Applied Psycholinguistics (2025), **46**, e34, 1–28
doi:10.1017/S0142716425100258

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Investigating the effects of prolonged exposure to textual enhancement on attention and learning: a pre-posttest measures eye-tracking study

Anastasia Pattemore, Vincent Fan, Laura Fiche and Hanneke Loerts

University of Groningen, Groningen, Netherlands
Corresponding author: Anastasia Pattemore; Email: a.pattemore@rug.nl

(Received 19 November 2024; revised 12 June 2025; accepted 10 September 2025)

Abstract

Research on the effects of textually enhanced (TE) subtitles on vocabulary acquisition through audiovisual input has yielded mixed results, primarily focusing on short viewing interventions. This study investigates the impact of TE on vocabulary meaning recall among 22 international students with limited knowledge of L3 Dutch. Participants watched an entire season of a comedy TV series in L2 English, accompanied by L3 Dutch subtitles as it would be broadcast on television. Using a within-subjects design, we assessed learning outcomes for 16 enhanced target words, 16 unenhanced target words, and 16 filler words absent from the subtitles. Two eye-tracking sessions were employed to measure participants' attention to both enhanced and unenhanced target words during the first and last episodes, addressing the limitations of previous studies that only included a single eye-tracking session and could not capture shifts in processing at pre- and posttest. The findings reveal that TE significantly increases fixations on enhanced words compared to unenhanced ones, with this difference remaining significant over the duration of the intervention, resulting in greater learning gains. Overall, the results highlight the potential of TE to facilitate vocabulary acquisition through subtitled audiovisual input.

Keywords: Additional languages; beginner learners; plurilingual audiovisual input; subtitles; vocabulary learning

Introduction

Subtitled audiovisual input, which combines images, sound, and on-screen text, is increasingly used for foreign language learning. To help learners notice and effectively acquire the target language, textually enhanced (TE) subtitles have been proposed as a powerful tool, as they can increase reading time fixation durations (as measured by eye-tracking, e.g., Puimège et al., 2023). However, studies reporting

positive effects of TE subtitles on learning generally use short video treatments and/or one-off data collections. In contrast, studies that incorporated longitudinal viewing treatments (Pattemore & Muñoz, 2022) have found no evidence of TE subtitles benefits, presumably due to a loss of attention toward TE subtitles over time, as suggested by Indrarathne et al. (2018). To date, research examining subtitles processing via eye-tracking over extended periods remains scarce.

So far, cross-sectional studies have been predominant in investigations of audiovisual input processing using eye-tracking methodology, although research on viewers' self-reported data suggests a significant change in viewing habits over time (Pattemore et al., 2024a). Thus, in this study, we assess changes in viewing behavior with eye-tracking sessions at pre- and posttest, following the methodology of Godfroid et al. (2018). The exploration of longitudinal exposure to textually enhanced words in subtitle lines is warranted not only by previous research but also by TechED tools, such as *Language Reactor* (Wilkinson & Apic, 2018), which allows users to highlight a word throughout an entire YouTube or Netflix episode—or even across a whole season of a Netflix TV series. Such technological affordances warrant further investigation into the effectiveness of such innovations.

This study addresses this gap by implementing a prolonged viewing intervention where participants watch an entire season of a TV show with TE subtitles. The participants' eye movements were recorded during the first and last episodes to measure potential shifts in attention to the TE target words. Over time, we also aimed to observe whether enhanced words consistently receive attention or whether they are increasingly skipped over time.

Literature review

Learning from audiovisual input

Audiovisual input such as films and TV series in a foreign language offers ample opportunities for foreign/second language (L2) development. For instance, it is widely acknowledged that exposure to audiovisual materials such as films and TV series leads to vocabulary gains (Montero Perez, 2022). These gains increase further when viewers are exposed to videos with captions (i.e., on-screen text in the same language as the soundtrack; Vanderplank, 2016).

The present study focuses on subtitles, on-screen text translation of the soundtrack, mostly to viewers' first language (L1). Previous research indicates that subtitles in the viewer's L1 support vocabulary learning and comprehension (e.g., Peters et al., 2016). However, little is known about the effects of watching with subtitles that are not in the viewer's L1, a common scenario in subtitling countries like the Netherlands, where this study is based. In these countries, non-native English and/or Dutch speakers are often exposed to L2 English films and TV series subtitled in L3+ Dutch on broadcast TV, in cinemas, and on streaming platforms, as dubbing is typically reserved for children's programs. Although this is a naturalistic viewing setting for many residents of subtitling countries, studies exploring this plurilingual subtitled input are limited to just three (Pattemore, 2023; Pattemore et al., 2024a, 2024b; Urbanek & De Vogelaer, 2025). These studies found significant effects of viewing with one foreign language in the audio and another in the subtitles

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

on learning vocabulary and multiword units. This learning might be attributed to the multilingual scaffolding that occurs when individuals are exposed to multiple languages simultaneously, allowing them to establish meaningful links between them (Duarte, 2020). Although previous research (Pattemore, 2023) using the eye-tracking methodology shows that viewers exposed to plurilingual subtitled audiovisual input process subtitles, it is still plausible that the challenging task of following a video in both L2 and L3 could benefit from additional support, especially in educational settings.

Textual enhancement and subtitling

To promote language learning from audiovisual input, researchers and language educators have explored additional methods to increase the noticing of target words and structures, particularly through textual enhancement (e.g., bolding, highlighting, and underlining). The benefit of input enhancement is theorized to lie in its ability to raise the salience of the target structures, thereby increasing the likelihood of them being learned (e.g., Sharwood Smith, 1993). The underlying assumption is that learners might miss target items without enhancement because these items may not naturally grab their attention. Consequently, increased attention to enhanced items is expected to lead to better language uptake (Leow & Martin, 2017). Although textual enhancement has been suggested to be beneficial for learning vocabulary from static texts (e.g., Vu & Peters, 2022), there is no overall consensus regarding its effectiveness. Systematic reviews have found mixed results, particularly for grammar, which does not consistently benefit from textual enhancement (e.g., Leow & Martin, 2017). Similarly, findings on the effects of textually enhanced (TE) on-screen text show inconsistencies.

One of the first studies that looked into the effect of TE captions (Montero Perez et al., 2014) explored L2 French vocabulary learning from watching three short video clips twice (about 21 minutes in total). While the TE group outperformed the no captions group, so did the unenhanced captions group, and there was no difference between the two captioning conditions. The authors suggested that unenhanced captions provided enough support to notice the target vocabulary and trigger the learning process. A more recent study on the learning of multiword units (Majuddin et al., 2021) obtained similar results. The experiment included a comparison between no captions, captions, and underlined TE captions while watching a 20-minute episode of a TV series where the target L2 English multiword units appeared once. The TE and unenhanced captions groups outperformed the no captions group in a form recall immediate posttest, but there was no difference between the two captioning groups. Interestingly, the comparison between the immediate and delayed posttest scores showed that the TE captions group performed better at the immediate posttest, indicating a short-term advantage of textual enhancement compared to unenhanced captions (cf. Finger-Bou & Muñoz, 2023; Pattemore & Muñoz, 2022). The authors also checked if there was a trade-off effect between paying attention to TE and fully attending to the content of the video. The results of the comprehension test suggested that the unenhanced captions group performed better than both the uncaptioned and TE captions groups, and there was no difference in content comprehension between the latter two groups.

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

This is surprising since captions were found to support comprehension (Montero Perez et al., 2013), but it seems that this did not apply to enhanced captions in Majuddin et al. (2021), suggesting that watching with textually enhanced captions might come at a cost.

Likewise, Finger-Bou and Muñoz (2023) compared TE (yellow and bold) and unenhanced captions, examining L2 English vocabulary learning from a 25-minute documentary. Participants were tested on meaning recall, meaning recognition, and form recall at the pretest, immediate posttest, and two-week delayed posttest. Both groups significantly improved their target vocabulary knowledge, with no significant difference between the TE and unenhanced conditions except for form recognition. However, this advantage was transient, as it did not hold at the delayed posttest. The results indicated that TE captions do not necessarily lead to better learning outcomes and that the effect of enhancement seems to be immediate rather than contributing to deeper learning (cf. Majuddin et al., 2021; Pattemore & Muñoz, 2022). Interestingly, half of the participants in the TE condition reported being distracted by the TE, focusing on the enhanced words and isolating themselves from the rest of the captions and the storyline of the video.

To shed light on L2 learners' focus of attention, reading patterns, and the division and shifting of attention in a complex task like simultaneously processing image, sound, and text, without relying on self-reports, researchers have turned to eye-tracking methodology (e.g., Bisson et al., 2014). This technique allows for studying learners' attention allocation on the screen in an online and unobtrusive way (Godfroid, 2020), providing insight into how L2 learners process audiovisual input and how this contributes to L2 gains.

Textually enhanced subtitles and eye-tracking

While a number of studies have applied the eye-tracking methodology to study processing (e.g., Kruger et al., 2022) and/or learning (e.g., Montero Perez, 2019) from videos and on-screen text, research on the effects of TE captions on learning and the attention allocation to textually enhanced target items remains limited. One such eye-tracking study (Lee & Révész, 2020) tested the effectiveness of textual enhancement (yellow font) over unenhanced captions on learning L2 grammatical constructions (present perfect and past simple) from 24 short video clips (20 to 50 seconds each). The eye-tracking data showed that TE constructions were more attention-grabbing than unenhanced constructions. While there was no significant learning for past simple grammatical constructions due to the participants' advanced proficiency level in L2 English, the results for the present perfect showed significant learning gains, particularly for the enhanced captions group. The authors suggested that the superiority of the enhanced over the unenhanced condition could be explained by the increased attention to the TE constructions, as enhancement increased their salience.

The eye-tracking study by Galimberti et al. (2023) tested the effect of textual enhancement (yellow font) on pronunciation training after watching four clips from a TV series (about 5 minutes total). The study showed that the textually enhanced captions group outperformed the unenhanced captions group in terms of pronunciation gains. The eye-tracking data also revealed that in the enhanced

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

conditions, once a sentence with a TE target word appeared on the screen, there was a significant delay between the first fixation time on the enhanced word and the time when the viewers actually heard the word. This suggests that textual enhancement grabs attention immediately after it is presented in the caption/subtitle line, regardless of when it will be produced by the narrators, taking at least some time away from other text or leading to audio-text unsynchronized processing of audiovisual input. This supports the findings in Finger-Bou and Muñoz (2023) that TE targets could be processed in isolation from the rest of the channels involved in audiovisual input (i.e., audio, video, on-screen text). Another eye-tracking study on pronunciation training and textual enhancement (TE) investigated whether different TE colors would lead to various learning outcomes when applied to two distinct pronunciation features, /ʌ/ and /æ/ (Mora & Fouz-González, 2024). The study involved four groups: a control group, an unenhanced captions group, a group with yellow-enhanced captions, and a group with captions contrastively enhanced in purple for /ʌ/ sound (e.g., cup) and yellow for /æ/ sound (e.g., cap). All groups that viewed a 30-minute TV series episode with either enhanced or unenhanced captions improved their perception of the target sounds significantly more than the control group. However, no clear advantage was observed among the different types of enhancement. Additionally, although textually enhanced words received longer fixation durations, the study found no correlation between the amount of attention participants paid to the targets and their learning gains. The authors suggested that the significant improvements in the unenhanced group could be due to participants being asked to focus on the target sounds. Without being distracted by the textual enhancement, they might have been able to focus not only on the highlighted targets but also on other words containing the same sounds.

Moving to vocabulary studies, Puimège et al. (2021) examined the learning of L2 English multiword units from underlined TE captions versus unenhanced captions. Using eye-tracking methodology, they measured the processing of TE-targeted multiword units and the effect of attention paid to these units on learning outcomes. A within-participants design allowed authors to gauge learning of both enhanced and unenhanced target words from the same audiovisual input. The eye-gaze movements during a 30-minute documentary showed that enhanced targets received more attention and were reread more than unenhanced multiword units. A form recall test showed that both enhanced and unenhanced targets had significantly higher gains than the comparison multiword units that did not appear in the video. A group comparison revealed that enhanced multiword units had significantly higher gains than the unenhanced, but this difference disappeared once total reading time was added to the model. This suggested that TE alone could not account for learning; factors such as engagement with the input (i.e., reading time), regardless of items being enhanced, led to significant learning. Similar to Majuddin et al.'s (2021) claim, due to the fast-paced nature of watching a video, participants might not have had enough time to fully process and reread the enhanced target multiword units.

A later multiword units eye-tracking study obtained conflicting results compared to Puimège et al. (2021). Choi (2023) focused on the effects of TE captions (in yellow) compared to unenhanced captions on learning L2 English collocations that appeared once in a short 7-minute TED talk video. The results showed that there

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

was no significant difference between the TE and unenhanced groups in terms of reading time for the target collocations, the sentences they appeared in, or the entire captions for the duration of the video. This means that textual enhancement did not lead to longer fixation durations, possibly due to the nature of the captioned audiovisual input where on-screen text is presented for a limited amount of time (cf. Majuddin et al., 2021). As for collocations learning, the TE group showed significantly higher results after watching the video than the unenhanced group. Interestingly, the author also implemented captions recall test to see if there was a trade-off effect of enhanced captions over unenhanced ones. The results revealed that attention to textual enhancement was not at the expense of unenhanced captions, as the group that watched the video with enhanced captions recalled a similar number of words in the unenhanced captions as the group that watched the video without enhancement. This, however, is not surprising since the reading time spent on the target collocations, whether in the enhanced or unenhanced group, was similar. A possible explanation for this could be the short video duration, where participants might have attended to all channels of input more or less equally. Differences might only appear after prolonged exposure. However, this finding contrasts with Majuddin et al. (2021), where textual enhancement caused a trade-off effect on video content comprehension. This discrepancy in results warrants further investigation into whether watching videos with textually enhanced targets comes at the expense of other linguistic features of the video.

Notably, all the studies discussed above were cross-sectional and provided short, single-viewing exposure to TE subtitles. Furthermore, the target items in those studies appeared in the videos only once. This means the studies examined potential learning from limited exposure to audiovisual input and did not employ a narrow viewing approach (Rodgers & Webb, 2011). This approach suggests that as viewers are continuously exposed to related L2 audiovisual input (e.g., a season of a TV series), they become more familiar with the content and characters, leading to increased comprehension and a higher likelihood of learning, especially as they are likely to encounter vocabulary from the same word families multiple times. Besides, daily exposure to television is a naturalistic experience, as evidenced by reports indicating that 79% of people living in Europe watch television on a daily or almost daily basis (European Commission, Directorate-General for Communication, 2023). Given this prevalence, it is imperative to shift from cross-sectional studies to examining the effects of audiovisual input in more ecologically valid settings.

Prolonged exposure to textual enhancement

Although several longitudinal viewing studies exist (e.g., Gesa & Miralpeix, 2023; Rodgers & Webb, 2011), only two studies have explored extensive exposure to textually enhanced audiovisual input (Pattemore & Muñoz, 2022; Finger-Bou & Muñoz, 2023). This represents a significant gap in research, particularly given that EdTech platforms like *Language Reactor* (Wilkinson & Apic, 2018) enhance words selected by viewers whenever they appear in the subtitles. It is therefore crucial to determine the effectiveness of textual enhancement during prolonged viewing.

The participants in Pattemore and Muñoz (2022) watched ten episodes of a TV series (227 minutes) over five weeks with captions, TE captions (in yellow), or

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

without captions. The authors measured learners' uptake of L2 English constructions, including multiword units (fully-filled constructions, e.g., do for a living), partially filled constructions (e.g., to be allowed to), and fully schematic constructions (e.g., passive voice). While the enhanced captions group outperformed both the unenhanced captions and no captions groups in the intermediate immediate posttests performed every week after watching two episodes, this advantage disappeared at the delayed posttest once the participants had watched all ten episodes. Unfortunately, the study did not employ eye-tracking to provide insights on the possible change in attention processing of TE constructions that might have affected the learning curve. Exposure to input over an extended period of time might influence reading behavior, processing, and attention allocation over time, as observed in naturalistic eye-tracking reading experiments where participants were exposed to several chapters or entire novels (Cop et al., 2015; Godfroid et al., 2018). The importance of tracking engagement with input over time is underscored by findings from Godfroid et al. (2018), where a decrease in reading time was observed with repeated exposure to unenhanced target non-English words (unenhanced). As participants became more familiar with these words, they were able to decode them faster. This aligns with the results of Indrarathne et al. (2018), where a reduction in attention to the enhanced target constructions (in bold) was observed as participants encountered them repeatedly (21 times) across three texts over three separate sessions. The authors suggested that attention decreased as the constructions became less novel, proposing that the increased attention to textually enhanced targets seen in previous cross-sectional studies could be due to textual enhancement attracting more attention during initial exposure rather than after repeated encounters. This, along with the observed decrease in performance for the enhanced group during the viewing intervention in Pattemore and Muñoz (2022)—not a few weeks after—suggests that learning from and processing textually enhanced targets is a dynamic process and highlights the need for research exploring naturalistic viewing with subtitles over time. Pattemore (2023), along with a recent study by Finger-Bou and Muñoz (2023), are among the first to extend naturalistic eye-tracking experiments (cf. Godfroid et al., 2018) to subtitling and audiovisual input research. In our previous study (Pattemore, 2023), we examined newcomers' learning of Dutch from 12 episodes of plurilingual audiovisual input, recording their eye gaze in the first and last episodes of the intervention. The eye-gaze behavior data indicated that participants allocated significantly more attention to subtitles and exhibited more reading-like behavior by the end of the intervention, possibly due to their growing familiarity with the plurilingual audiovisual input.

Similarly, in the study by Finger-Bou and Muñoz (2023), participants' viewing patterns were recorded with an eye tracker at the beginning and end of the intervention in their textually enhanced captions study. Primary school students (ages 10–11) watched 11 episodes of an L2 English animated TV series under three conditions: textually enhanced captions (L2 English), unenhanced captions, or no captions. The results showed that textually enhanced words, which were bolded and highlighted in yellow, initially had significantly longer fixation durations compared to the unenhanced captions group, but this difference disappeared by the end of the intervention. Additionally, vocabulary learning was significantly greater in the

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

textually enhanced captions condition than in both the unenhanced and no captions conditions.

Taking into account the above-mentioned gaps in research, the current study is the first of its kind to go beyond exploring captions textual enhancement and focus on plurilingual subtitled input textual enhancement. Moreover, no previous research implemented extensive exposure to TE target words with multiple eye-tracking sessions and within participants design with both enhanced and unenhanced target words in the audiovisual input. We address the following research questions:

1) To what extent does textual enhancement affect attention to enhanced and unenhanced target words in the subtitles at pre- and posttest?
2) To what extent are enhanced and unenhanced target items learnt from the intervention?
3) Is there an association between attention paid to the enhanced and unenhanced target items and learning outcomes?

Methodology

This study aimed to explore the effectiveness of TE over a prolonged period and implemented a pre-post eye-tracking design where the first and last episodes were viewed with an eye-tracker (EyeLink Portable Duo), while the episodes in between were viewed at home over three weeks. Using this setup with a within-participants design where participants were exposed to both enhanced and unenhanced target items, we aimed to determine the extent to which TE increases or decreases fixations on words, as well as the acquisition of these words, as indicated by pre- and posttesting. In addition, using a within-subjects design rather than a between-subjects design, where one group viewed enhanced and another unenhanced target items (e.g., Pattemore & Muñoz, 2022), allowed us to assess attention allocation to both enhanced and unenhanced items within the same viewing sessions (e.g., Puimège et al., 2021).

Participants

The initial pool of participants consisted of 30 university students (19–31 years old, $M = 23$, $SD = 3.95$); however, eight participants either did not watch all the episodes or their eye-tracking data were of low quality and therefore were excluded from the analysis, leaving 22 participants. Those 22 participants were international students at the University of Groningen in the Netherlands (20 undergraduate and 2 postgraduate levels) with a variety of L1 backgrounds (Italian, Spanish, Russia, Mandarin, Greek, Polish, Korean, Estonian, Bahasa, Farsi, Albanian, Hungarian, Romanian, Hindi, and Thai), but caution was made to avoid Germanic L1s (see procedure). Eighteen of the participants (82%) were comfortable with watching with subtitles in their daily lives, while four of them were not used to watching with subtitles. On average, the participants spent around one year in the Netherlands at the time of the data collection, and the majority of them did not take Dutch classes in the past (59%, 13/22). The intervention testing (Table 1, see proficiency measures

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Table 1. Participants' proficiency levels in English and Dutch

	Mean (SD)	Range	Max possible score
English LexTALE	81.87 (9.28)	56.25–97.50	100
Dutch PPVT	52.45 (14.65)	22–73	108

below) showed that the participants' Dutch proficiency was around A1, and English proficiency was advanced (C1–C2, based on Lemhöfer & Broersma, 2012).

Audiovisual input

The first season of the comedy fantasy TV series The Good Place (Schur, 2016) was used as the audiovisual input in this study. We chose this show because it has been used in multiple audiovisual input studies (e.g., Galimberti et al., 2023; Pattemore & Muñoz, 2022) and is suitable for a wide range of viewers, as it does not contain restricted content such as violence or nudity that could affect the viewers. The thirteen episodes (284 minutes in total) were shown to the participants in their original version with English audio and Dutch subtitles, similar to how they would be viewed on Dutch television or a Dutch Netflix account. Exposure to English L2 audio and L3+ Dutch subtitles provided participants with plurilingual subtitled audiovisual input, which has been shown to be beneficial for L3+ beginner vocabulary and multi-word units learning (Pattemore et al., 2024b).

The enhanced target items (see below) appeared throughout the episodes in yellow (see Figure 1) and were textually enhanced using TextEdit.

Target vocabulary

We chose 32 regularly occurring target items on the basis of the subtitle scripts of the first season of the TV series. A Welch's *t*-test confirmed that the enhanced ($n = 16$) and unenhanced ($n = 16$) target items appeared approximately equally often in the subtitles (Enhanced: $M = 42.4$, $SD = 19.4$; Unenhanced: $M = 53.2$, $SD = 22$) throughout the season (t (29.55) = -1.48, $p = .15$). To assess the target item's overall frequency in film and television subtitles, we used the log-transformed word frequency values (Lg10WF) from the SUBTLEX-NL database (Keuleers et al., 2010). These values were also comparable for enhanced ($M = 4.24$, $SD = 0.8$) and unenhanced ($M = 4.47$, $SD = 0.33$) items (t (20.06) = –1.04, $p = .31$).

A full overview of the targets can be found in Table 2. The list contains the stems of the targets and, in some cases, the occurrences of these items in the subtitles constitute inflections or conjugations, e.g., zielsverwant–zielsverwanten (soulmate–soulmates).

Pre-/posttest

Participants were asked to perform a standardized pre-/post vocabulary knowledge scale test (VKS, Wesche and Paribakht, 1996) containing 48 items in total: 16 target words that were enhanced in the subtitles throughout the season, 16 unenhanced target words, and 16 filler words that did not appear in the subtitles at all (see

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Figure 1. Example of textually enhanced subtitles and eye-tracking areas of interest.

Table 2). Participants were presented with the items in a written form, and they had to indicate their level of familiarity with the test items: 1) not having seen the word before; 2) recalling the word but not remembering the meaning; 3) recalling the meaning and providing it as a definition, translation, or description using synonyms/antonyms, but being unsure about it; 4) knowing the meaning and providing it; and 5) providing a use of the test item in context, along with the answer to option 4. Figure 2 shows how the test items were presented on the computer screen.

Following Wesche and Paribakht's (1996) guidelines, we evaluated participants' responses, scoring them from 1 to 5. If the participants wrote "1," they received 1 point. If the participants wrote "2," they received 2 points. If the participants wrote "3" and gave a correct translation or definition, they received 3 points. If they indicated "3" but the answer was incorrect (for example, writing "good" as a translation of 'gezellig'), they received 2 points. If the participants chose "4" and provided a correct translation or definition, they received 4 points. If the translation or definition was incorrect or nothing was filled in, they received 2 points. If the participants chose "5" and provided a correct translation or definition and used the word correctly in a phrase or sentence that revealed the definition of the word, they received 5 points (participants could see the written form of the word on the screen, and the spelling mistakes in the phrases and sentences were not penalized). If they used the word incorrectly in context but gave a correct definition or translation, they received 4 points. If the meaning was incorrect, they received 2 points.

A dichotomous score was created to classify items as known (1) or unknown (0): if participants received a score of 3 to 5, the word was considered known; a score of 1 or 2 indicated it was unknown.

Both the pretest and posttest were completed online using Qualtrics (Provo, UT). The same 48 items appeared in both the pretest and posttest and were automatically randomized. Participants were only provided with the textual presentation of the

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Table 2. Target vocabulary

	Dutch	English translation	Frequency of occurrence in *The Good Place*	Corpora frequency of occurrence (Lg10WF)
Enhanced (n = 16)	zielsverwant	soulmate	52	1.6
	trouwens	by the way	16	3.6
	vreselijk	awful	13	3.7
	geloven	believe	25	4.1
	gebeurt	happen	37	4.2
	vermoord	killed	27	4.2
	jongens	guys	19	4.2
	natuurlijk	of course/natural	34	4.4
	leuk	fun	61	4.4
	dood	dead/death	30	4.6
	mensen	people	72	4.6
	iemand	somebody	55	4.7
	misschien	maybe	53	4.8
	alleen	alone	58	4.8
	alles	everything	69	4.9
	waarom	why	57	5.0
Unenhanced (n = 16)	baas	boss	9	3.9
	buurt	neighbourhood	74	3.9
	slecht	evil/bad	97	4.1
	geweldig	great	62	4.2
	dingen	things	48	4.3
	blijven	stay	41	4.4
	elkaar	each other	32	4.4
	bedankt	thank you	31	4.5
	iedereen	everybody	68	4.5
	omdat	because	58	4.6
	gewoon	just/ordinary	76	4.6
	zeggen	say	61	4.7
	terug	back	40	4.8
	nooit	never	44	4.8
	toch	right/yet/still	37	4.9
	iets	something	73	5.0
Fillers (n = 16)	gezellig	cozy	NA	3.0
	lawaai	noise		2.9
	nauwelijks	barely		3.2
	rotzooi	mess		3.2
	spelletje	game		3.2
	langzaam	slowly		3.3
	opschieten	hurry up		3.4
	toekomst	future		3.7
	leger	army		3.7
	beneden	downstairs		3.9
	volgens	according to		4.1
	zonder	without		4.4
	klaar	ready		4.4
	zoals	like		4.5
	anders	different/otherwise		4.5
	ooit	someday		4.3

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

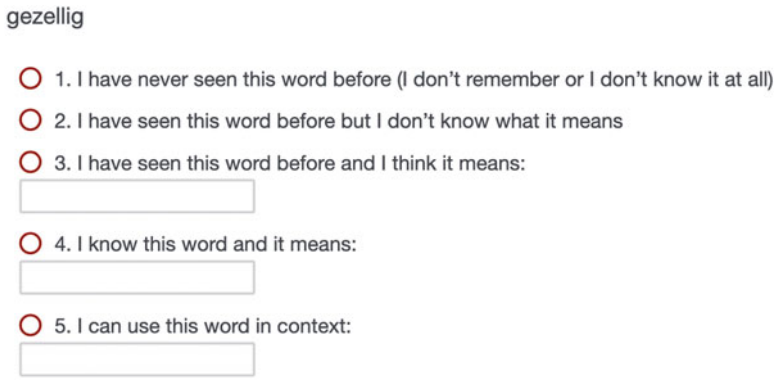

Figure 2. Example of a test item in the pre-/posttest.

test items, as they were exposed to L3 Dutch solely through textual subtitles and did not receive any auditory input (see Jelani & Boers, 2020).

Post-viewing questionnaire

To obtain participants' perceptions of the intervention and their learning/viewing experience, participants completed a post-viewing questionnaire after watching episodes 1 and 13 in the lab with the eye-tracker. They responded to a series of statements regarding their experience of watching the TV series in English with Dutch subtitles. The questions explored participants' views on the language differences between subtitles and audio, their attention levels, and the effectiveness of the TV series in promoting their vocabulary knowledge in L3 Dutch. We were also interested in whether watching in two foreign languages caused distraction and comprehension difficulties. Finally, participants also provided examples of what they learnt from the episodes. The results of the questionnaire were used for data triangulation, and the questionnaire can be found on the Github page associated with this study.

Proficiency measures

The participants completed vocabulary size tests as a proficiency proxy for English and Dutch (see Table 1).

To measure English proficiency, we used the English version of LexTALE (Lemhöfer & Broersma, 2012), a test of receptive vocabulary knowledge where participants are presented with 60 English-looking words and must decide whether they know each word. Forty of these words are existing English words, while 20 are non-words. If the test-taker claims to know a non-word, this response is penalized. The maximum score is 100, with scores below 59 indicating B1 proficiency level or lower, scores between 60 and 80 considered B2 level, and scores between 80 and 100 indicating C1–C2 level. Overall, LexTALE distinguishes between lower intermediate, upper intermediate, and advanced L2 speakers.

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Although LexTALE is also available in Dutch, the test is not reliable for proficiency levels below intermediate (Lemhöfer & Broersma, 2012), and we expected our participants to be at the beginning of their L3 Dutch learning process. Considering that, we measured Dutch receptive skills using the Dutch version of the Peabody Picture Vocabulary Test (PPVT-III-NL, Dunn et al., 2005), a test that was used by previous research on adult L2 Dutch learners (e.g., Deygers & Vanbuel, 2022). Following Deygers and Vanbuel (2022), we used the first nine sets of the test, each containing 12 items, resulting in a total of 108 items. For each item, participants were presented with an auditory word and four pictures to identify the target. The target items were audio recorded by a Dutch native speaker and were played by the participants at their own pace as a PowerPoint presentation. Participants could also see the textual representation of the target on the answer sheet to provide a multimodal test that included both auditory and textual information. This was done to avoid situations where a participant might know a written form of the word but not its auditory form, and vice versa. Each item received either a correct or incorrect score, with a maximum possible score of 108. Since there is no standardized grouping of PPVT scores into the Common European Framework of Reference (CEFR), we used Deygers and Vanbuel's (2022) results for secondary-level and higher educated learners of L2 Dutch literate in Latin script as a baseline. Thus, a score around 64 could be roughly attributed to a learner at an A1 level and a score of 71 to a learner at an A2 level.

Eye-tracking experiment

To record viewing of episode 1 and episode 13, we used EyeLink Portable Duo version 6.12 (SR Research Ltd.) with a 500 Hz sampling rate mounted on an MSI laptop with a 17.3 inch display and 1920 × 1080 pixels resolution. The display laptop was connected to a Lenovo T480 host laptop with Ethernet cable. The subtitles were styled using Aegisub (version 3.2.2) in deference to the Timed Text Style Guide of Netflix (2022) to best recreate natural watching experience. The subtitles were presented in NetflixSans Bold font, size 74 (in Aegisub unit). An area of interest (AOI) was assigned to each of the words in the subtitles (see Figure 1). The AOIs had a height of approximately 87 pixels, which are 1.5 times as high as the subtitles and had varied width depending on the length of the word. This size of the AOIs was determined by applying AOI sets of various heights to pilot data, and the set that captured as many data points as possible with minimum noise from surrounding distractions was chosen as the optimal set. The Python programs used to create the dynamic word-level AOI sets are available on the Github page associated with the paper. The experiment was created with free, open access software OpenSesame (Mathôt et al., 2012).

Procedure

Ethical clearance was obtained prior to the beginning of the intervention. All materials and procedures were piloted beforehand with 13 international students from the same university who shared characteristics with the participants in this study. The pilot was conducted in the context of a master's thesis (Fiche, 2023).

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

After analyzing the results, adjustments were made to several target items (e.g., unenhanced target items were added), and a second eye-tracking session was included, as the previous study only featured one session for the first episode. Once all materials were revised, an additional pilot with one participant was conducted before the recruitment process began.

Participants were recruited through on-campus flyers, university online media, and by word of mouth. Interested students received a link to information about the study and filled out a background questionnaire. Participants whose L1 was neither English, Dutch, nor German, had a low level of Dutch (no knowledge of Dutch—pre-A1), and had not watched *The Good Place* before were invited to the in-person session at the university eye-tracking lab.

During the first in-person session that took about 1.5 h, participants filled out consent forms, completed the VKS pretest, Dutch PPVT, English LexTALE, and English OPT prior to watching the first episode of the TV series with the eye-tracker. The participants were tested individually in a silent, darkened room, and they sat comfortably in front of the display laptop. The participants' dominant eye was recorded based on a determination of their eye dominance during the data collection session. The data were collected in the remote mode, and participants wore a sticker on their forehead, sitting approximately 60 centimeters from the screen. We opted for a remote rather than chin rest mode to create a more naturalistic viewing setting. A 13-point calibration was performed twice: At the beginning of the viewing, and in the middle of the episode, 11 minutes after viewing, when the participants had a chance to rest and readjust their position. Both calibrations were followed by the drift correction.

After watching the first episode, participants completed the post-viewing questionnaire and received instructions for viewing the remaining episodes. Over the next three weeks, participants watched episodes 2 to 12 on their own using the EdPuzzle website (Sabrià, 2013), which allows instructors to upload videos for students and monitor their progress. The system stops the video if a viewer changes tabs, does not allow rewinding or fast-forwarding, and shows instructors how many episodes participants watched and how long it took them to do so. To avoid binge watching and space out the viewing sessions (Pattemore & Muñoz, 2023), participants received links to episodes 2 to 5 during the first week, episodes 6 to 9 during the second week, and episodes 10 to 12 during the third week. After watching episode 12, participants returned to the eye-tracking lab to watch episode 13 with the eye-tracker, followed by a post-viewing questionnaire and a VKS posttest. This second in-person session took approximately 45 minutes and followed the same eye-tracking procedure as the first episode.

The intervention procedure is summarized in Figure 3. Participants received 25 euros for completing all parts of the intervention and were debriefed at the end of the second eye-tracking session.

Data preparation and analysis

Eye-tracking data for episode 1 (pretest) and episode 13 (posttest) were visually inspected in EyeLink Data Viewer to find excessive drifts. Data from one participant were removed due to the poor quality of the recording. For valid data, dwell time

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

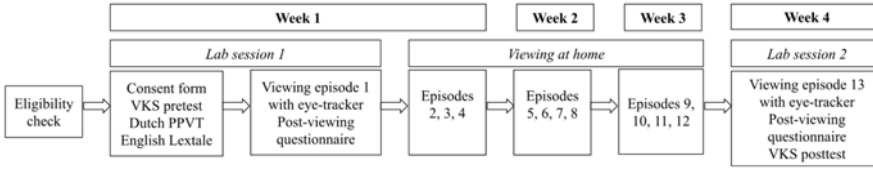

Figure 3. Study timeline and procedure.

(total time in milliseconds spent on an area of interest) for both enhanced and unenhanced targets was obtained from Data Viewer. One enhanced target ("Vreselijk") was removed from the first and the third research question analyses as it did not appear in Episode 13, and consequently, there was no data for it in the second viewing with the eyetracking. We kept this target for research question 2 since it was concerned with learning throughout the whole season. Zero-fixations were kept in the process, and were treated as skipping in the subsequent analysis. Skipping could be also informative, as looking at or not looking at a target are the only ways to visually interact with the stimuli (Godfroid, 2020, p. 213), hence the decision not to separate zero-fixations as usually done by previous studies (e.g., Liao et al., 2021; Wang & Pellicer-Sánchez, 2022). Rather, we adopted zero-inflation models to include these zero values while analyzing dwell time.

For the first research question examining attention allocation, the initial data sheet of dwell time contained 6044 observations. Only one datapoint had a dwell time under 100 ms and was therefore removed, leaving 6043. Following Godfroid (2020, pp. 267–269), outlier removal was performed through model criticism. After first fitting the models, data points with an absolute standardized residual exceeding 2.5 SD were removed. Eventually, the Gamma Hurdle model reported in this paper contains 5859 observations (97% of the original data) and was fitted using the "glmmTMB" package (Brooks et al., 2017).

For the second and third research question focusing on vocabulary learning data from the VKS pre- and posttest were coded either as known words (i.e., words for which participants scored 3 or more points) or unknown words (i.e., words for which participants scored 2 points or less). We then predicted this binary variable reflecting word knowledge using a generalized mixed-effects logistic regression using the 'glmer' function from the "lme4" package (Bates et al., 2015).

The data for the second research question contained all words from the VKS and, based on hypotheses, word knowledge was predicted based on the categorical predictors *Time* (Pretest vs Posttest) and *Word Category* (Enhanced, Unenhanced, and Filler) and their interaction. The word knowledge outcome variable comprised 2,112 observations from 22 participants, each completing a 48-item vocabulary test before and after the intervention (see Figure 5). The initial models used ordinal test scores from 1 to 5, but they did not converge. Therefore, we opted to predict learning using a binomial model.

Additionally, the potential impact of continuous variables reflecting participant's proficiency (PPVT score for Dutch and LexTALE for English) and item frequency in the show were assessed. The third research question aimed at predicting word learning based on attention was addressed in a similar fashion, except that fillers

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

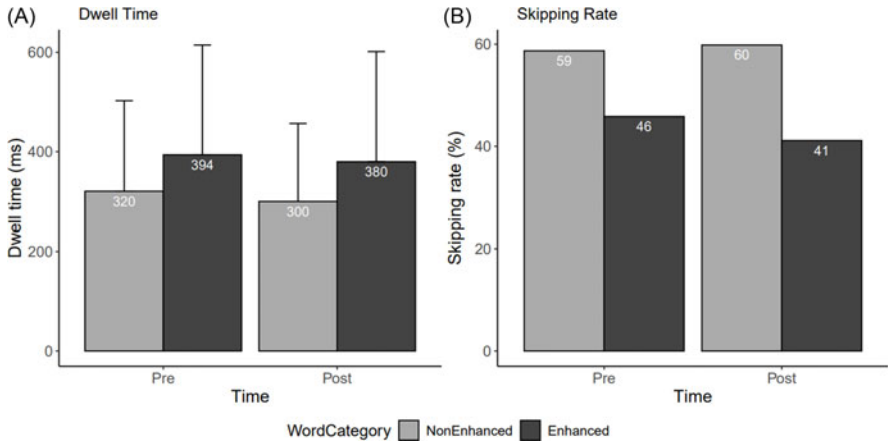

Figure 4. Barplots showing (**A**) how long participants spent looking at the words (in milliseconds) and (**B**) how often words were skipped depending on the *Word Category* (NonEnhanced vs Enhanced) and the *Time* of testing (Pretest vs Posttest).

were removed from the dataset (as there is no eye-tracking data available for these items) and eye-tracking measures (i.e., dwell time and skipping rate) were added as continuous fixed effects to the model.

All best-fitting models included *Participant* and *Item* as random effects, also to account for dependencies, and all coefficients reported are log-odds on log odds scale except for the Gamma part of the hurdle model whose coefficients are on log scale. To examine the differences between multiple groups, Tukey post hoc pairwise comparisons were performed using the "emmeans" package (Lenth, 2024). All analyses were performed in R version 4.4.1 (R Core Team, 2024). The data and scripts containing the most parsimonious models can be found on the Github page associated with this study.

Results

Does target enhancement affect attention?

The first model, the Binomial-Gamma Hurdle mixed model, addressed the question of whether *Word Category* (Enhanced vs NonEnhanced) and/or *Time* (Pretest vs Posttest) affected attention to the target words in the subtitles.

A significant effect of *Word Category* was found in both the conditional model (see Table 3 on the left) and the zero-inflation model (see Table 3 on the right). The latter model revealed that NonEnhanced words were skipped more often than Enhanced words (also see the barplot in Figure 4B). The main effect of *Word Category* in the Conditional model additionally revealed that, when targets were looked at (and not skipped), people overall spent more time looking at Enhanced words than Unenhanced words (also see Figure 4A).

As revealed by the significant negative estimate for *Time*, expected dwell time was 8% shorter at post- (episode 13) as compared to pretest (episode 1). No significant interaction between *Time* and *Word Category* was found in the conditional model

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

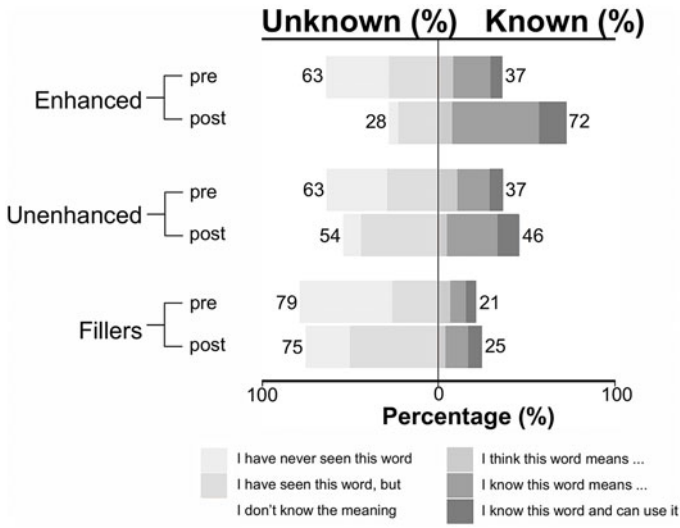

Figure 5. Plots showing the percentages of known and unknown words per *Word Category* and separated per *Time* of testing (Pretest vs Posttest). The data are depicted here both in its original ordinal scale as well as in the binary (Unknown—Known) distinction used in the analyses.

Table 3. Results of the Binomial-Gamma Hurdle mixed model predicting the chance that participants pay attention to words *(formula: DwellTime ~ Time * WordCategory + (1 | Participant) + (1 | Item), family = ziGamma(link = "log"))*

	Conditional Model		Zero-inflation Model	
Predictor	Coefficients (SE)	*z*	Coefficients (SE)	*z*
(Intercept)	5.74 (0.05)	115.01***	0.43 (0.17)	2.50*
Time = Post	−0.08 (0.03)	−2.75**	0.008 (0.08)	0.10
WordCategory = Enhanced	0.21 (0.05)	4.12***	−0.61 (0.14)	−4.24***
Time = Post* WordCategory = Enhanced	0.05 (0.04)	1.15	−0.28 (0.12)	−2.27*

Note: *$p < .05$, **$p < .01$, ***$p < .001$

predicting dwell time, but the zero-inflation model did reveal an interaction between these two variables. The barplot in Figure 4B suggests that this interaction is caused by a decrease in skipping of Enhanced words, but not Unenhanced words. Tukey post hoc pairwise comparisons confirmed this idea by showing a significant decrease in the probability of skipping Enhanced words ($p = .018$) and no significant change in skipping for NonEnhanced words over time ($p = .99$).

Does target enhancement affect word learning?

A second model was fit to examine the hypothesis that learning is affected by a combination of *Time* of testing (Pretest vs Posttest) and *Word Category* (Enhanced,

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Table 4. Results of the first mixed-effects binary logistic regression model predicting the chance that participants know a word *(formula: Known ~ Time * WordCategory + PPVT + (1 | Participant) + (1 | Item), family = "binomial")*

Random effects	Variance	*SD*	Correlation with Intercept
Item	3.74	1.94	
Participant	4.35	1.79	
Time = Post	1.07	1.03	−0.45

Fixed Effects	Estimate	SE	*z*-value	*p-value*
Intercept	−2.97	0.66	−4.5	<.001 ***
Time = Post	0.58	0.35	1.67	0.095.
WordCategory = NonEnhanced	192	0.74	2.59	0.009 **
WordCategory = Enhanced	1.68	0.74	2.28	0.023 *
PPVT	0.10	0.03	3.96	<.001***
Time = Post * WordCategory = NonEnhanced	0.36	0.34	1.087	0.28
Time = Post * WordCategory = Enhanced	2.79	0.37	7.45	<.001***

Unenhanced, and Filler items). The percentage of known words at pre- and posttest is visualized in Figure 5, and the details of the binary logistic mixed model predicting the probability of knowing a word can be found in Table 4.

A main effect of *Word Category* was found revealing that, when compared to fillers items, the chance of knowing a word was significantly higher for NonEnhanced and Enhanced words. This effect seemed to be driven by an interaction between *Time* and *Word Category* suggesting that a stronger learning effect from pre- to posttest was only found for Enhanced words (see Figures 5 and 6).

The interaction was further investigated using Tukey post-hoc comparisons and revealed that the chance of knowing a word was unaffected by *Word Category* at pre-test (*p*s. > .10). The chance of knowing fillers did not increase from pre- to posttest ($p = .55$), but the chance of knowing NonEnhanced and Enhanced words significantly increased over time ($p = .038$ and $p < .0001$, respectively). When compared to filler words, the chance of knowing NonEnhanced and Enhanced words was significantly larger at posttest ($p = .024$ and $p < .0001$, respectively). Additionally, the chance of knowing a word at posttest was significantly higher for Enhanced as compared to NonEnhanced words ($p = .029$). The interaction between *Time* and *Word Category* is also visualized in Figure 6.

Dutch vocabulary proficiency (i.e., *PPVT* score) additionally significantly increased the overall chance of knowing words. As can be seen in Table 4, no main effect of *Time* was found, but a random slope for *Time* did improve model fit ($\chi 2\ (2) = 17.9$, $p < .001$) suggesting that the overall learning effect from pre- to posttest differed between individuals.

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

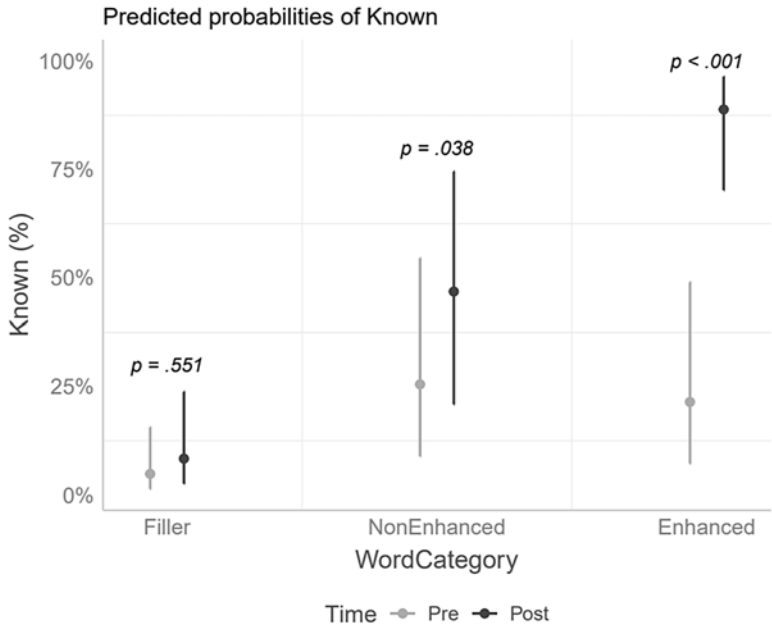

Figure 6. The predicted probabilities of the binary logistic mixed model predicted the chance of knowing a word based on *WordCategory* (Filler vs NonEnhanced vs Enhanced) and the *Time* of testing (Pretest vs Posttest).

To further explore learning, we analyzed responses from the post-viewing questionnaire completed after Episodes 1 and 13. Overall, participants expressed a positive attitude toward plurilingual audiovisual input, with 77% (17) agreeing that drawing on multiple languages enhanced their understanding of the TV series. In contrast, 5 (23%) participants disagreed with this statement. Notably, a majority (77%, 17) reported finding it challenging and distracting to watch in two different languages. Despite this difficulty, 68% (15) indicated that they would continue engaging with TV content in this format to improve their Dutch proficiency.

As for learning outcomes, the majority of participants reported a positive experience after watching the episodes, with 86% (19) agreeing that they learned from viewing the TV series. Only 14% (3) remained neutral, and no participants disagreed with the statement. Participants were also asked to indicate any words they felt they had learned from the episodes, and Table 5 lists these reported words along with the number of students who noted each one (in brackets). After Episode 1, a total of 39 words were recalled by the 22 participants, of which 31 (79.5%) were Enhanced target words, 3 (7.7%) were NonEnhanced target words, and 5 (12.8%) were non-target words not included in the pretest. After completing the full season, participants collectively recalled 72 words, with 45 (62%) being Enhanced target words, 7 (10%) NonEnhanced target words, and 20 (28%) non-target words. These results indicate that most items reported as learned were textually enhanced, although participants also noted some unenhanced and non-target words and expressions.

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Table 5. Self-reported learning data

	Reported learning after Episode 1 (n)	Reported learning after Episode 13 (n)
Enhanced target word	mensen—people (7) dood—dead (5) leuk—fun (4) vreselijk—awful (2) trouwens—by the way (3) soulmate[a] (4) natuurlijk—of course (1) iemand—somebody (1) vermoord—killed (1) baas—boss (1) gebeurt—happen (1) geloven—believe (1)	mensen—people (6) dood—dead (5) alleen—only (4) trouwens—by the way (4) zielsverwant—soulmate (4) natuurlijk—of course (4) misschien—maybe (3) alles—everything (4) geloof—believe (2) waarom—why (2) leuk—fun (2) vermoord—killed (2) jongens—guys (2) gebeurt—to happen (1)
NonEnhanced target words	zeggen—say (1) slecht—bad (1) toch—right/yet/still (1)	iedereen—everybody (4) slecht—bad (3)
Non-target words	ik—I (1) jij– you (1) goed—good (1) plek—place (1) schat—dear (1)	ik—I (2) houd van jou—love you (1) niet—no (1) zon—sun (1) paradijs—paradise (1) kijken—look (1) verlieds—lost (1) kom binnen—welcome (1) nodig—required (1) bijlengrijk—important (1) plek—place (1) goeie—good (1) martelen—torture (1) grot—big (1) anderen—others (1) vraag—question (1) schat—dear (1) ik kom uit—I come from (1) goede plaats—good place (1)

[a]Participants did not provide the Dutch form of the word, but reported learning it in Dutch. However, as shown in Table 6, when completing the questionnaire after watching Episode 13, participants were able to write the word in Dutch.

Is this learning associated with eye-tracking measurements?

When fillers were removed from the dataset in order to assess the potential impact of eye-tracking measures on learning, the interaction between *Time* and *Word Category* remained and showed a significantly larger increase in learning for Enhanced as compared to NonEnhanced target words (see Table 6). Without fillers, the main effect of *Word Category* was not significant, but the main effect of *Time* did reach significance in this model with an overall higher chance of knowing target words at posttest. As in the second model, a higher *PPVT* score increased the chance of knowing words. This time, we could also assess the potential effect of the frequency of occurrence of target words on the show. The significant positive

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Table 6. Results of the second mixed-effects binary logistic regression model predicting the chance that participants know a word *(formula: Known ~ Time * WordCategory + PPVT + FrequencyShow + (1 | Participant) + (1 | Item), family = "binomial")*

Random effects	Variance	SD		
Item	1.93	1.4		
Participant	2.20	1.5		
Fixed Effects	**Estimate**	**SE**	***z*-value**	***p*-value**
Intercept	−2.71	0.87	−3.11	.002**
Time = Post	0.73	0.21	3.48	<.001***
WordCategory = Enhanced	0.33	0.56	0.59	0.56
PPVT	0.09	0.02	3.82	<.001***
FrequencyShow	0.03	0.01	2.52	0.012*
Time = Post* WordCategory = Enhanced	2.36	0.33	7.04	<.001***

Estimate for *FrequencyShow* revealed that words that occurred more often in the subtitles had more chance of being known. *Dwell Time* and the *Skipping* of words did not significantly impact the chance of knowing a word, and therefore were removed for a better model fit (Table 6).

Discussion

This study set out to explore the processing and effectiveness of textually enhanced (TE) subtitles during exposure to plurilingual audiovisual input, specifically with L2 English audio and L3+ Dutch subtitles, over the course of a 13-episodes TV series season. Participants' processing of both enhanced and unenhanced subtitles was recorded at the beginning and end of the intervention, along with pre- and posttest measurements of learning unenhanced, enhanced, and filler words. The study's driving hypothesis was the assumption that attention to textually enhanced on-screen text might decrease over time and that the effectiveness of TE subtitles may diminish in longitudinal interventions as opposed to cross-sectional ones (Indrarathne et al., 2018; Pattemore & Muñoz, 2022).

Attention allocation to target words over time

The first research question examined the attention paid to (un)enhanced target words over the course of the intervention. Results showed that, during the first episode, participants spent significantly more time reading enhanced words compared to unenhanced ones, as measured by dwell time and skipping rates. This aligns with previous cross-sectional studies showing that textual enhancement attracts more attention than unenhanced targets when viewers engage in a single viewing (e.g., Puimège et al., 2021). However, when analyzing eye-tracking data for the 13th episode, we found that the significant difference in attention allocation between enhanced and unenhanced targets remained. Although participants spent

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

on average 8% less time on target words in Episode 13—contrary to our predictions—textually enhanced (TE) target words continued to attract more attention and were skipped less often than unenhanced targets. This finding does not confirm our hypothesis that there would be a stronger decrease in attention to TE target words. The insignificant decrease in attention to textually enhanced (TE) captions does not align with the results of another eye-tracking study that recorded eye-tracking data at the beginning and end of the intervention (Finger-Bou & Muñoz, 2023). In that study, attention to textually enhanced words significantly decreased between the first and second eye-tracking sessions. Although we observed an overall 8% decrease in attention, this reduced dwell time could not be attributed solely to textual enhancement, as attention to unenhanced targets also decreased. It is possible that by the end of the intervention, participants were relying more on the English audio than on the Dutch subtitles, which may explain the decrease in attention, rather than continuous exposure to textual enhancement. Additionally, differences in the results of the two studies could be attributed to the age difference in their samples, as previous research suggests that subtitle processing differs between adults and children (Muñoz, 2017). It is possible that adults in our study continued to attend to textual enhancement, whereas primary school children in Finger-Bou and Muñoz (2023) reduced their attention to it. Given that only two studies, including the present one, have investigated extended exposure to TE target words over time, conclusions remain premature. More research with varied learner populations and input conditions is needed to determine whether attention to textual enhancement changes over time and, if so, in what direction. In addition, future research should include more than two eye-tracking sessions to capture potential changes in attention to TE in greater detail.

Meanwhile, our study provides initial evidence that TE subtitles continue to promote a significant amount of attention to enhanced words even after prolonged exposure.

Learning from (un)enhanced subtitles

Regarding learning outcomes from the intervention, our binomial model indicated that both enhanced and unenhanced words were more likely to be known by participants after the intervention than filler words that did not appear in the episodes, providing additional evidence of the benefits of exposure to plurilingual subtitled audiovisual input (cf. Pattemore et al., 2024a, 2024b; Urbanek & De Vogelaer, 2025). Notably, enhanced words were significantly more likely to be known than unenhanced words at the posttest. In addition, when we asked our participants to report what they learnt from the episodes, the majority of the words that participants recalled were textually enhanced after watching both Episode 1 and Episode 13.

This finding contrasts with most previous cross-sectional vocabulary studies, which found no significant difference between enhanced and unenhanced groups (Montero Perez et al., 2014; Majuddin et al., 2021; Finger-Bou & Muñoz, 2023). However, it aligns with a within-subjects study by Puimège et al. (2021), which observed significantly greater learning of textually enhanced than unenhanced multiword units. Our results also support the prolonged exposure study by Finger-

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Bou and Muñoz (2024), which found a significant effect of textual enhancement on vocabulary learning compared to unenhanced and uncaptioned conditions. These findings suggest that short interventions may not fully capture the learning potential of textual enhancement and that extended exposure and within-subject design may be necessary to observe significant effects.

In contrast, our findings diverge from those of Pattemore and Muñoz (2022), who also implemented a prolonged viewing intervention and found that the uncaptioned group outperformed the enhanced group after watching ten episodes of the same TV series. Since Pattemore and Muñoz (2022) focused on constructions (e.g., *to be allowed to*) rather than single words, as in the present study, this discrepancy may stem from differences in target items. Although Puimège et al. (2021) also focused on multiword units and found significantly higher learning gains for enhanced targets, their assessment took place immediately after viewing, capturing immediate gains. Pattemore and Muñoz (2022) also observed immediate learning gains when testing constructions after every two episodes, though these gains diminished over time as demonstrated by the posttest taken after watching all ten episodes.

Comparing the findings of the present study and Pattemore and Muñoz (2022), it seems that while isolating single words through textual enhancement may be effective over time, constructions may require additional contextual cues that are harder to access when most attention is directed to the enhanced segment of a sentence. This issue is particularly relevant in dynamic subtitles, where text appears on screen for only a limited time, making it likely that excessive focus on enhanced targets could lead to overlooking other parts of the subtitle.

Regarding the third research question on the effect of eye-tracking measures on learning, we found that including eye-tracking data in the model did not affect the likelihood of words being known, but enhanced words were still more likely to be known than unenhanced ones. Interestingly, in the study by Puimège et al. (2021), adding eye-tracking measures caused the significant effect of textual enhancement to disappear, underscoring differences between cross-sectional and prolonged viewing studies. Overall, our findings suggest that textual enhancement was effective regardless of the time participants spent on the words. Instead, the factors that influenced learning outcomes were participants' Dutch vocabulary size and the frequency of target word occurrence across the thirteen episodes. This aligns with previous research, which indicates that vocabulary size, proficiency level, and word frequency are significant predictors of learning outcomes in prolonged audiovisual input interventions (Muñoz et al., 2021).

These findings, showing an advantage for textually enhanced (TE) targets over unenhanced ones and relatively stable attention allocation to enhanced words even after watching 13 episodes of the TV series (284 minutes of input), raise the question of whether TE targets were learned at the expense of unenhanced items. Several studies have identified a trade-off effect with textual enhancement (Galimberti et al., 2023; Majuddin et al., 2021), either in terms of audio-text synchronisation or overall comprehension. Textual enhancement can attract attention to specific words, prompting viewers to read the enhanced words before the corresponding audio, which can lead to reduced attention to the other, unenhanced words in the same subtitle line (Galimberti et al., 2023). This isolated focus on TE words may, in turn,

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

decrease content comprehension (Majuddin et al., 2021) and result in lower learning gains for unenhanced words, as observed in the present study.

To better distinguish the positive effects of textual enhancement from its potential trade-off effects on unenhanced words, future research should include a comparison group that follows the same research design but watches the episodes without any textual enhancement. First, this approach would clarify whether enhanced words were learned significantly more due to textual enhancement rather than inherent word properties. Although we carefully matched the enhanced, unenhanced, and filler items as closely as possible (see methodology above), other factors may still have influenced learning outcomes. Second, if a follow-up study shows that, in the absence of textual enhancement, the difference between what are now enhanced and unenhanced targets disappears, this would suggest that textual enhancement indeed diverted attention away from other words and that removing it allows a more balanced learning of all words.

Limitations

This study is not without limitations. First, the relatively small number of participants, all from the same university, may limit the generalizability of our findings. While participants came from different countries and fields of study, a multisite study in the future would allow for greater diversity and broader applicability. Second, our assessment primarily measured meaning recall, as we opted to use a binary scoring system instead of the ordinal measure provided by the Vocabulary Knowledge Scale due to model convergence issues. It means that this test does not capture other dimensions of word knowledge, particularly form recall. In addition, presenting the target words during the pretest, which occurred in the same session as the start of the viewing, may have inadvertently primed participants' attention toward those words. Although we attempted to mitigate this by including filler items in the pre- and posttests, future research could consider scheduling the pretest a few weeks prior to the start of the viewing phase to further reduce potential testing effects.

Finally, while we included metrics such as dwell time and skipping to gauge reading behavior, these measures alone may not offer a sufficiently nuanced picture of how textual enhancement impacts vocabulary learning. A more comprehensive analysis—beyond the word limit of this paper—could incorporate measures such as fixation duration, regression patterns, or sequential attention to enhanced and unenhanced words to yield deeper insights into the interaction between reading behavior and vocabulary acquisition.

Despite these limitations, our study provides valuable evidence on the effects of textual enhancement in a prolonged exposure context using subtitled audiovisual input. Our findings suggest that textual enhancement consistently attracts attention over time and that textually enhanced words are learned significantly more than unenhanced words.

https://doi.org/10.1017/S0142716425100258 Published online by Cambridge University Press

Conclusion

Our study provides new insights into the impact of textual enhancement on vocabulary learning within a prolonged exposure to plurilingual audiovisual input. Findings reveal that enhanced subtitles not only consistently attracted attention throughout the intervention but also facilitated significantly higher learning of enhanced words compared to unenhanced words. This suggests that prolonged exposure is essential to capture the full potential of textual enhancement, which may not be as evident in shorter interventions as evidenced by previous research. Additionally, the stable attention to enhanced words, even after extended viewing, indicates that enhanced subtitles can maintain engagement without substantial declines over time. Future research should further examine the long-term trade-offs between attention to enhanced versus unenhanced words, particularly by including a group with unenhanced subtitles to assess the effect of enhancement on the same target words learning. Additionally, examining textual enhancement effects on more complex linguistic targets, like constructions or multiword expressions, could clarify whether these require more contextual support rather than textual enhancement for effective learning in subtitling settings. Together, these findings underscore the potential of textual enhancement in promoting vocabulary learning in audiovisual input contexts, especially for single-word meaning recall learning.

Replication package. Replication data and materials for this article can be found at https://github.com/Vincenzofan/textual_enhancement.

Standardized assessment (PPVT) and TV series episodes cannot be publicly shared because these materials are copyrighted by the publisher. LexTale is an open-access proficiency measure and is freely available at https://www.lextale.com/.

Acknowledgments. We would like to thank the Center for Language and Cognition (CLCG) at the University of Groningen for their financial support with this study. We are grateful to Tongyao for all the help at the beginning of the project.

Funding statement. Open access funding provided by University of Groningen.

Competing interests. The authors declare no conflict of interest.

Ethics statement. This study was approved by the University of Groningen Ethics Committee [ID 92373223]. The participants were informed about the nature of the study and signed informed consent prior to data collection.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: an eye tracking study. *Applied Psycholinguistics*, **35**(2), 399–418. https://doi.org/10.1017/S0142716412000434

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, **9**(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Choi, S. (2023). Visual saliency in captioned digital videos and learning of English collocations: an eye-tracking study. *Language Learning & Technology*, **27**(1), 1–21. https://hdl.handle.net/10125/73536

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: a comparison of monolingual and bilingual reading of a novel. *PLoS ONE*, **10**(8). https://doi.org/10.1371/journal.pone.0134008

Deygers, B., & Vanbuel, M. (2022). Gauging the impact of literacy and educational background on receptive vocabulary test scores. *Language Testing*, **39**(2), 191–211. https://doi.org/10.1177/02655322211049097

Duarte, J. (2020). Translanguaging in the Context of Mainstream Multilingual Education. *International Journal of Multilingualism*, **17**(2), 232–247. https://doi.org/10.1080/14790718.2018.1512607

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). Peabody Picture Vocabulary Test-III-NL. Handleiding. Pearson Assessment and Information B.V.

European Commission, Directorate-General for Communication. (2023). Standard Eurobarometer 100 – Autumn 2023. Media use in the European Union. European Union. Retrieved August 7, 2024, from: https://op.europa.eu/en/publication-detail/-/publication/953c23c3-c3d9-11ee-95d9-01aa75ed71a1

Fiche, L. (2023). The effectiveness of subtitle enhancement in TV series on target language vocabulary learning. [Unpublished Master's thesis]. University of Groningen. https://arts.studenttheses.ub.rug.nl/id/eprint/33520

Finger-Bou, R., & Muñoz, C. (2023). The effects of regular and enhanced captions on incidental vocabulary acquisition. *ELIA: Estudios de lingüística inglesa aplicada*, **23**, 15–50. https://doi.org/10.12795/elia.2023.i23.01

Galimberti, V., Mora, J. C., & Gilabert, R. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System*, **116**, 103078. https://doi.org/10.1016/j.system.2023.103078

Gesa, F., & Miralpeix, I. (2023). Extensive viewing as additional input for foreign language vocabulary learning: a longitudinal study in secondary school. *Language Teaching Research*. https://doi.org/10.1177/13621688231169451

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge. https://doi.org/10.4324/9781315775616

Godfroid, A., Ahn, J., Choi, I. N. A., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition*, **21**(3), 563–584. https://doi.org/10.1017/S1366728917000219

Indrarathne, B., Ratajczak, M., & Kormos, J. (2018). Modelling changes in the cognitive processing of grammar in implicit and explicit learning conditions: insights from an eye-tracking study. *Language Learning*, **68**(3), 669–708. https://doi.org/10.1111/lang.12290

Jelani, N. A. M., & Boers, F. (2020). Examining incidental vocabulary acquisition from captioned video: does test modality matter? In S. Webb (Ed.), *Approaches to learning, testing and researching L2 vocabulary* (pp. 169–189). John Benjamins. https://doi.org/10.1075/bct.109.itl.00011.jel

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, **42**(3), 643–650.

Kruger, J.-L., Wisniewska, N., & Liao, S. (2022). Why subtitle speed matters: evidence from word skipping and rereading. *Applied Psycholinguistics*, **43**(1), 211–236. https://doi.org/10.1017/S0142716421000503

Lee, M., & Révész, A. (2020). Promoting grammatical development through captions and textual enhancement in multimodal input-based tasks. *Studies in Second Language Acquisition*, **42**(3), 625–651. https://doi.org/10.1017/S0272263120000108

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of english. *Behavior Research Methods*, **44**, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lenth, R. (2024). _emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.10.4, <https://CRAN.R-project.org/package = emmeans>.

Leow, R. P., & Martin, A. (2017). Enhancing the input to promote salience of the L2: a critical overview. In S. Gass, Spinner, P., & Behney, J (Eds.) *Salience in second language acquisition* (pp. 167–186). Routledge. https://doi.org/10.4324/9781315399027-9

Liao, S., Yu, L., Reichle, E. D., & Kruger, J. L. (2021). Using eye movements to study the reading of subtitles in video. *Scientific Studies of Reading*, **25**(5), 417–435. https://doi.org/10.1080/10888438.2020.1823986

Majuddin, E., Siyanova-Chanturia, A., & Boers, F. (2021). Incidental acquisition of multiword expressions through audiovisual materials: The role of repetition and typographic enhancement. *Studies in Second Language Acquisition*, **43**(5), 985–1008. https://doi.org/10.1017/S0272263121000036

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, **44**(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: an eye-tracking study. *The Language Learning Journal*, **47**(4), 460–478. https://doi.org/10.1080/09571736.2019.1638623

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, **55**(2), 163–192. https://doi.org/10.1017/S0261444821000501

Montero Perez, M., Peters, E., Clarebout, G. & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, **18**(1), 118–141. http://dx.doi.org/10125/44357

Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: a meta-analysis. *System*, **41**(3), 720–739. https://doi.org/10.1016/j.system.2013.07.013

Mora, J. C., & Fouz-González, J. (2024). Contrastive input enhancement in captioned video for L2 pronunciation learning. In C. Muñoz & Miralpeix, I. (Eds.) *Audiovisual input and second language learning* (pp. 150–175). John Benjamins. https://doi.org/10.1075/lllt.61.07mor

Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, **67**, 77–86. https://doi.org/10.1016/j.system.2017.04.015

Muñoz, C., Pujadas, G., & Pattemore, A. (2021). Audio-visual input for learning L2 vocabulary and grammatical constructions. *Second Language Research*, **39**(1), 13–37. https://doi.org/10.1177/02676583211015797

Pattemore, A. (2023, December). Watching TV in two foreign languages: What does it bring to language learning? [Conference presentation]. Vocab@Vic 2023 conference, Wellington, New Zealand.

Pattemore, A., Cabrera Fernández, B., Lopez de Artola, M., & Michel, M. (2024b). Maximising the potential of plurilingual subtitled audiovisual input: learning L3 words and multiword units with L2 audio. *Innovation in Language Learning and Teaching*, 1–16. https://doi.org/10.1080/17501229.2024.2366254

Pattemore, A., & Muñoz, C. (2022). Captions and learnability factors in learning grammar from audio-visual input. *The JALT CALL Journal*, **18**(1), 83–109. https://doi.org/10.29140/jaltcall.v18n1.564

Pattemore, A., & Muñoz, C. (2023). The effects of binge-watching and spacing on learning L2 multi-word units from captioned TV series. *The Language Learning Journal*, **51**(4), 401–415. https://doi.org/10.1080/09571736.2023.2211614

Pattemore, A., Suárez, M.-M., & Muñoz, C. (2024a). Perceptions of learning from audiovisual input and changes in L2 viewing preferences: the roles of on-screen text and proficiency. *ReCALL*, **36**(2), 135–151. https://doi.org/10.1017/S0958344024000065

Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: the differential effect of L1 subtitles and captions. *System*, **63**, 134–148. https://doi.org/10.1016/j.system.2016.10.002

Puimège, E., Montero Perez, M., & Peters, E. (2021). Promoting L2 acquisition of multiword units through textually enhanced audiovisual input: an eye-tracking study. *Second Language Research*, **39**(2), 471–492. https://doi.org/10.1177/02676583211049741

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rodgers, M. P. H., & Webb, S. (2011). Narrow viewing: the vocabulary in related television programs. *Tesol Quarterly*, **45**(4), 689–717. https://doi.org/10.5054/tq.2011.268062

Sabrià, Q. 2013. Edpuzzle. https://edpuzzle.com/.

Schur, M. (Creator). (2016). *The good place*. [TV series]. Fremulon.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA: theoretical bases. *Studies in Second Language Acquisition*, **15**, 165–179. https://doi.org/10.1017/S0272263100011943

Urbanek, L., & De Vogelaer, G. (2025). Subtitling in passive and active use in L3-Dutch teaching in German schools and its potential for incidental vocabulary acquisition. In: A. Pattemore, & F. Gesa (Eds.),

Foreign language learning from audiovisual input. educational linguistics, vol. **66**. Springer. https://doi.org/10.1007/978-3-031-91001-2_3

Vanderplank, R. (2016). *Captioned media in foreign language learning: subtitles for the deaf and hard-of-hearing as tools for language learning*. London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-50045-8

Vu, D. V., & Peters, E. (2022). Learning vocabulary from reading-only, reading-while-listening, and reading with textual input enhancement: insights from Vietnamese EFL learners. *RELC Journal*, **53**(1), 85–100. https://doi.org/10.1177/0033688220911485

Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: an eye-tracking study. *Language Learning*, **72**(3), 765–805. https://doi.org/10.1111/lang.12495

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth vs. breadth. *The Canadian Modern Language Review*, **53**(1), 13–40. https://doi.org/10.3138/cmlr.53.1.13

Wilkinson, D., & Apic, O. (2018). Language Reactor [Computer Software]. Retrieved from https://www.languagereactor.com/

Cite this article: Pattemore, A., Fan, V., Fiche, L., & Loerts, H. (2025). Investigating the effects of prolonged exposure to textual enhancement on attention and learning: a pre-posttest measures eye-tracking study. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425100258