10

SALIENCE, COGNITIVE EFFORT, AND WORD LEARNING

Insights from Pupillometry

Kelli Ryan, Phillip Hamrick, Ryan T. Miller, and Christopher A. Was

Introduction

Understanding the roles of general cognitive processes in language learning is one of the central goals of research in second language acquisition (SLA). Of these cognitive processes, the role of attention in second language (L2) learning are central to numerous theoretical frameworks (e.g., Ellis, 2006; Gass, 1997; Leow, 2015; Schmidt, 1990; Tomlin & Villa, 1994). As attention plays a critical role in these frameworks, factors that constrain and direct learners' attention should, therefore, also play critical roles. One such factor is salience (e.g., Ellis, 2006; Leow, Egi, Nuevo, & Tsai, 2003). In a given learning context, allocation of attention—and therefore the amount of attention-driven learning—should be at least partly determined by the salience of a given form in the input. On this assumption, researchers have employed a variety of methods aimed at promoting attentional processing, most of which involve making linguistic stimuli more salient. Perhaps the most widely employed method of increasing salience involves some kind of input enhancement (for overviews, see Han, Park, & Combs, 2008; Lee & Huang, 2008; Leow & Martin, Chapter 9).

Input enhancement is typically conceived of as a way for researchers or educators to direct learners' attention to specific information in the input. Sharwood Smith (1991, 1993) argued that this input enhancement makes target information more perceptually salient and more likely to be noticed (see also Gass, 1988; Schmidt, 1990). His theory builds on the assumption that noticing is necessary for learning, and that internally- and externally-driven salience govern the learner's ability to notice particular types of information in the input. To promote learners' noticing in visual modalities, which is our present focus, researchers and educators can manipulate the visual perceptual salience of text-based linguistic features by manipulating the text through techniques like color coding, highlighting, and boldfacing, henceforth referred to as textual enhancement (TE).

The efficacy of externally-generated salience-enhancing techniques, such as TE, has long been the subject of empirical investigation. However, research on the effectiveness of TE has yielded largely mixed results. For example, a meta-analysis conducted by Lee and Huang (2008) found that TE may be beneficial for form learning, but not for meaning comprehension. Specifically, Lee and Huang (2008) confirmed a small positive effect for TE on form learning across 16 studies encompassing 20 samples. In these studies, L2 learners who were exposed to various TE manipulations outperformed their unenhanced counterparts learning the same target forms with a small positive effect size (d = 0.22), while a small, negative effect size was obtained for learners' meaning comprehension (d = –0.26).

In an attempt to shed more light on these complex findings, researchers have resorted to using a variety of techniques to look at the interplay among TE, attention, and learning. One such technique is eye-tracking (Godfroid, Boers, & Housen, 2013; Godfroid & Winke, 2015). For example, Winke (2013) conducted a modified replication of Lee (2007) using eye-tracking technology. Lee's (2007) original study reported that while TE supported the learning of target grammatical forms (the passive voice), it did not improve comprehension of meaning. In fact, Lee (2007) reported a negative effect of TE on a comprehension recall task. In Winke's (2013) modified replication, she also investigated learning of the passive voice in university-level ESL students, via training consisting of form and meaning comprehension. These results indicated that TE did not significantly increase form learning, and it had no negative comprehension effects. Additionally, eye-tracking data suggested that participants exposed to input with TE noticed the passive to a greater degree, as measured by gaze time and rereading. However, this did not itself lead to better learning in the absence of explicit instruction. Thus, Winke's (2013) findings are consistent with Sharwood Smith's (1991) proposal that TE may promote noticing; however, in this case, noticing alone did not result in learning.

Even with eye-tracking studies, though, the reported effects of TE on attention, noticing, and learning have been mixed. For example, contrary to Winke (2013), Loewen and Inceoglu's (2016) eye-tracking study of L2 Spanish learners provided no evidence that enhancement changed learners' attentional processing, as measured by fixation time and self-reported awareness. Additionally, while both enhanced and unenhanced groups improved their knowledge of the Spanish past tense, there was no significant difference between groups in overall performance. Thus, there was no clear link between enhancement and attentional processing or learning gains. Similarly, an eye-tracking study by Indrarathne and Kormos (2016) reported that increased attentional processing was not necessarily a function of TE. Indeed, between enhanced (in the form of bold text) and unenhanced groups, there was no significant difference in learning on production and comprehension testing of the target grammatical forms.

Taken together, the available evidence is mixed with respect to whether TE influences subsequent noticing, attentional processing, or learning. It is possible that other processes beyond noticing and attention are responsible for the varied learning outcomes across studies. Certainly, the degree of elaboration (number of instances) or explicitness of metalinguistic information (Sharwood Smith, 1991) may moderate the efficacy of TE. Similarly, one might also expect that cognitive effort, or the related construct of depth of processing (Leow, 2015; Leow & Mercer, 2015), could moderate the effectiveness of TE.

Depth of Processing and Cognitive Effort

Although attention is important to their account, Leow's depth of processing approach (Leow, 2015; Leow & Mercer, 2015) emphasizes stages in information processing beyond detection (Tomlin & Villa, 1994). Indeed, critical for the present study, Leow and Mercer (2015, p. 2) argue that deeper processing is, among other things, employing "greater cognitive effort during processing while using prior knowledge to strengthen the process." In relation to TE, lower depth of processing may explain some results where enhanced groups do not outperform unenhanced (such as Winke, 2013). For example, when L2 learners see novel enhanced input, they may spend more time wondering why the text is boldface than they do learning or comprehending the text or form (e.g., Bowles, 2003; Leow, 2001). Thus, depth of processing may interact with the salience of information in the L2 input, and there is some evidence for this. For example, in a study on the effects of TE, Leow et al. (2003) reported no significant benefit for learning Spanish past tense forms using enhanced text over unenhanced text at posttest. Despite no apparent benefits of TE on the immediate recognition and comprehension posttest, Leow et al. (2003) did find that the salience of forms, regardless of TE, seemed to be related to the amount of deeper processing reported.

Thus, there may be some interaction between depth of processing or amount of cognitive effort and enhancement, and it is possible that these factors play a role in learning. Indeed, several researchers argue that attentional processing needs to be accompanied by some (at a minimum) low-level cognitive effort (e.g., Gass, 1997; Hulstijn & Laufer, 2001; Leow, 2015; Truscott & Sharwood Smith, 2011; VanPatten, 2004). However, more research is needed to examine these claims, and, of course, establishing whether and how much cognitive effort learners expend while processing the L2 input can be difficult to establish. Researchers have employed a variety of methods to do so, including concurrent data elicitation from think-aloud protocols (e.g., Leow et al., 2003) and pupillometry.

Pupillometry

While eye-tracking data, such as fixations and regressions, have increasingly been used to study the attentional processes involved in L2 learning, other eye-related data can be used to shed light on different cognitive processes beyond attention.[1] Pupillometry, which involves measuring changes in pupil dilation in response to different stimuli, has been extensively used in cognitive psychology as an index of cognitive effort and processing load (for an overview, see Sirois & Brisson, 2014) but has not, to our knowledge, seen much use in SLA (although, see Schmidtke, 2014).

The pupil is the open region of the iris which allows light to reach the retina, and the muscles controlling the pupil are sensitive to a wide variety of factors of interest to psychologists and SLA researchers, including cognitive effort (e.g., Kahneman & Beatty, 1966) and amount of mental activity (e.g., Wierda, van Rijn, Taatgen, & Martens, 2012). Kahneman (1973) described the validity of pupillometric measures of cognitive task demands and the use of pupillometry to capture variability in a person's effort both during a task and across tasks of varying challenge. Because increases in dilation are associated with higher levels of mental effort, dilation is often used to measure cognitive effort or load. For example, during an attentional blink task, which requires attention across a series of distractors, dilation increases significantly as the time between stimuli and number of distractors increases (Wierda et al., 2012). Cabestrero, Crespo, and Quirós (2009) used pupillometry as in index of mental effort and allocation of cognitive resources under several load conditions. Using a voice-specificity paradigm often used in recognition memory experiments, Papesh, Goldinger, and Hout (2012) found pupil dilation was greater when items were subsequently recognized with more confidence, relative to those recognized with less confidence or not recognized. The items successfully recognized with greater confidence were items which participants apparently expended greater cognitive effort during encoding.

Thus, because pupillary responses are so tightly linked to cognitive effort, pupillometry is ideally suited for assessing the role of cognitive effort in attention-related language learning phenomena.

The Present Study

The interactions among TE, attention, and deeper processing make for a complex picture that is in need of more investigation and clarification. Pupillometry is poised to assist in addressing this need. To that end, the aim of the present study was to begin investigating whether the efficacy of one type of input enhancement (textual enhancement) interacts with one learner-internal factor (cognitive effort), indexed by pupil dilation. We did so by focusing on the domain of incidental word form learning. In particular, this study aimed to address several related research questions:

RQ1: Does textual enhancement lead to better learning of and memory for novel word forms?
RQ2: Does cognitive effort, indexed by changes in pupil dilation, predict learning of and memory for novel word forms?
RQ3: Is there an interaction between the effects of textual enhancement and cognitive effort in predicting learning of and memory for novel word forms?

Given that the effects of input enhancement have yielded mixed results, we made no specific hypotheses regarding the first research question. Regarding the second research question, and following suggestions that deeper processing (e.g., Leow, 2001, 2015; Sharwood Smith, 1991) may be critical to the effectiveness of input enhancement and that involvement effort may be critical for incidental word learning (e.g., Hulstijn & Laufer, 2001), we hypothesized that greater cognitive effort during the training phase—as indexed by greater pupil dilation—would predict memory for the

TABLE 10.1 Biodata for Participants by Group

Group	*Gender*	*Mean Age*	*Mean L2 Experience in Years (total)*
+TE	10 females, 4 males	20.71 (19–23)	3.67 (2–5)
–TE	10 females, 2 males	20.58 (18–27)	3.25 (2–4)

Note: Ranges in parentheses.

new words at test. Regarding the third research question, which follows from using a factorial design with an interaction, we predicted that textual enhancement would interact with cognitive effort in predicting learning of and memory for novel words. We predicted an interaction on the expectation that the salience of a form (i.e., whether or not it is enhanced) should influence the probability that it gets processed further, and that word learning may depend on whether enhancement has this effect.

Methods

Participants

Thirty-six volunteer undergraduates participated in the experiment in exchange for extra credit in their introductory linguistics courses. Participants were randomly assigned to a group that received input with textual enhancement (+TE) or without textual enhancement (–TE). Data from 10 participants were unable to be used because of loss of pupil information (n = 8), for having uncorrected medical conditions involving the eye (n = 1), and for not being a native speaker of English (n = 1). This left 14 participants in the +TE group and 12 participants in the –TE group. All of the remaining participants reported having normal or corrected-to-normal vision and hearing.

Materials

Training Phase

The experimental task was programmed and displayed with E-Prime 2.0 ® (Schneider, Eschman, & Zuccolotto, 2012). Participants were visually presented with 30 English sentences (Table 10.2). Each sentence was presented twice, once for a control condition in which the sentence-final word was a normal English word and once for an experimental condition in which the sentence-final word was a target pseudoword. Presenting each sentence this way allowed us to control several factors, including angle of the target to the eye. It has been shown that measurement of pupil diameter can be influenced by gaze position as a participant's eyes move across the screen, even when the pupil itself remains the same size (Hayes & Petrov, 2016).[2] Presenting the same sentences twice, once in a control condition and once in a target condition, allowed us to rule out stimulus-eye angle as a confound between

TABLE 10.2 Sample Sentences from the Training Phase

Sentence-final control word	*Sentence-final target pseudoword*
She could tell he was mad by the tone of his voice.	She could tell he was mad by the tone of his tolve.
After every meal it's good to brush your teeth.	After every meal it's good to brush your thrig.
His boss refused to give him a raise.	His boss refused to give him a septh.

conditions. The English sentences were taken from a set of sentences with reported cloze probability and completion norms made publicly available (Block & Baldwin, 2010). In order to keep participants' focus on meaning, we chose sentences with sentence-final words that had high cloze probabilities (M = 0.87, SD = 0.08) in order to make it easier to induce meaning from context even in the presence of a sentence-final pseudoword (e.g., *The bride smiled as she walked down the bealm*).

Pseudowords were produced via search queries to the English Lexicon Project website database (Balota et al., 2007). The search was aimed at retrieving pseudowords that were similar to control words in terms of their lexical characteristics. Control words were each monosyllabic, five letters long, with mean bigram (i.e., two-letter combinations) frequencies of 1400 (SD = 488). Therefore, the following search criteria were used: word length = 5 letters, syllables = 1, average bigram frequency = 1,000 to 3,000.[3]

To control for memorability, all possible pseudowords were normed using a recognition memory task. A sample of 13 students who did not participate in the main experiment was given one of two lists of 30 pseudowords (list A or list B). In the study phase, participants were instructed to remember as many pseudowords as they could. They were then given a recognition task that contained all the words from both lists in random order. Memory performance for list A (M = 0.66, SD = 0.20) was not significantly different from memory performance for list B (M = 0.63, SD = 0.18), $t(58) = 0.59$, $p = .55$. Since participants appeared to have no pre-existing bias making one list more memorable than the other, we assigned list A to be the target pseudowords in the training phase and list B to be the foil (new) words for the recognition memory test for the main experiment. In the final version of the stimuli, both control words and pseudowords all contained five letters and one (apparent) syllable. The difference in mean bigram frequency between control words (M = 1400, SD = 488) and pseudowords in the training phase (M = 1285, SD = 280) was not statistically significant, Welch's $t(46.31) = 1.11$, $p = .27$, $d = 0.29$, 95% CI [-92, 321].

In the training phase, all participants in both the +TE and −TE groups read the same exact sentences the same number of times, with the order of presentation randomized for each participant. The only difference in the stimuli between groups was the presence of textual enhancement, operationalized as yellow highlighting. Highlighting was achieved using the Visual Basic RGB function in E-Prime for yellow (255 red, 255, green and 0 blue). For the +TE group, both sentence-final control words and pseudowords were highlighted (so that any luminance effects from highlighting on pupil dilation would be balanced across control and target stimuli). The –TE group saw the sentences without any highlighting. All pupillometry measures came from changes to pupil diameter measured during this training phase.

Recognition Memory Task

After the training phase, participants were given a surprise recognition memory test for the pseudowords from the training phase. They were instructed that they would see 60 words presented in the middle of the computer screen. They were informed, correctly, that half the pseudowords occurred in the sentences from the training phase and that these should be classified as "old" while the other half of the pseudowords would be new ones that they had not yet seen. These should be classified as "new." Participants were instructed to make their old/new decision as quickly and as accurately as they could by pressing the computer keys corresponding to "old" and "new." Reaction times and accuracy were recorded.

Apparatus

Pupil size and other eye-tracking data were recorded with an Applied Science Laboratories ASL 6 eye-tracker, sampling at 60 Hz from participants' left eye, and stimuli were presented in in black print on a white background in Arial font size 24. Participants were seated so that their eyes were approximately 65 cm from the eye-tracker and such that the participants' gaze angle was less than 42° to the screen as recommended by the manufacturer.

Procedure

Participants were seated in a quiet, dimly lit room. They placed their chins in a chinrest that minimized head movement. A nine-point calibration of eye movement was performed at the beginning of the experiment for each participant. Each training phase trial began with a 250 ms fixation cross placed at the left edge of the screen where the first word of each stimulus sentence would appear. Then a stimulus sentence appeared for 2 seconds followed by a blank screen ISI of 1 second. Participants were told that the goal of the study was to understand how people process sentences for meaning via the use of eye-tracking technology. They were told that their task was to read each sentence in order to understand its meaning just like they "would read a book, a news article, or a blog." Participants were not informed that they would be tested, nor were they instructed to try to learn anything. In other words, participants were exposed to the target pseudowords under incidental learning conditions.[4]

Results

All statistical analyses were performed in the statistics program R (R Core Team, 2015) and figures were produced with the package ggplot2 (Wickham, 2009).

Accuracy and Reaction Time on the Recognition Task

The recognition memory task was first analyzed to establish whether any learning took place. We computed *d′* scores for each participant (Wickens, 2001). The *d′* measure is commonly used to assess recognition memory because it can discriminate between "signal" (e.g., memory for a target word) and "noise" (e.g., other factors that lead participants to make one kind of response over another) that jointly contribute to recognition judgments. Because we were most interested in memory for previously seen target pseudowords, and because participants may be expected to perform differently on target pseudowords (old items) and foil pseudowords (new items), we analyzed hits (correct classification of old items as old) and correct rejections (correct classification of new items as new) separately. Table 10.3 reports the descriptive results. Visual inspection of the descriptive statistics indicated opposite patterns of accuracy for hits and correct rejections between the +TE and –TE groups. Therefore, we conducted a mixed ANOVA with Group (2 levels: +TE, –TE) as a between-subjects factor and Item Type (two levels: hits, correct rejections) as within-subjects factor. The results of the mixed ANOVA revealed no significant effects of Group, $F(1, 24) = 0.02$, $p = .87$, $\eta_p^2 = 0.001$, or Item Type, $F(1, 24) = 0.18$, $p = .67$, $\eta_p^2 = 0.01$, but there was a significant Group*Item Type interaction, $F(1, 24) = 6.35$, $p = .02$, $\eta_p^2 = 0.21$. However, Bonferroni-adjusted post-hoc *t*-tests (revised *p* value = 0.025) did not show significant between group differences in accuracy on hits, $t(24) = 1.71$, $p = .09$, or on correct rejections, $t(24) = 1.92$, $p = .06$.

We analyzed participants' reaction times (RT) in the same fashion as their accuracy, focusing on RT for hits and correct rejections. A mixed ANOVA with Group (2 levels: +TE, –TE) as a between-subjects factor and Item Type (two levels: hits, correct rejections) as a within-subjects factor revealed a significant effect of Item Type, $F(1, 24) = 5.17$, $p = .03$, $\eta_p^2 = 0.18$, Group, $F(1, 24) = 4.24$, $p = .05$, $\eta_p^2 = 0.15$, and no significant Group*Item Type interaction, $F(1, 24) = 0.11$, $p = .74$, $\eta_p^2 = .004$. Bonferroni-adjusted post-hoc *t*-tests (revised *p* value = 0.025) did not show significant between group differences in RT for hits, $t(24) = 2.21$, $p = .03$, or for correct rejections, $t(24) = 1.83$, $p = .07$.

Taken together, the results indicate that both groups were accurate to similar degrees on the recognition task, but their patterns of accuracy differed. On the other hand, the RT data appeared to show that the +TE group was generally faster than

TABLE 10.3 Descriptive Results for the Recognition Memory Task

	+Textual Enhancement	*–Textual Enhancement*
Hit Accuracy	0.61 (0.13, 0.03)	0.69 (0.07, 0.02)
Correct Rejection Accuracy	0.71 (0.12, 0.03)	0.62 (0.10, 0.03)
d′	0.89 (0.50, 0.13)	0.83 (0.39, 0.12)
Hit RT	953 (294, 78)	1219 (318, 91)
Correct Rejection RT	1031 (367, 98)	1323 (443, 127)

Note: Standard deviations and standard errors are reported in parentheses, respectively. Reaction times (RT) are reported in milliseconds.

the –TE group at classifying items in the recognition task, regardless of whether those items were old or new.

Cognitive Effort as a Predictor of Recognition Task Performance

To investigate our second and third research questions, we built a series of regression models testing whether pupil diameter during the training phase predicted performance on the recognition memory test. Due to the interaction between groups and accuracy on old and new items at test, we built separate regression models[5] for hits and correct rejections and separate models for accuracy and RT. Models were fit with an outcome variable being predicted by main effects of the categorical predictor Group (two levels: +TE, –TE) and the continuous predictor of pupil dilation (measured during the training phase). The –TE group served as the reference level for all analyses. The regression model for accuracy on hits showed that pupil dilation during the training phase did not predict accuracy on hits for the –TE group, but

TABLE 10.4 Regression Model Results for Accuracy for Hits in the Recognition Test

	Estimate	***SE***	***t-value***
Intercept	1.032	0.221	4.677**
+Textual Enhancement	–1.172	0.440	–2.663*
Pupil Dilation	–0.008	0.005	–1.556
+Textual Enhancement × Pupil Dilation	0.027	0.011	2.499*

Note: ***p* < .001. **p* < .05.

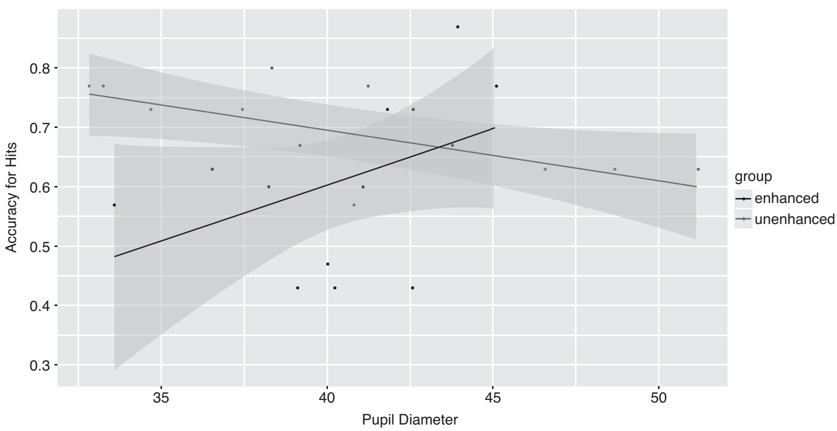

FIGURE 10.1 Scatterplot depicting relationships between recognition test accuracy for hits and mean pupil dilation in the training phase by group. Each dot represents a single participant. The gray region around the regression lines indicates the 95% confidence interval.

the significant interaction indicates that increased pupil dilation during the training phase predicted increased accuracy on hits for the +TE group (Table 10.4). Thus, cognitive effort predicted accuracy on hits for the +TE group, but not for the –TE group. This regression model was statistically significant, $F(3, 22) = 3.28$, $p = .04$, and it accounted for 21% of the variance in accuracy on hits. However, this model did not survive a Bonferroni correction for multiple comparisons ($p > .025$). Regression models for the other three outcome variables (accuracy for correct rejections, RT for hits, RT for correct rejections) were not statistically significant, all *p*s > .17.

Discussion

This study investigated three research questions. The first question asked whether textual enhancement lead to better learning of and memory for novel words. While the +TE group was faster than the –TE group at correctly classifying items in the recognition task, overall accuracy between the two groups on the recognition task was not significantly different. However, closer inspection of the accuracy data revealed that the –TE group outperformed the +TE group specifically on the recognition of novel words from the training phase (i.e., hits). Conversely, the +TE group was better than the –TE group at classifying new items that had not occurred in the training phase (i.e., correct rejections). What accounts for this pattern of results? One simple explanation is perceptual priming. Perceptual priming is a phenomenon whereby subsequent recognition or classification of a repeated stimulus is affected by its perceptual similarity to (or difference from) a previous stimulus. For example, presentation "dog" followed by "dog" is more likely to result in perceptual priming than a presentation of "dog" followed by "***dog***." On this view, the visual similarity between training and test phase items may have been higher for the –TE group because the target pseudowords were unenhanced in both the training and recognition test phases. In this scenario, if participants used perceptual similarity as a response heuristic then this would lead to a response bias favoring more "old" judgments. That is, there would be more hits (correct classification of old items as old) and more false alarms (incorrect classification of new items as old). This pattern can be seen in the data (bearing in mind that false alarms = 1 – correct rejections). Relative to the –TE group, the similarity between highlighted target pseudowords from training and unenhanced words in the recognition task may have resulted in the opposite bias for the +TE group, namely, a response bias favoring more "new" judgments. That is, there would be more misses (incorrect classifications of old items as new) and correct rejections (correct classifications of new items as new). This was precisely the pattern of results found (again, bearing in mind that misses = 1 – hits). Thus, we propose that perceptual priming may account for these findings. Importantly, this perceptual priming account can easily be tested in future replication and extension studies by, for example, manipulating the perceptual modality of the stimulus items between training and testing.

Our second and third research questions asked whether cognitive effort during initial learning, as indexed by pupil dilation during training, predicted performance in the recognition task. Our results indicated that cognitive effort during training did not uniformly predict recognition task performance. Instead, there was an interaction whereby cognitive effort predicted accuracy for hits for the +TE group, but not the –TE group. That is, correct identification of previously seen words among participants who received textual enhancement depended on degree of cognitive effort (with greater cognitive effort, as indexed by pupil dilation, being related to increased accuracy). On the other hand, for those who did not receive textual enhancement, increased cognitive effort was not related to increased recognition accuracy. What explains this pattern of findings? Once again, we turn to the possibility of perceptual priming. It has been noted elsewhere (e.g., Jacoby & Dallas, 1981) that perceptual priming is insensitive to changes in levels of processing. If this holds for the current study, then increased reliance on perceptual similarity in the –TE group may account for the reduced influence of cognitive effort on recognition performance. Conversely, since the recognition task stimuli were perceptually less similar to the training stimuli, the +TE group may have been more likely to rely on other sources of information that might have benefited from increased cognitive processing during training. Moreover, it is also worth noting that the range of pupil dilation during training was larger for the –TE group. This large amount of individual variability coupled with our small sample sizes may have obscured the effects of cognitive effort in the –TE group.

Whatever the ultimate explanation for these results, our findings are consistent with the wider notion that the efficacy of input enhancement depends, at least in part, upon the depth of cognitive processing that learners engage in during initial encoding. Our results are consistent with several approaches that emphasize the importance of cognitive effort (e.g., Leow, 2015) and other processes beyond noticing in L2 learning (e.g., Sharwood Smith, 1991) and incidental vocabulary learning (e.g., Hulstijn & Laufer, 2001). Our results are also somewhat consistent with other eye-tracking studies on TE (Indrarathne & Kormos, 2016; Winke, 2013) in that we also found that participants' cognitive effort beyond noticing may have been a factor in their learning outcomes. These results suggest that cognitive effort after noticing may also play a role in learning beyond the well-established effects of attention on word learning (Godfroid et al., 2013). These results—particularly if they are replicated and extended in future research—may go some way toward improving our understanding of whether and how cognitive effort and deeper processing more broadly may moderate the effectiveness of TE. However, using the present results to shed light on the hitherto mixed findings on TE should be done with great caution. At the very least, one must bear in mind that our study focused on word form learning, which obviously limits its comparability to the majority of TE studies which have focused primarily on grammar and/or meaning comprehension.

Limitations

Several other limitations to our study warrant caution in generalization. First, there are obvious limitations in sample size, which might obscure the smaller effect sizes associated with pupillometry, increasing Type II error risk. Likewise, although we controlled for stimulus-level confounds in pupil dilation by presenting the target and control words in the exact same sentential context twice, this procedure could have led to larger variability in pupil dilation measurement because pupil dilation reflected differences that could have occurred due to other effects (e.g., repetition of the same sentential context). The study is also limited by the narrow, semi-artificial scope of our experimental design and materials. To some degree, these limitations were necessary due to the delicate nature of pupillometry itself. Although the use of recognition memory tasks and semi-artificial languages certainly have precedent in SLA research (e.g., Hamrick, 2014, 2015), they come with their own limitations and validity threats that necessitate future research using a range of complementary tasks and materials. This study was also limited in that it only investigated incidental learning conditions. Given the ample evidence that different learning conditions often result in different learning outcomes (for an overview, see Leow & Zamora, 2017) and that they may interact with different cognitive processes (e.g., Hamrick, 2015), it will be important for future research to investigate whether and how cognitive effort moderates the efficacy of input enhancement under different learning and instructional conditions.

Conclusion

The present study employed a novel methodology, pupillometry, for investigating the interaction between TE and participants' cognitive effort during training. The results suggest that it may be cognitive effort (or some other deeper mental processing) that plays an important role in moderating the effects of TE on word learning. If such results continue to be found—particularly if they are found across studies using a variety of different research methods—then it would suggest that salience—especially salience that is constructed by researchers and teachers in the form of TE—may be useful for language learners, but critically when it is accompanied by deeper processing (Leow, 2015), more elaboration or instruction (Sharwood Smith, 1991), or more cognitive effort.

Notes

1 Although traditional eye-tracking measures (e.g., fixations) may tap more than just attentional processing, for present purposes we assume that traditional eye-tracking measures tap distinct processes from pupillometry measures, which index cognitive effort—the focus of this study—more specifically.
2 We thank an anonymous reviewer for this point and reference.
3 This was the final set of parameters for our search. Several different mean bigram frequency ranges had to be tested multiple times during the search to establish a mean and standard deviation value that was comparable to the control words.
4 Incidental learning conditions were used to minimize the likelihood of ceiling effects (i.e., where performance is so high that no further gains can be made).
5 We avoided using mixed effects or multilevel models due to the fact that the sheer volume of data (over 6 million individual data points) was computationally too great.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Block, C., & Baldwin, C. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*, 665–670.

Bowles, M. A. (2003). The effects of textual input enhancement on language learning: An online/offline study of fourth semester Spanish students. In P. Kempchinsky & C. Piñeros (Eds.), *Theory, practice, and acquisition: Papers from the 6th Hispanic Linguistic Symposium and 5th Conference on the Acquisition of Spanish and Portuguese* (pp. 395–411). Somerville, MA: Cascadilla Press.

Cabestrero, R., Crespo, A., & Quirós, P. (2009). Pupillary dilation as an index of task demands. *Perceptual & Motor Skills*, *109*, 664–678.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*, 164–194.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, *9*, 198–217.

Gass, S. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Erlbaum.

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*, 483–517.

Godfroid, A., & Winke, P. (2015). Investigating implicit and explicit processing using L2 learners' eye-movement data. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 325–348). Philadelphia: John Benjamins.

Hamrick, P. (2014). Recognition memory for novel syntactic structures. *Canadian Journal of Experimental Psychology*, *68*, 2–7.

Hamrick, P. (2015). Declarative and procedural memory as individual differences in incidental language learning. *Learning and Individual Differences*, *44*, 9–15.

Han, Z., Park, E. S., & Combs, C. (2008). Textual enhancement of input: Issues and possibilities. *Applied Linguistics*, *29*, 597–618.

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*, 510–527.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the Involvement Load Hypothesis in vocabulary acquisition. *Language Learning*, *51*, 539–558.

Indrarathne, B., & Kormos, J. (2016). Attentional processing of input in explicit and implicit learning conditions: An eye-tracking study. *Studies in Second Language Acquisition*, *39*, doi: 10.1017/S027226311600019X.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, *110*, 306–340.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585.

Lee, S. K. (2007). Effects of textual enhancement and topic familiarity on Korean EFL students' reading comprehension and learning of passive form. *Language Learning*, *57*, 87–118.

Lee, S. K., & Huang, H. T. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition*, *30*, 307–331.

Leow, R. P. (2001). Do learners notice enhanced forms while interacting with the L2? An online and offline study of the role of written input enhancement in L2 reading. *Hispania*, *84*, 496–509.

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

Leow, R. P., & Zamora, C. (2017). Intentional and incidental learning. In S. Loewen & M. Sato (Eds.), *Routledge handbook of instructed second language acquisition*, pp. 33–49. New York: Routledge.

Leow, R. P., Egi, T., Nuevo, A. M., & Tsai, Y. C. (2003). The roles of textual enhancement and type of linguistic item in adult L2 learners' comprehension and intake. *Applied Language Learning*, *13*, 1–16.

Leow, R. P., & Mercer, J. D. (2015). Depth of processing in L2 learning: Theory, research, and pedagogy. *Journal of Spanish Language Teaching*, *2*, 69–82.

Loewen, S., & Inceoglu, S. (2016). The effectiveness of visual input enhancement on the noticing and L2 development of the Spanish past tense. *Studies in Second Language Learning and Teaching*, *6*, 89–110.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, *83*(1), 56–64.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: Evidence from pupillometry. *Frontiers in Psychology*, *5*, 1–16.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.

Sharwood Smith, M. S. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, 7, 118–132.

Sharwood Smith, M. S. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, *15*, 165–179.

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs in Cognitive Science*, *5*, 679–692.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203.

Truscott, J., & Sharwood Smith, M. S. (2011). Input, intake, and consciousness. *Studies in Second Language Acquisition*, *33*, 497–528.

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5–31). New York: Routledge.

Wickens, T. (2001). *Elementary signal detection theory*. Oxford: Oxford University Press.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wierda, S., van Rijn, H., Taatgen, N., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*, 8456–8460.

Winke, P. M. (2013). The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition*, *35*, 323–352.