doi: 10.1016/j.cogbrainres.2005.04.007.
- Fernandez, L.B., Allen, S., & Scheepers, C. (in press). Crosslinguistic differences in parafoveal semantic and orthographic processing. *Attention, Perception, & Psychophysics*.
- Fernandez, L. B., Scheepers, C., & Allen, S. E. (2020). The impact of uninformative parafoveal masks on L1 and late L2 speakers. *Journal of Eye Movement Research*, *13*. doi: 10.16910/jemr.13.6.3.
- Friel, B. M., & Kennison, S. M. (2001). Identifying German-English cognates, false cognates, and non-cognates: Methodological issues and descriptive norms. *Bilingualism: Language and Cognition*, *4*, 249-274. doi:10.1017/S1366728901000438.

- Hohenstein, S., Laubrock, J., & Kliegl, R. (2010). Semantic preview benefit in eye movements during reading: A parafoveal fast-priming study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1150–1170. doi: 10.1037/a0020233.
- Hohenstein, S., & Kliegl, R. (2014). Semantic preview benefit during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 166-190. doi: 10.1037/a0033670.
- Hoversten, L.J., & Traxler, M.J. (2020). Zooming in on zooming out: Partial selectivity and dynamic turning of bilingual language control during reading. *Cognition*, *195*, 104118. doi:10.1016/j.cognition.2019.104118
- Hyönä, J., & Häikiö, T. (2005). Is emotional content obtained from parafoveal words during reading? An eye movement analysis. *Scandinavian Journal of Psychology*, *46*, 475–483. doi: 10.1111/j.1467-9450.2005.00479.x.
- Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTenCorpus Family. In *7th International Corpus Linguistics Conference CL* (pp. 125-127).
- Jouravlev, O., & Jared, D. (2018). Cross-script orthographic and phonological preview benefits. *Quarterly Journal of Experimental Psychology*, *71*(1), 11-19. doi: 10.1080/17470218.2016.1226906.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633.
- Kuznetsova, A., Brockhoff, P. B., & Bojesen, C. (2019). *lmerTest: Tests in linear effects models*, R package version 3.1-0.

- Laubrock, J., & Hohenstein, S. (2012). Orthographic consistency and parafoveal preview benefit: A resource-sharing account of language differences in processing of phonological and semantic codes. *Behavioral and Brain Sciences*, *35*(5), 292-293. doi:10.1017/S0140525X12000209.
- Lauro, J., & Schwartz, A. I. (2017). Bilingual non-selective lexical access in sentence contexts: A meta-analytic review. *Journal of Memory and Language*, *92*, 217–233. doi: 10.1016/j.jml.2016.06.010.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325-343. doi: 10.3758/s13428-011-0146-0.
- Landauer, T.K., Foltz, P.W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284. doi: 10.1080/01638539809545028.
- Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381–390. doi: 10.1037/a0014875.
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. doi: 10.3389/fpsyg.2015.01171.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203–208 (1996). doi: /10.3758/BF03204766.
- Marian, V., & Shook, A. (2012). The cognitive benefits of being bilingual. *Cerebrum: the Dana forum on brain science*, *2012*, 1-11.

Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 787–796. doi: 10.1037/a0035583.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*, 65–81. doi: 10.1016/0010-0285(75)90005-5.

Rayner, K., Balota, D. A., & Pollatsek, A. (1986). Against parafoveal semantic preprocessing during eye fixations in reading. *Canadian Journal of Psychology*, *40*, 473–483. doi: 10.1037/h0080111.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology*, *12*, 206–226. doi: 10.1016/0010-0285(80)90009-2.

Rayner, K. & Schotter, E.R. (2014). Semantic preview benefit in reading English: The effect of initial letter capitalization. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1617-1628. doi: 10.1037/a0036763.

Rayner, K., Schotter, E.R., & Drieghe, D. (2014). Lack of semantic parafoveal preview benefit in reading revisited. *Psychonomic Bulletin & Review*, *21*, 1067-1072. doi: 10.3758/s13423-014-0582-9

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157. doi: 10.1037/e520602012-979.

- Reichle, E. D., & Sheridan, H. (2015). E-Z Reader: An overview of the model and two recent applications. In A. Pollatsek & R. Treiman (Eds.), *Oxford library of psychology. The Oxford handbook of reading* (p. 277–290). Oxford University Press.
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. doi: 10.1016/j.jml.2019.104038.
- Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition* *37*, 569–586. doi: 10.3758/MC.37.5.569
- Schotter, E.R. (2013). Synonyms provide semantic preview benefit in English. *Journal of Memory and Language*, *69*, 619–633. doi: 10.1016/j.jml.2013.09.002.
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*, 5–35. doi:10.3758/s13414-011-0219-2.
- Schotter, E.R., & Fennell, A.M. (2019). Readers can identify the meanings of words without looking at them: Evidence from regressive eye movements. *Psychonomic Bulletin & Review*, *26*, 1697–1704. doi: <https://doi.org/10.3758/s13423-019-01662-1>
- Schotter, E. R., & Jia, A. (2016). Semantic and plausibility preview benefit effects in English: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(12), 1839–1866. doi: <https://doi.org/10.1037/xlm0000281>
- Schotter, E. R., Lee, M., Reiderman, M., & Rayner, K. (2015). The effect of contextual constraint on parafoveal processing in reading. *Journal of Memory and Language*, *83*, 118–139. doi: <http://dx.doi.org/10.1016/j.jml.2015.04.005>

- Schwartz, A.I., & Kroll, J.F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, 55, 197-212. doi: 10.1016/j.jml.2006.03.004.
- Tsai, J., L., Kliegl, R., & Yan, M. (2012). Parafoveal semantic information extraction in traditional Chinese reading. *Acta Psychologica*, 141, 17–23. doi: 10.1016/j.actpsy.2012.06.004.
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, 64, 88–107. doi: 10.1016/j.jml.2010.08.006.
- Van Assche, E. Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology*, 3, 1-8. doi: 10.3389/fpsyg.2012.00174.
- Van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, 128, 431-451. doi: 10.1016/j.actpsy.2008.03.010.
- Vasilev, R. M., & Angele, B. (2017). Parafoveal preview effects from word N+1 and word N+2 during reading: A critical review and Bayesian meta-analysis. *Psychonomic Bulletin and Review*, 24, 666-689. doi: 10.3758/s13423-016-1147-x.
- Vaughan-Evans, A, Liversedge, S.P, Fitzsimmons, G. & Jones, M.W. (2020). Syntactic co-activation in natural reading. *Visual Cognition*, 28, 541-556. doi: 10.1080/13506285.2020.1841866
- Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, 67, 703-727. doi: 10.1080/17470218.2013.826258.

- Veldre, A., & Andrews, S. (2016a). Semantic preview benefit in English: Individual differences in the extraction and use of parafoveal semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 837-854. doi: 10.1080/17470218.2013.826258.
- Veldre, A., & Andrews, S. (2016b). Is semantic preview benefit due to relatedness or plausibility? *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 939-952. doi: 10.1037/xhp0000200
- von der Malsburg, T., & Angele, B. (2017). False Positives and Other Statistical Errors in Standard Analyses of Eye Movements in Reading. *Journal of Memory and Language*, *94*, 119-133. doi: 10.1016/j.jml.2016.10.003.
- Wakeford, L. & Murray, W. (2013). Effects of parafoveal plausibility during reading. In M. Knauff, N. Sebanz, M. Pauen, & I. Wachsmuth, I. (Eds.), 35th annual meeting of the Cognitive Science Society: *Cooperative minds: Social interaction and group dynamics* (p. 3723-3728). Austin, TX: Cognitive Science Society.
- Wang, A., Yeon, J., Zhou, W., Shu, H., & Yan, M. (2016). Cross-language parafoveal semantic processing: Evidence from Korean-Chinese bilinguals. *Psychonomic Bulletin & Review*, *23*, 285-290. doi: 10.3758/s13423-015-0876-6.
- Wang, A., Zhou, W., Shu, H., & Yan, M. (2014). Reading proficiency modulates parafoveal processing efficiency: Evidence from reading Chinese as a second language. *Acta Psychologica*, *152*, 29-33. doi: 10.1016/j.actpsy.2014.07.010.
- Whitford, V., & Titone, D. (2015). Second-language experience modulates eye movements during first- and second-language sentence reading: Evidence from a gaze-

contingent moving window paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1118-1129. doi: 10.1037/xlm0000093.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review*, *16*, 561–566. doi: 10.3758/PBR.16.3.561.

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2012). Lexical and sub-lexical semantic preview benefits in Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1069–1075. doi:10.1037/a0026935.

Yan, M., Wang, A., Song, H., & Kliegl, R. (2019). Parafoveal processing of phonology and semantics during the reading of Korean sentences. *Cognition*, *193*, 104009.

Yang, J., Wang, S., Tong, X., & Rayner, K. (2012). Semantic and plausibility effects on preview benefit during eye fixations in Chinese reading. *Reading and Writing*, *25*, 1031–1052. doi:10.1007/s11145-010-9281-8.

¹Only parafoveal N+1 preview effects are considered here given that it is the focus of the current study. Please see Schotter et al. (2012) for a more comprehensive review of parafoveal processing, for example parafoveal on foveal effects.

²For similar results using a different paradigm (parafoveal fast-priming), see Hohenstein, Laubrock, & Kliegl (2010)

³While they did not directly compare `arrow` and `pfexk`, inspection of their descriptive statistics and figures suggest that `arrow` and `pfexk` evoked surprisingly similar reading rates. The first fixation duration of `arrow` was 216 ms in the Identical condition (`arrow`), 220 ms in the German orthographic condition (`pfexk`), and 256ms in the German translation condition (`pfeil`) for L2 speakers.

⁴We would like to thank an anonymous reviewer for pointing out the potential confounds that led to the addition of the English pseudo-word and English orthography word conditions.

⁵While the pseudo-word masks were based on English words, given that both English and German are Germanic languages, the masks (particularly those words shorter in length) may be possible pseudo-words in both German and English. We do not believe this will impact eye-movement behavior given that neither L1 English speakers or L2 English speakers (with a German L1) seem to be sensitive to parafoveally viewed language-specific sub-lexical information (Fernandez et al., 2020).

⁶Note that two words were not in the HAL/LSA database (`airholes` and `caulk`) and therefore the missing values were replaced with mean of the remaining 88 values in the dataset.

⁷Maximal fitting model: $(DV \sim 1 + \text{condition} * \text{centered_LexDec} * \text{centered_proficiency} + (1 + \text{condition} | \text{Participant}) +$

(1 + condition| Item)

Table 1. *Fernandez et al. summary of hypotheses and findings.*

Mask comparisons	Fernandez et al. hypotheses	Fernandez et al. findings
arrow vs. arrzm	N+1 orthographic facilitation	No differences - evidence of N+1 orthographic facilitation
arrzm vs. xezne	N+1 inhibition - due to the mask being a non-word	No difference - perhaps due to shared letter shape (i.e., masks not different enough to be detected and lead to inhibition)
arrow vs. pfeil	N+1 semantic facilitation for the L2 group, and N+1 inhibition for the L2 group	N+1 inhibition for both groups - perhaps due to switching costs (English to German to English) for L2 group and due to the mask being a non-word for the L1 group
pfeil vs. pfexk	N+1 orthographic facilitation (via semantics)	N+1 semantic/orthographic facilitation - perhaps because both lexicons remained active
pfexk vs. ylctb	N+1 inhibition - due to the mask being a non-word	N+1 inhibition - due to the mask being a non-word
QLR - spelling	Higher QLR - more efficient parafoveal information extraction	Impact of QRL – reading durations (gaze duration) decreased as QRL increased
English proficiency	Higher proficiency - more efficient parafoveal information extraction	No impact

Table 2. Example stimulus. *The sentence frame is shown in the leftmost column where the critical word (here the underscored word Pfeil) was embedded; “|” indicates the position of the invisible boundary. Before crossing this boundary with an eye-movement, the critical word was masked with one of five different types of letter strings (columns 2-6).*

	Identical	English translati on	English ortho- graphy	English word	English pseudo- word
Der Mann fand de n ... in einem Baum.	Pfeil	Arrow	Arrzm	Array	Clain

Table 3. *Current study hypotheses.*

Current mask study comparisons	Hypotheses based on original Fernandez et al. theoretical hypotheses	Hypotheses based on Fernandez et al. findings	
Pfeil vs. Arrow	N+1 semantic facilitation	N+1 inhibition - perhaps due to switching costs (German to English to German)	
Arrow vs. Arrzm	N+1 orthographic facilitation (via semantics)	N+1 facilitation - perhaps because both lexicons remained active	
New comparisons			
	Hypothesis if facilitation is due to “word-likeness”	Hypothesis if facilitation is due to orthographic overlap	
Arrzm vs. Clain	No differences - due to both words being non-words that provide no helpful information	N+1 inhibition - due to the mask not sharing orthographic information with the translation	
	Hypothesis if facilitation is due to purely semantic information	Hypothesis if all masks are treated as non-words	Hypothesis if there is semantic access and a switch cost is incurred
Arrow vs. Arrzm + Clain	N+1 semantic facilitation (via semantics)	No difference – all words are uninformative for L1 German speakers	N+1 inhibition – switch cost from the English translation but not from the non-words
	Hypothesis if Pfeil vs. Arrow inhibition is due to semantic overlap	Hypothesis if Pfeil vs. Arrow inhibition is due to an English word being in the parafovea	Hypothesis if Pfeil vs. Arrow inhibition is due to semantic overlap and a switch cost
Arrow vs. Array	N+1 inhibition – suggesting that the Pfeil vs. Arrow inhibition is indeed from the costs of the language switch	No difference – the Pfeil vs. Arrow comparison is due to the fact that there is an English word in the parafovea	N+1 facilitation - due to the language switch, reader must switch to English and back to German but will have no semantic information to narrow down the lexical candidates in German

Table 4. *Participant information (standard deviation in parentheses).*

N	Male/ female	Mean Age	Mean OPT (in English)	Mean LexTALE (in German)
100	45/55	25.30 (4.40)	77.62 (11.21)	86.35 (8.11)

Table 5. Mean values and standard deviations per measure (FFD = First Fixation Duration, GD = Gaze Duration, and Skipping = Skipping Rate).

Condition	FFD (ms)		GD (ms)		Skipping (%)	
	M	SD	M	SD	M	SD
Identical	210.32	69.91	245.05	120.47	25.71	43.72
English translation	222.66	85.42	257.75	116.98	24.17	42.82
English orthography	225.57	88.43	264.88	124.42	23.20	42.23
English word	220.56	86.60	261.92	129.04	22.34	41.66
English pseudo-word	224.05	85.01	266.22	127.85	23.10	42.15



- (A) The pensioner mowed his s xxxx on Saturday.
- (B) The pensioner mowed his lawn on Saturday.

Figure 1. Illustration of the Gaze Contingent Boundary (GCB) paradigm (taken from Fernandez et al., 2021). The green ellipses represent hypothetical fixation points and the red dotted lines represent a pre-defined boundary that is invisible to the reader. As soon as the eye crosses the invisible boundary, the mask ‘xxxx’ in (A) is permanently replaced with the critical word ‘lawn’ in (B).

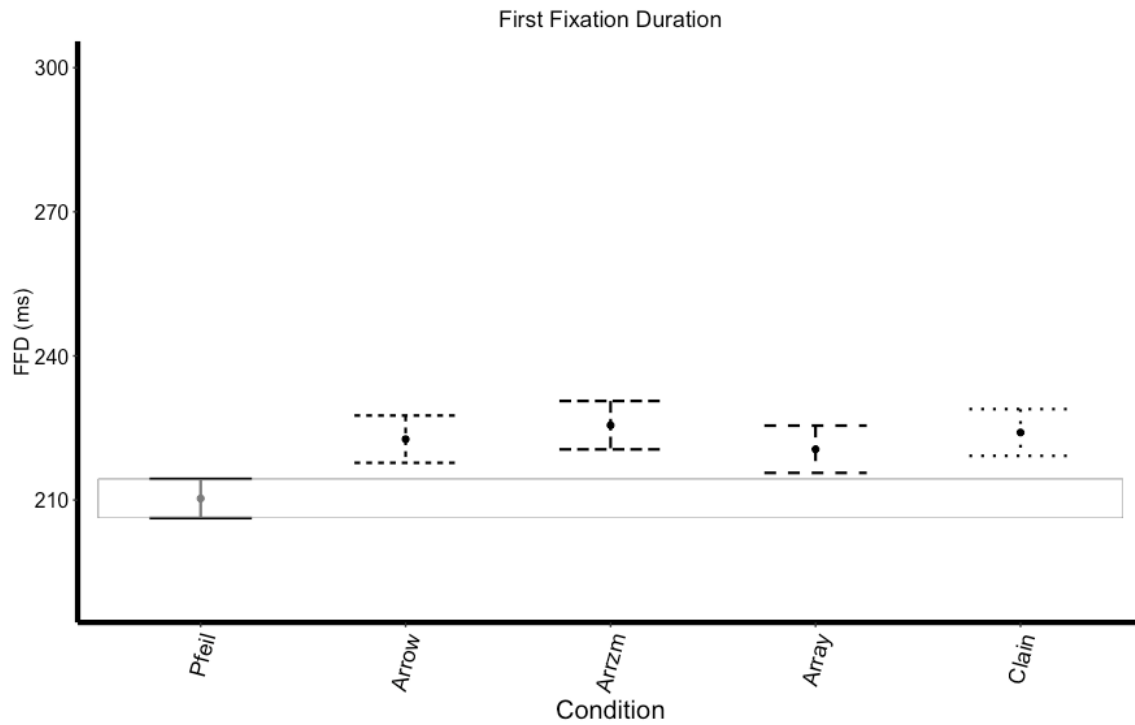


Figure 2. Mean FFD across all conditions (error bars reflect 97.5% confidence intervals)

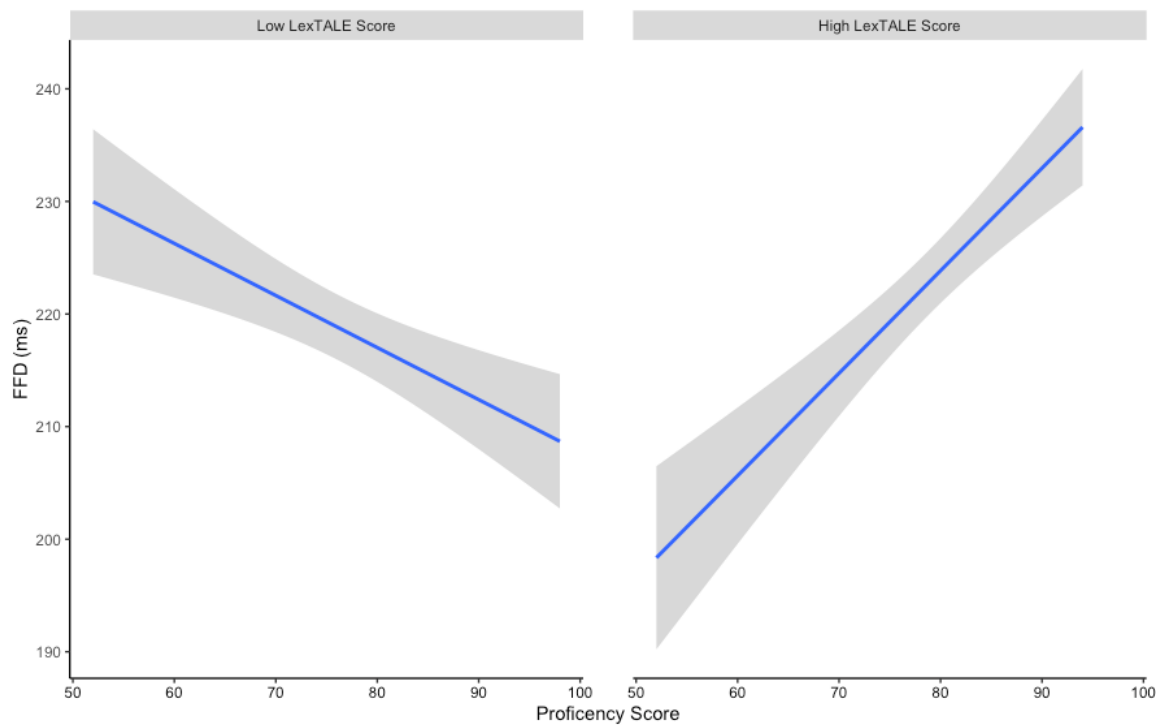


Figure 3. FFD and English proficiency split by LexTALE score (grey shading reflect 95% confidence intervals)

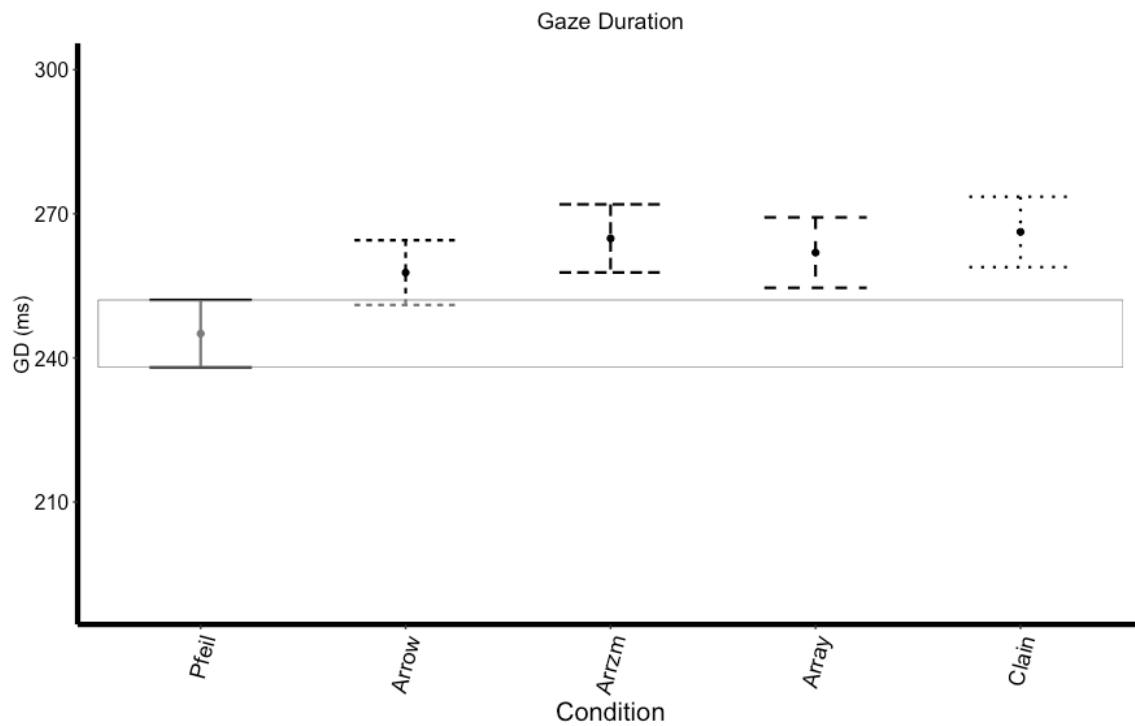


Figure 4. Mean GD across all conditions (error bars reflect 97.5% confidence intervals)

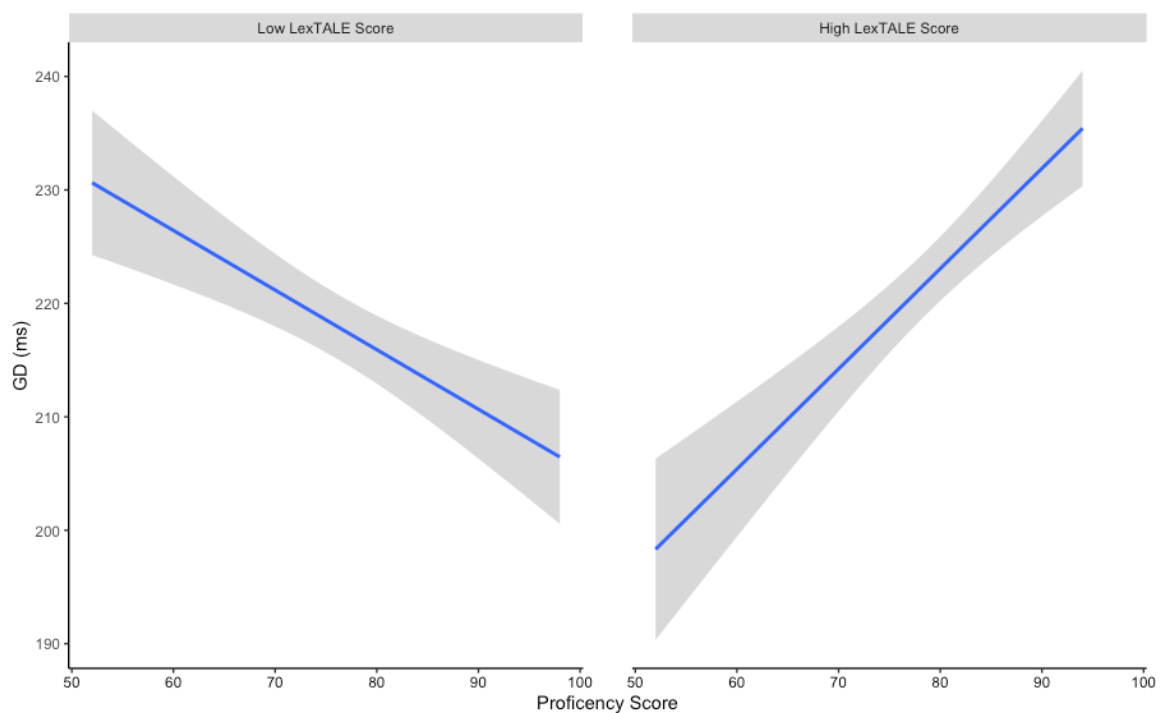


Figure 5. GD and English proficiency split by LexTALE score (grey shading reflect 95% confidence intervals)

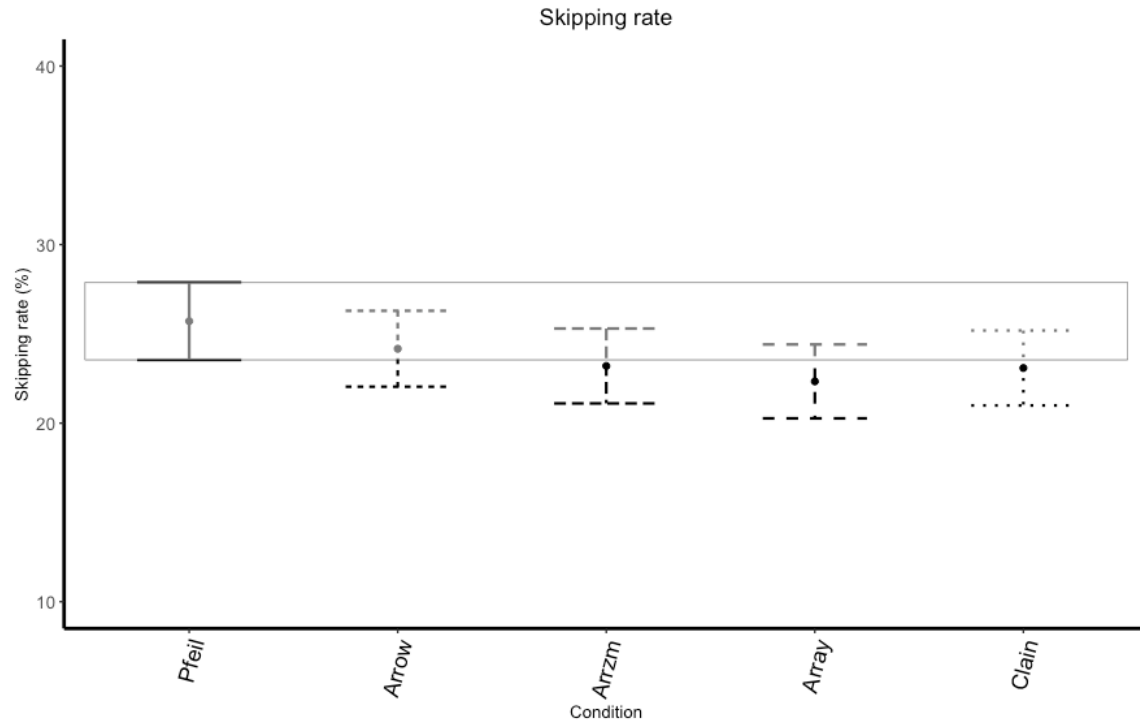


Figure 6. Mean skipping rate across all conditions (error bars reflect 97.5% confidence intervals)

Appendix A

Table A1. *Percent of data points removed per measure (FFD = First Fixation Duration, GD = Gaze Duration, and Skipping = Skipping Rate) and amount the of data points remaining per condition*

	Percentage removed	Identical	English translation	English orthography	English word	English pseudo word
FFD	25.00%	1249	1272	1299	1307	1286
GD	25.00%	1253	1272	1299	1306	1284
Skipping	NA	1715	1713	1711	1710	1702

Appendix B

Table B1. FFD (model 1)

	Estimate	Std. error	t-value	p-value
Intercept	222.55	2.45	90.72	<.001
Identical vs. English translation	13.77	2.92	4.71	<.001
English translation vs. English orthography	2.61	4.06	0.64	0.52
English orthography vs English pseudo-word	-1.93	4.23	-0.46	0.65
English translation vs. English orthography & pseudo-word	-1.42	3.21	-0.44	0.66
LexTALE	0.29	0.28	1.01	0.31
Proficiency	0.15	0.21	0.71	0.48
Identical vs. English translation*LexTALE	-0.21	0.37	-0.57	0.57
English translation vs. English orthography*LexTALE	0.24	0.43	0.56	0.57
English orthography vs English pseudo-word*LexTALE	-0.23	0.43	-0.54	0.59
English translation vs. English orthography & pseudo-word*LexTALE	0.01	0.36	0.02	0.98
Identical vs. English translation	0.05	0.26	0.18	0.86
English translation vs. English orthography*Proficiency	-0.35	0.31	-1.13	0.26
English orthography vs English pseudo-word*Proficiency	0.22	0.31	0.71	0.48
English translation vs. English orthography & pseudo-word*Proficiency	-0.03	0.26	-0.13	0.90
LexTALE*Proficiency	0.08	0.02	3.28	<.01
Identical vs. English translation*LexTALE*Proficiency	0.06	0.03	1.96	0.05
English translation vs. English orthography*LexTALE*Proficiency	-0.04	0.04	-1.06	0.29
English orthography vs English pseudo-word*LexTALE*Proficiency	-0.02	0.04	-0.65	0.52
English translation vs. English orthography & pseudo-word*LexTALE*Proficiency	-0.05	0.03	-1.77	0.08

Table B2. FFD (model 2)

	Estimate	Std. error	t-value	p-value
Intercept	222.71	2.51	88.56	<.001
English translation vs. English word	-2.44	3.98	-0.61	0.54
LexTALE	0.39	0.29	1.35	0.18
Proficiency	0.15	0.21	0.73	0.47
Semantic score	-2.23	12.75	-0.18	0.86
English translation vs. English word*LexTALE	-0.01	0.45	-0.03	0.98
English translation vs. English word*Proficiency	0.02	0.32	0.05	0.96
LexTALE*Proficiency	0.07	0.02	3.04	<.01
English translation vs. English word*LexTALE*Proficiency	-0.04	0.04	-1.11	0.27

Table B3. GD (model 1)

	Estimate	Std. error	t-value	p-value
Intercept	258.10	3.62	71.37	<.001
Identical vs. English translation	15.80	4.91	3.22	<.001
English translation vs. English orthography	7.16	6.00	1.20	0.23
English orthography vs English pseudo-word	-0.36	6.21	-0.06	0.95
English translation vs. English orthography & pseudo-word	5.05	5.19	0.97	0.33
LexTALE	0.39	0.38	1.03	0.30
Proficiency	-0.11	0.28	-0.39	0.70
Identical vs. English translation*LexTALE	-0.13	0.51	-0.25	0.80
English translation vs. English orthography*LexTALE	0.59	0.63	0.94	0.35
English orthography vs English pseudo-word*LexTALE	-0.63	0.66	-0.96	0.34
English translation vs. English orthography & pseudo-word*LexTALE	0.10	0.55	0.19	0.85
Identical vs. English translation	0.57	0.37	1.54	0.12
English translation vs. English orthography*Proficiency	-0.69	0.46	-1.50	0.13
English orthography vs English pseudo-word*Proficiency	0.12	0.49	0.24	0.81
English translation vs. English orthography & pseudo-word*Proficiency	-0.56	0.40	-1.39	0.17
LexTALE*Proficiency	0.08	0.03	2.63	<.01
Identical vs. English translation*LexTALE*Proficiency	0.04	0.04	0.88	0.38
English translation vs. English orthography*LexTALE*Proficiency	-0.06	0.05	-1.18	0.24
English orthography vs English pseudo-word*LexTALE*Proficiency	-0.01	0.06	-0.23	0.82
English translation vs. English orthography & pseudo-word*LexTALE*Proficiency	-0.06	0.05	-1.31	0.19

Table B4. GD (model 2)

	Estimate	Std. error	t-value	p-value
Intercept	259.30	3.76	68.93	<.001
English translation vs. English word	4.54	5.12	0.89	0.38
LexTALE	0.55	0.39	1.39	0.16
Proficiency	-0.08	0.28	-0.28	0.78
Semantic score	3.74	23.89	0.16	0.88
English translation vs. English word*LexTALE	0.46	0.51	0.91	0.36
English translation vs. English word*Proficiency	-0.54	0.38	-1.43	0.15
LexTALE*Proficiency	0.08	0.03	2.34	<.025
English translation vs. English word*LexTALE*Proficiency	-0.04	0.04	-0.84	0.40

Table B5. Skipping (model 1)

	Estimate	Std. error	z-value	p-value
Intercept	-1.38	0.11	-12.52	<.001
Identical vs. English translation	-0.14	0.1	-1.46	0.15
English translation vs. English orthography	0.00	0.1	0.02	0.98
English orthography vs English pseudo-word	-0.08	0.1	-0.76	0.45
English translation vs. English orthography & pseudo-word	-0.08	0.09	-0.93	0.35
LexTALE	0.00	0.01	-0.35	0.72
Proficiency	0.01	0.01	1.04	0.30
Identical vs. English translation*LexTALE	0.01	0.01	1.21	0.23
English translation vs. English orthography*LexTALE	-0.01	0.01	-1.03	0.30
English orthography vs English pseudo-word*LexTALE	0.00	0.01	-0.08	0.93
English translation vs. English orthography & pseudo-word*LexTALE	-0.01	0.01	-1.44	0.15
Identical vs. English translation	0.01	0.01	0.82	0.41
English translation vs. English orthography*Proficiency	0.00	0.01	0.29	0.78
English orthography vs English pseudo-word*Proficiency	0.00	0.01	-0.42	0.67
English translation vs. English orthography & pseudo-word*Proficiency	0.00	0.01	-0.32	0.75
LexTALE*Proficiency	0.00	0.00	0.23	0.82
Identical vs. English translation*LexTALE*Proficiency	0.00	0.00	1.68	0.09
English translation vs. English orthography*LexTALE*Proficiency	0.00	0.00	-0.34	0.74
English orthography vs English pseudo-word*LexTALE*Proficiency	0.00	0.00	-0.22	0.83
English translation vs. English orthography & pseudo-word*LexTALE*Proficiency	0.00	0.00	-0.32	0.75

Table B6. Skipping (model 2)

	Estimate	Std. error	z-value	p-value
Intercept	-1.38	0.11	-12.63	<0.001
English translation vs. English word	-0.09	0.09	-0.99	0.32
LexTALE	-0.01	0.01	-0.73	0.47
Proficiency	0.01	0.01	1.21	0.22
Semantic score	0.93	0.50	1.88	0.06
English translation vs. English word*LexTALE	-0.02	0.01	-1.48	0.14
English translation vs. English word*Proficiency	0.00	0.01	-0.42	0.68
LexTALE*Proficiency	0.00	0.00	0.06	0.96
English translation vs. English word*LexTALE*Proficiency	0.00	0.00	-0.04	0.97