Combining eye-tracking and verbal reports in vocabulary research: Benefits and challenges

Andi Wang and Ana Pellicer-Sanchez

Abstract

Eye-tracking has been increasingly used as a reliable measure of second language learners' attention in vocabulary learning studies. While providing a rich measure of online processing behaviour and attention to lexical items, the examination of eye movements is not enough to explore all underlying cognitive subprocesses. Researchers have recently suggested combining eye-tracking with verbal reports to obtain a fuller picture of learners' cognitive processes. The current commentary discusses how the combination of eye-tracking and stimulated recalls can shed light into the cognitive processes underlying vocabulary learning. Using data from a mixed methods study on incidental vocabulary learning from viewing, we present detailed examples of how eye-tracking and stimulated recalls can be combined in vocabulary learning research. A discussion of the methodological benefits and challenges of combing these two research methods is also provided.

1. Introduction

Eye-tracking allows the real-time, online, and direct recording of an individual's eye-movement behaviour during information processing, and has been widely used in a range of

applied linguistics domains (Godfroid, 2020). It allows researchers to obtain a rich record of the rapid movements of the eyes (i.e., *saccades*), when the eyes stop to process information (i.e., *fixations*), and movements back in a text when reading (i.e., *regressions*; see Conklin, et al., 2018 and Godfroid, 2020, for detailed discussions). Eye movements are direct measures of allocation of overt attention and are closely related to covert attentional processes (Rayner, 2009). Eye-tracking is now frequently used to examine the cognitive processes underlying second language (L2) vocabulary learning from different input conditions, providing new insights into the vocabulary learning process. Eye-tracking studies on L2 vocabulary learning from reading have shown increased attention to unknown lexical items when they are first encountered in a text (e.g., Godfroid, et al., 2013; Pellicer-Sánchez, 2016), with number and duration of fixations decreasing with increased exposures to both single words (e.g., Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez et al., 2021) and formulaic sequences (e.g., Pellicer-Sánchez et al., 2022). Recent studies have also shown that textual enhancement techniques seem to further increase initial attention to unknown lexical items (e.g., Puimège et al., 2023). A few studies on learning from reading have also suggested that reading times on unknown vocabulary are significant predictors of learning gains (e.g., Godfroid et al., 2013; Pellicer-Sánchez, 2016; Pellicer-Sánchez et al., 2021). However, other studies have failed to show the predictive role of the amount of attention on learning gains (e.g., Elgort et al., 2018; Ouyang et al., 2020).

In the context of vocabulary learning from viewing (i.e., watching audio-visual materials, such as television shows and movies), studies have shown that adult learners process the animation and on-screen text regardless of the language of the text, with similar processing patterns for first language (L1) subtitles (i.e., on-screen text in viewers' L1) and captions (i.e., on-screen text in the same language as the soundtrack; e.g., Bisson, et al., 2014). Empirical evidence has also suggested that early processing of unknown lexical items facilitates learners'

knowledge of word form, whereas the predictive role of late measures is still unclear (e.g., Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022b).

Eye-tracking has clearly contributed to a better understanding of the vocabulary learning process, allowing researchers to observe attention allocation to unknown lexical items during learning and gaining a clearer (yet inconclusive) picture of the relationship between attention and vocabulary learning. However, although eye movements are believed to be a robust physiological measure to study attention (Leow, 2015), eye-tracking cannot always disambiguate the various subprocesses that underlie eye-movement measures. For example, as shown above, eye-tracking studies on vocabulary learning have reported increased processing times on unknown lexical items when reading and viewing, and this has been interpreted as a reflection of increased cognitive effort. However, those increased reading times could reflect participants' application of different processing strategies. They could reflect readers' conscious effort to guess the meaning of the new lexical item from contextual cues and encode it to memory, or they could also be reflecting comprehension difficulties. Previous studies have indeed shown that learners use various strategies to process unknown words in reading (e.g., Fraser, 1999; Hu & Nassaji, 2012) and viewing (e.g., Sydorenko, 2010), but the different processing strategies cannot be unpacked in eye movement data. Similarly, previous research has shown that learners' awareness of unknown vocabulary in reading is a stronger predictor of vocabulary recognition (Godfroid & Schmidtke, 2013), but different levels of awareness cannot be identified in eye movement data. This has led researchers to claim that other data sources should be used to gain a more comprehensive understanding of the vocabulary learning process (Pellicer-Sánchez, 2020).

Apart from eye-tracking, which relies on objective observation of individual's behaviour to infer cognition, other types of introspection methods have been used in the field. One of the most frequently used forms of introspection is verbal reports, which assume that individuals

can observe and verbalise their own internal processes at some level as dealing with external events (Gass & Mackey, 2017). Researchers have suggested the use of verbal reports as a direct method to elicit learners' thoughts and tap into their internal processes during their engagement with language input (Hiver et al., 2021). Think-aloud protocols and stimulated recalls are the most frequently used introspective methods in L2 research. They are helpful in gathering information about thought processes that occur during an event (Gass & Mackey, 2017). Think-aloud protocols require collecting data concurrently with language production, whereas stimulated recalls are conducted after a language event using a prompt to support learners' memory retrieval (Gass & Mackey, 2017). Stimulated recalls aim at examining cognitive processes through asking participants to recall the thoughts they had while completing a task through a posteriori recall session (Gass & Mackey, 2017). The main limitation of stimulated recall is veridicality, as participants may not accurately recall their thought processes after the task due to potential memory decay, but this limitation can be alleviated with a short amount of intervening time between the task and the recall (Gass & Mackey, 2017). Although think-aloud protocols are helpful for providing valid data on participants' spontaneous task-related thoughts without corrupting or changing their memory (Ericsson, 2002), they are not suitable for research recording eye-movement data due to reactivity, as verbalising one's thoughts increases the time on task, resulting in the distortion of eye movements (Godfroid & Schmidtke, 2013). Therefore, the various subprocesses reflected in eye-movement data could be better isolated and identified through the combination of eye-tracking and stimulated recall data.

An initial attempt to combine eye-tracking and stimulated recalls in vocabulary learning research was conducted by Godfroid and Schmidtke (2013). Results of their triangulation of eye movements, vocabulary tests scores, and stimulated recalls showed longer fixations and better learning for words that participants remembered having seen in context, providing

empirical evidence for the relationship between eye movements, level of awareness, and vocabulary learning. Later, Jung and Révész (2018) also triangulated data from eye-tracking (as a measure of attention) and stimulated recalls (as a measure of awareness) to examine the effects of different reading activities on L2 learners' reading processes and their processing of glossed pseudowords in L2 reading. Although their findings revealed no impact of activity type on learners' attention or awareness to pseudowords, their findings further suggested the benefits of combining these two methods in gaining a fuller picture of reading operations. It should be noted, however, that different from Godfroid and Schmidtke (2013), who used quantitative approach to analyse and triangulate the eye-tracking and stimulated recall data, Jung and Révész (2018) analysed these two types of data separately, and the triangulation of data was based on a discussion of general patterns rather than running statistical analyses focusing on each pseudoword. Different research questions would require different methods to triangulate eye-tracking and stimulated recall data. However, while these initial studies show the potential of the triangulation of eye-movement data with verbal reports, there has been very little discussion in the field of how this could be done. Thus, the aim of the present commentary is to start this discussion, illustrating how eye-tracking and stimulated recalls can be combined to explore the cognitive processes underlying vocabulary learning.

2. Methodological options in the combination of eye-tracking and stimulated recalls

Against this background, we turn to discussing in detail how recordings of eye movements and stimulated recalls can be combined to gain a fuller picture of the cognitive processes that underlie vocabulary learning. As argued above, fixation durations and amount of attention paid to unknown vocabulary during reading and viewing can be related to various processing strategies and different levels of awareness, but eye-tracking data is not sufficient to unpack these subprocesses. In this section, we present two methodological options to

investigate these two key aspects of cognitive processing during vocabulary learning. We present two worked examples using data collected by Wang (2022) on learning from subtitled viewing. Wang's (2022) research aimed to investigate the effects of different subtitling types (i.e., captions, L1 subtitles, and bilingual subtitles) on L2 learners' comprehension (Wang & Pellicer-Sánchez, 2022a) and incidental vocabulary learning (Wang & Pellicer-Sánchez, 2022b), and explored learners' engagement with unknown words during viewing using eye-tracking and stimulated recalls. In Wang's (2022) research, a number of unknown words from the video were selected as target words (TWs) and participants' prior knowledge of those words was tested by means of pre-tests. Participants' eye movements were recorded during viewing and, after the viewing, all participants participated in stimulated recall interviews to explore whether they were aware of the TWs and what they had done when encountering the TWs in the video. Analyses were conducted only with the items that were unknown to participants, as indicated in both the pre-test and participants' self-reports.

2.1. Example 1

2.1.1. Research Question

Is participants' reported awareness of the unknown vocabulary associated with the amount of attention paid during captioned viewing?

2.1.2. Rationale and Aim

As argued earlier, eye-tracking is considered a reliable measure of learners' overt attention (Godfroid, 2020; Rayner, 2009). Attention is crucial for L2 input to become intake, which contributes to L2 learning (Schmidt, 2001; Tomlin & Villa, 1994). Ellis (1994) claims that attention is necessary and sufficient for learning word forms, but learning word meanings requires both attention and awareness. This seems to suggest that attention itself might not

sufficiently contribute to vocabulary learning. Although attention and awareness are closely related (Schmidt, 2001), some researchers argue that awareness entails attention, but attention can occur without awareness (Tomlin & Villa, 1994). Eye-tracking, as a measure of learners' attention, cannot distinguish learners' awareness of the language item itself from only attention to the textual layout (Godfroid & Schmidtke, 2013). This could explain why some eye-tracking studies have failed to find a relationship between amount of attention, as reflected in eye-movement measures, and scores in vocabulary tests. One way to tease apart attention from awareness is to collect information about awareness through qualitative methods and separately explore eye movements for those words for which participants report awareness and for those for which no awareness is reported (see Godfroid & Schmidtke, 2013). In this commentary, awareness is operationalised as the availability for self-reporting either during or immediately after exposure to input (Leow, 2015).

Another important limitation of eye-tracking identified in the context of learning from viewing is that, while eye-tracking provides a reliable measure of over attention to written stimuli, it cannot provide information about learners' attention to the auditory stimuli and their awareness to the aural form of unknown vocabulary. The combination of eye-tracking with stimulated recalls would allow researchers to obtain a more comprehensive picture of learners' awareness of unknown words in captioned viewing, contributing to a better understanding of the relationship between learners' attention and vocabulary learning gains.

The main aim of Example 1 is to explore whether learners' reported awareness of L2 unknown words is associated with the amount of visual attention paid to those words during captioned viewing. Learners' visual attention is measured using eye-tracking, and awareness is measured using stimulated recalls (see also Jung & Révész, 2018). In this context, stimulated recall is considered suitable to examine learners' reported awareness (both auditory and visual awareness) of the unknown words for three reasons: 1) as one type of introspective verbal

report, stimulated recall has been used to examine learners' awareness in second language acquisition (Gass & Mackey, 2017); 2) in natural viewing settings, the real-time nature of watching audio-visual material, where new information is continuously provided, does not allow participants to pause and verbalise their thoughts during viewing; 3) think-aloud protocols are not suitable for research using eye-tracking during viewing as verbalising one's thoughts would distort participants' eye movements (Godfroid & Schmidtke, 2013).

2.1.3. Design

Participants would be asked to watch the captioned video while their eye movements are recorded. Immediately after their viewing, participants would be asked to watch the parts of the video containing the TWs to recall their awareness of each TW during viewing one by one and answer the question: "Were you aware of this word when watching this part of the video?". Importantly, at this stage, it would be preferable not to show participants' own eye movement recordings as prompts in order to prevent participants' from reporting their awareness according to the eye movements they see in the recording, rather than recalling their awareness at the time of the initial viewing (see also Godfroid & Schmidtke, 2013). For each TW presented in the video, we would have a measure of attention, reflected in the recordings of eye movements, and a measure of awareness in stimulated recalls.

To analyse the eye-tracking data, dynamic interest areas covering the presentation time of each TW should be first created (for a more detailed explanation, see Wang & Pellicer-Sánchez, 2022b). Then, researchers should select and export eye-tracking measures for all the unknown TWs. The commonly reported measures focusing on lexical items in viewing include: first-pass reading time, second-pass reading time, and total reading time (e.g., Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022b), as they represent participants' early and late eye movements, covering a more complete picture of learners' cognitive processes (Godfroid,

2020). To analyse the stimulated recall data, qualitative data analysis software, such as NVivo could be used for coding. Participants' awareness could be coded inductively using content analysis to generate codes based on participants' answers according to the coding steps suggested by Bryman (2012) and Selvi (2020): 1) concepts are generated by coding data at the level of open coding; 2) categories are generated through a constant comparison of concepts, micro-categories are grouped into more general categories; 3) saturated categories are listed; 4) categories are applied back to the stimulated recall data pertaining each word. A second coder with experience of coding stimulated recall data should also code parts of the data to examine the inter-coder reliability of the coding.

In the coding of the stimulated recall data collected by Wang (2022), following the above data analysis procedure, three awareness categories emerged: forgot (i.e., participants reported that they forgot/did not remember/were uncertain about whether they were aware of the TW while viewing), no reported awareness (i.e., participants reported that they were not aware of the TW during viewing), and reported awareness (i.e., participants reported that they were aware of the TW, either visually, aurally, or both). Then, as shown in Table 1, each of the unknown TWs could be assigned into one of the awareness categories and presented in a table alongside its corresponding eye-movement data. For example, Table 1 shows that participant A16 indicated awareness of the word *sedated* in the stimulated recall and spent 621 milliseconds (ms) processing it while viewing. However, 0 ms were reported for the word *waddled,* for which participant A16 reported lack of awareness.

Table 1

Examples of Participant A16's Reported Awareness and Eye-Tracking Data for Each Unknown Target Word

Participant	TW	Awareness	Total reading time (ms)	1st-Pass reading time (ms)	2nd-Pass reading time (ms)
A16	confiscated	+	1067	156	911
A16	barneys	+	0	0	0
A16	sedated	+	621	621	0
A16	gland	–	0	0	0
A16	ulcers	–	323	323	0
A16	foal	+	150	150	0
A16	dinky	–	0	0	0
A16	nuzzle	+	357	357	0
A16	waddled	–	0	0	0
A16	foraging	–	426	145	281
A16	midwife	+	37	37	0
A16	surrogate	+	1031	240	419

Note. + indicates reported awareness, – indicates reported no awareness

For an initial analysis, descriptive data could be used to summarise and compare the eye-tracking data across the three awareness categories to examine if the TWs with reported awareness involve a different amount of attention than those without reported awareness. For example, participant A16 had an average of 466 ms for the words for which she reported awareness whereas an average of 150 ms was reported for those without reported awareness. For a more robust analysis of the relationship between attention and awareness, mixed-effects models could be constructed on item-level with eye-tracking data as the outcome variable and awareness categories as the predictor variable. Previous eye-tracking research has found that longer words (e.g., Montero Perez et al., 2015; Puimège & Peters, 2019; Wang & Pellicer-

Sánchez, 2022b) and words with higher frequency of occurrence (e.g., Montero Perez et al., 2015; Wang & Pellicer-Sánchez, 2022b) lead to longer processing times, resulting in higher learning gains. Besides, nouns seem to be more likely to be learned from reading than verbs and adjectives (e.g., Godfroid et al., 2018), implying different processes with items of different parts of speech. Thus, in order to ascertain that differences in reading times are due to the different levels of awareness, these potentially confounding factors (e.g., word length, frequency of occurrence, part of speech) would need to be controlled for in the analyses. This analysis would allow us to examine if different awareness categories are statistically associated with the amount of attention paid to an unknown word (see also Godfroid & Schmidtke, 2013).

The addition of qualitative data would also enable us to further explore the reasons behind the mismatch between participants' recorded attention and reported awareness. For example, as shown in Table 1, despite the fact that no fixations were recorded on the TW *barneys*, data from Wang (2022) showed that participant A16 still reported her awareness of that word in the stimulated recall: "I noticed this word [*barneys*], I heard it, but I didn't know its meaning". As this example shows, the addition of verbal reports in studies on learning from viewing would allow researchers to explore attention to the auditory input, addressing the limitation of eye-tracking.

2.2. Example 2

2.2.1. Research Question

What is the relationship between participants' reported vocabulary processing strategies and the amount of attention paid to unknown vocabulary?

2.2.2. Rationale and Aim

Previous research has revealed that L2 learners make use of various strategies to process unknown words in L2 reading (e.g., Rott, 2000) and viewing (e.g., Sydorenko, 2010), but the reported strategies might not always lead to successful guessing or contribute to learning. To the best of our knowledge, only the study conducted by Sydorenko (2010) has explored L2 learners' vocabulary processing strategies in captioned viewing. However, this research only investigated the overall strategies used by participants during viewing using an open-ended questionnaire rather than focusing on the strategies for each unknown word. It is possible that the general strategies did not represent what happened with each individual word. Importantly, the various strategies that learners employ to engage with unknown vocabulary could be reflected in differences in the amount of attention to the unknown items. Since attention is a crucial component for vocabulary learning, the examination of the connection between types of strategies and amounts of attention would allow us to investigate if certain types of processing strategies lead to increased attention to novel vocabulary and potentially to higher learning gains. Thus, the aim of Example 2 is to explore the relationship between amount of attention and underlying processing strategies.

2.2.3. Design

Similar to Example 1, eye-tracking would be used to record participants' attention to each TW during viewing. Immediately after their viewing, stimulated recalls should be administered individually. An important methodological decision in this design is whether participants' eye movements recordings should be shown as prompts in the stimulated recalls. In Example 1, we suggested not using eye movement recordings when the aim was to examine awareness, as it has the drawback of potentially induce participants to report awareness by describing their eye movements. However, Example 2 uses stimulated recalls to elicit participants' thoughts when processing the unknown words. Thus, using participants' eye movements as prompts could

potentially facilitate their memory recall (see also Jung & Lee, 2022). The recorded stimuli could be played at a 50% speed to avoid participants' difficulty in following their rapid eye movements (Wang, 2022). In the pilot study conducted by Wang (2022), participants reported that the slowed down play of their eye movements aided their recall of thoughts.

Participants would be asked to watch the recorded stimuli containing each TW with a pause in between target items. After watching each stimulus, participants would answer the question: "What were you thinking at that time when encountering this word during viewing?" and would be asked to report the thoughts that they had during viewing rather than their thoughts at the time of the stimulated recall. Following Gass and Mackey (2017), no concrete responses should be given to the participants' answers, except repeating their responses, or providing "back-channelling cues or nonresponses" such as "Oh, mhm, great, good, I see, uh-huh, ok" (p. 55). If there is an indication that the participant was talking about his/her current thoughts of the unknown word, the researcher could bring the participant back on track by asking: "Is this what you were thinking at that time during viewing or your current thoughts?". No further questions should be asked if participants are unable to recall their thoughts relating to the words. We recommend to hold all stimulated recall interviews in participants' L1, when the L1 is shared, to avoid language constrains (see also Gass & Mackey, 2017).

To analyse the stimulated recall data, since no previous research has provided a framework to categorise learners' vocabulary processing strategies in viewing, we recommend to code the data following an inductive approach and adopting a qualitative content analysis method (Selvi, 2020), as suggested in Example 1. Therefore, themes and categories would be generated in a data-driven approach. In the study by Wang (2022), data for each TW was coded following an inductive approach, and general and specific strategies were identified. As Table 2 shows, for the captions group, 17 specific strategies emerged from the data and were grouped

into five general strategies. For each unknown TW, all reported processing strategies linked to the word should be coded.

Table 2

Coding Categories for Participants' Reported Vocabulary Processing Strategies (Wang, 2022)

General strategies	Specific strategies
1. Word features analysis	1.1. Analysing part of speech
	1.2. Analysing word-structure
	1.3. Form association
	1.4. Analysing word pronunciation
	1.5 Word usage
2. Using context	2.1. Using auditory cues
	2.2. Using images
	2.3. Using global understanding
	2.4. Using local contextual cues
3. Guessing without reported strategies	3.1. Meaning fully guessed
	3.2. Meaning partially guessed
	3.3. Meaning unsuccessfully guessed
4. Other strategies	4.1. Pre-test impact
	4.2. Dictionary use
	4.3. Visualizing
5. Reported awareness but without reported strategies	5.1. No reported meaning guessed
	5.2. Forgot thoughts

An initial and simple approach to analyse the data would involve the comparison of participants' eye movements across different strategies by calculating the average for the eye-

movement measures for each strategy. The descriptive statistics would allow us to find out if using certain strategies would lead to relatively longer processing times. A second and more robust option is to analyse the data at item level. For each participant, we would record the strategies used alongside the eye-tracking data for each TW, as shown in Table 3. Instead of calculating average data per strategy, we would code the data at the item level and construct mixed-effects models with the eye-tracking measures as dependent variables and the strategy as the independent variable. This would allow us to examine whether certain types of strategies predict the amount of attention paid to novel vocabulary while controlling for the item- and participant-level differences. To combat the issue of having large number of strategies as independent variable, researchers can consider using general strategies (as those presented in Table 2) to analyse the data; or using Lasso (least absolute shrinkage and selection operator), which performs variable selection and regularization, to reduce the types of strategies by combining strategies which do not substantially differ in terms of participants' eye movements (Gertheiss & Tutz, 2010). With non-normally distributed data, researchers can consider using bootstrapping for analysis (Plonsky et al., 2015).

Table 3

Examples of Participants' Reported Vocabulary Processing Strategies and Eye-Tracking Data for Each Unknown Target Word

Participant	TW	Strategy categories	Total reading time (ms)	1st-Pass reading time (ms)	2nd-Pass reading time (ms)
A11	barneys	2.2	873	867	6
A11	sedated	3.3	1083	300	365
A11	dinky	3.1	383	383	0
A11	foal	2.2; 2.4	519	339	180
A11	traumatised	1.1	262	190	72

A11	ulcers	2.4	326	326	0
A11	endearing	2.3	926	104	408
A11	foraging	5.1	921	92	829
A11	purring	2.1	301	254	47
A11	surrogate	2.2	257	257	0
A13	barneys	5.1	544	544	0
A13	sedated	2.4	481	100	381
A13	foal	2.4	0	0	0
A13	hump	2.2; 2.3	0	0	0
A13	surrogate	2.2; 2.3; 2.4	216	216	0

Finally, in addition to converting qualitative data into numbers and analyse them quantitatively, the final option involves adopting "qualitizing techniques" (p. 126) to analyse the eye-tracking data qualitatively and create different profiles for participants (Tashakkori & Teddlie, 1998). This approach has also been recommended by Révész et al. (2021) to better capture the details of L2 learning process by examining individual-level data. For example, as shown in Table 4, based on participants' frequency of strategy use, we used the quantile function in R and categorised participants as *active strategy users* (used strategies for 50% – 100% unknown TWs), *moderate strategy users* (25% – 49% unknown TWs), or *minimal strategy users* (0% – 24% unknown TWs). We could then explore what participants thoughts were, and how their processing strategies might potentially vary across different types of strategy users when they spent relatively long processing time on the unknown words. In this example, we could set 700 ms as the threshold for relatively long processing time, as L2 readers' total reading time on the first occurrence of an unknown word has been reported as around 700 ms in reading studies (e.g., Mohamed, 2018; Pellicer-Sánchez, 2016). Table 4 presents one

example from each type of strategy user. For each participant, it presents the unknown TWs that had a total reading time longer than 700 ms during the viewing activity, alongside the corresponding stimulated recall comments and the coded strategies.

Table 4

Examples of Participants' Stimulated Recall Comments and Their Corresponding Eye-Tracking Data for the Unknown Target Words During Viewing With Total Reading Time Longer Than 700 ms (Wang, 2022)

Active strategy user - A15					
TW	**Coding of strategies**	**Stimulated recall comments**	**Total reading time (ms)**	**1st-Pass reading time (ms)**	**2nd-Pass reading time (ms)**
surrogate	Using images; Using global understanding; Using local contextual cues	I have guessed this, because I thought this scene is too funny. What kind of duck could grow this fast?! It is bigger than that [the cat], and then I thought *surrogate* mum could mean something, something, such as replacement.	1079	387	417
sedated	No reported strategies	I didn't know what it means, so I would pay attention to the words I don't know.	926	926	0
traumatised	Analysing word-structure;	Actually, these three, I was unfamiliar with them, but because the middle one [*traumatised*] is the most unfamiliar one. Because the spelling of it is very	813	545	268

	Analysing word pronunciation	complicated, and it's hard to pronounce, *tra*, and then there's a *u* there, so I didn't know how to pronounce it. I would pay most attention to the most complicated and unfamiliar one.			
barneys	Dictionary use	I didn't know it's meaning. I wanted to know what it means, and I even wanted to memorize it, so that I could look it up later. But then I found that's impossible, because I would forget the previous ones when continue.	745	367	378
foraging	No awareness	I have no impression at all now, this word, I don't even have impression of this sentence.	719	719	0

Moderate strategy user - A20

TW	**Coding of strategies**	**Stimulated recall comments**	**Total reading time (ms)**	**1st-Pass reading time (ms)**	**2nd-Pass reading time (ms)**
nuzzle	No reported strategies	This was an unknown word, so it was just passed without much attention.	1142	1142	0

confiscated	No reported strategies	I really didn't know this word. I paused for a moment when I saw this word, and then I probably just looked away.	1073	203	97
endearing	Analysing part of speech; Analysing word-structure	Well, I guessed it was a positive adjective, to describe something good. And there is a *dear* in it, I guessed this might be a word describing the closeness of relationships, or the feeling of getting along well with others.	789	154	195
barneys	Using images	At that time, I guessed it should be… because the image showed that house, I thought it should be the name of the residence, or that kind of building, um.	712	712	0

Minimal strategy user - A23

TW	**Coding of strategies**	**Stimulated recall comments**	**Total reading time (ms)**	**1st-Pass reading time (ms)**	**2nd-Pass reading time (ms)**
sedated	No awareness	I didn't pay attention. I remember that I paid more attention to the images at that time.	952	262	253
bizarre	No awareness	Hmm, I don't know, I don't remember seeing this at that time, ah, I didn't pay much attention to this word.	906	201	359

barneys	No reported strategies	I have no idea. I might have seen it at that time, but I didn't quite know it's meaning.	872	223	649
twirls	No awareness	This one, I didn't pay attention.	739	213	526

As can be observed in Table 4, for those novel words with relatively long processing time, the active strategy user A15 used images, textual context, or word features to engage with three of them, and combined different strategies to guess the meaning of some TWs. A15 also demonstrated willingness to learn the TW *barneys* by trying to memorise and look it up in the dictionary after the experiment. In the two remaining cases, A15 reported no strategy use or no awareness of the unknown words. The moderate strategy user A20 used word features or images to engage with two unknown words but did not report any strategies for the other two unknown words. Unlike A15 and A20, although the minimal strategy user A23 spent relatively longer processing time on four TWs, the unknown words tended to be processed unconsciously as A23 reported either no awareness or no strategy used for those words.

Therefore, for the active and moderate users, the longer processing time on the unknown words seems to either reflect learners' active engagement with the word or processing difficulty without deep engagement. However, for the minimal strategy users, the longer processing time on a novel item was more likely to indicate processing difficulty or no awareness. Analysing data at individual-level, instead of relying on group-level summaries, would allow us to explore potential individual differences, which might also help to explain the relationship between attention and vocabulary learning gains.

3. Methodological challenges of the combination of eye-tracking and stimulated recalls

In the previous section, we have seen how eye-tracking and stimulated recalls might be combined to reach a better and more comprehensive understanding of the various processes involved in vocabulary learning, highlighting the advantages of such approach. In this section, we consider some of the main challenges that may arise.

The main challenges of combing eye-tracking and stimulated recalls lie in the research design and data analysis. It is still arguable whether participants' eye movements should be

used as prompts in stimulated recalls, as some researchers included eye movements for a better memory recall (e.g., Jung & Lee, 2022), while others have argued that eye movements might be an overly strong stimulus which may distract participants' memory thereby obscuring recall (Gass & Mackey, 2017). As explained in the two examples above, the decision to use eye-movement recordings as prompts in stimulated recalls will ultimately depend on the research purpose of a particular study, but more research is needed to further investigate this issue and understand the consequences of their use. In addition, when interpreting stimulated recall data, researchers should take veridicality into account, as memory decay could potentially affect the accuracy of participants' recall, and stimulated recalls should be conducted as close to the activity as possible to retain the memory (Gass & Mackey, 2017). Moreover, in large-scale research including post-tests to assess learners' vocabulary knowledge, the reactivity issue of using post-tests between the activity and stimulated recall should also be considered when interpreting the stimulated recall findings (Bowles, 2018). Lastly, in terms of data analysis, researchers should not only be skilful at analysing both quantitative and qualitative data but should also be capable of finding the best way to integrate different types of data to answer research questions. It might also be challenging to interpret conflicting results when the data collected from these two methods contradicts to each other (Johnson & Christensen, 2012). While these limitations and challenges need to be taken into account, the better and more comprehensive understanding of the vocabulary learning process that the combination of eye-tracking and stimulated recalls can bring merits its further exploration.

4. Conclusion

The use of eye-tracking as a measure of attention has brought about an important turn in vocabulary learning research, allowing researchers to gain a clearer understanding of the underlying cognitive processes and of the relationship between attention and word learning.

The research accumulated thus far has also been instrumental in showing the limitations of eye-tracking in this context. Researchers have argued for the combination of eye-tracking with verbal reports to address these limitations, but research adopting such approach is still scarce. The suggested studies and analyses have illustrated how this combination can shed light into the subprocesses encoded in eye-movement data, contributing to our understanding of the relationship between attention, awareness, and word learning, and of how the use of different processing strategies might modulate this relationship. Future research could also explore L2 learners' emotional processes via verbal reports to inspect their potential relationship with eye movements as well as with learning gains (Deng & Gao, 2022). The triangulation of various data sources, along the lines of what is suggested in this paper, will help to move the field forward.

References

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

Bowles, M. A. (2018). Introspective verbal reports: Think-Alouds and stimulated recall. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 339–357). Palgrave Macmillan UK.

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford University Press.

Deng, R., Gao, Y. (2022). A review of eye tracking research on video-based learning. *Education and Information Technologies*. *28*, 7671–7702 https://doi.org/10.1007/s10639-022-11486-7

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition, 40*(2), 341–366. https://doi.org/10.1017/S0272263117000109

Ellis, N. C. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review*. *11*, 37–56.

Ericsson, K. A. (2002). Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology, 16*(8), 981–987. https://doi.org/10.1002/acp.925

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21*(2), 225–241. http://doi.org/10.1017/S0272263199002041

Gass, S. M., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research* (2nd ed.). Routledge.

Gertheiss, J. & Tutz, G. (2010). Sparse modeling of categorial explanatory variables. *The Annals of Applied Statistics*. *4*(4), 2150–2180. http://doi.org/10.1214/10-AOAS355

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge.

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*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(03), 483–517. https://doi.org/10.1017/s0272263113000119

Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 183–205). University of Hawaii, National Foreign Language Resource Center.

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2021). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, *0*(0), 1–30. https://doi.org/10.1177/13621688211001289

Hu, H. M., & Nassaji, H. (2012). Ease of inferencing, learner inferential strategies, and their relationship with the retention of word meanings inferred from context. *Canadian modern language review*, *68*(1), 54–77. http://doi.org/10.3138/cmlr.68.1.054

Johnson, B., & Christensen, L. B. (2012). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). SAGE.

Jung & Révész (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*(4), 755–780. http://doi.org/10.1017/S0272263118000165

Jung, J., & Lee, M. (2022). Second language reading and recall processes under different reading purposes: An eye-tracking, keystroke-logging, and stimulated recall study. *Language Awareness*, 1–23. https://doi.org/10.1080/09658416.2022.2069251

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge.

Mohamed, A. A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, *40*(2), 269–293. https://doi.org/10.1017/S0272263117000092

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*(2), 308–328. https://doi.org/10.1111/modl.12215

Ouyang, J., Huang, L., & Jiang, J. (2020). The effects of glossing on incidental vocabulary learning during second language reading: Based on an eye-tracking study. *Journal of Research in Reading*, *43*(4), 496–515. https://doi.org/10.1111/1467-9817.12326

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking Study. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2020). Expanding English vocabulary knowledge through reading: Insights from eye-tracking studies. *RELC Journal*, *51*(1), 134–146. https://doi.org/10.1177/0033688220906904

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning, 71*(1), 162–203. https://doi.org/10.1111/lang.12430

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics. 43*(3), 727–756. https://doi.org/10.1017/S014271642200011X

Plonsky, L., Egbert, J., & Laflair, G. T. (2015). Bootstrapping in applied linguistics: Assessing its potential using shared data. *Applied Linguistics*, *36*(5), 591–610. https://doi.org/10.1093/applin/amu001

Puimège, E., Montero Perez, M., & Peters, E. (2023). The effects of typographic enhancement on L2 collocation processing and learning from reading: An eye-tracking study. *Applied Linguistics*, First view: https://doi.org/10.1093/applin/amad003

Puimège, E., & Peters, E. (2019). Learning L2 vocabulary from audiovisual input: An exploratory study into incidental learning of single words and formulaic sequences. *The Language Learning Journal, 47*(4), 424–438. http://doi.org/10.1080/09571736.2019.1638630

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Révész, A., Lu, X., & Pellicer-Sánchez, A. (2021). Directions for future methodologies to capture the processing dimension of L2 writing and written corrective feedback. In R. M. Manchón & C. Polio (Eds.), *The Routledge handbook of second language acquisition and writing* (pp. 339–355). Routledge.

Rott, S. (2000). Relationships between the process of reading, word inferencing, and incidental word acquisition, in assigning meaning to form. In J. E. Lee & A. Valdman (Eds.), *Form and meaning: Multiple perspectives* (pp. 255–282). Heinle & Heinle Publishers.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Selvi, A. F. (2020). Qualitative content analysis. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 440–452). Routledge.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. SAGE.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 183–203. https://doi.org/10.1017/S0272263100012870

Wang, A. (2022). Investigating the effectiveness of bilingual subtitles for incidental vocabulary learning: A mixed methods study [Unpublished doctoral dissertation]. University College London.

Wang, A., & Pellicer-Sánchez, A. (2022a). Examining the effectiveness of bilingual subtitles for comprehension: An eye-tracking study. *Studies in Second Language Acquisition*, 1–24. https://doi.org/10.1017/S0272263122000493

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