The effect of gap awareness in L2 word learning

Nikos Amvrazis
Aristotle University of Thessaloniki
amvrazis@lit.auth.gr

Περίληψη

Τριάντα δύο ενήλικες σπουδαστές της ελληνικής ως δεύτερης γλώσσας πήραν μέρος στην παρούσα μελέτη που ερευνά το ρόλο του γλωσσικού εξαγόμενου (Swain 1993) στην εκμάθηση λεξιλογίου. Πιο συγκεκριμένα, εξετάστηκε, σε πρώτη φάση, η σχέση των δραστηριοτήτων γλωσσικού εξαγόμενου με τη λειτουργία της επισήμανσης κενών στη διαγλώσσα. Στη συνέχεια ανιχνεύτηκε ο αντίκτυπος αυτής της λειτουργίας στον τρόπο που ο μη φυσικός ομιλητής αξιοποιεί το γλωσσικό εισαγόμενο προκειμένου να καλύψει τα κενά που έχει επισημάνει. Τέλος, αναζητήθηκε η συσχέτιση της λειτουργίας της επισήμανσης με τη συγκράτηση των νέων λέξεων στο λεξικό.

Λέξεις-κλειδιά: επισήμανση, τύπος δοκιμασίας, εκμάθηση λεξιλογίου

1 Introduction

Noticing is a function incorporated within the Output Hypothesis (Swain 1993, 1995) according to which acquisition in the second language (henceforth L2) is facilitated by the learner's awareness of her own state of linguistic limitations while speaking or writing in the target language. In most cases, such awareness could be evoked by unsuccessful retrieval of words or grammar structures in the L2. If the above failure is followed by relevant input containing the target forms, the learners are supposed to process it in a more exhaustive way. The term was also critical in Noticing Hypothesis (Schmidt 1990, 2001) However, the latter is not aligned with the Output Hypothesis as it associates awareness with focal attention on linguistic input the learner is exposed to in the L2.

It turns out that noticing can be triggered by two different mechanisms; an internal one (i.e. Output Hypothesis) that is motivated by the speaker's intention to speak or to respond to another speaker and an external one that directs attentional resources to linguistic input. In this study, we test the impact of noticing as an internal mechanism on focal attention. Our chief goal is to measure the causal impact of asking an L2 learner (henceforth L2er) to produce a word when she does not know it.

2 On Noticing in SLA

The effects of awareness on subsequent lexical embedding have been tested by a large number of studies from a psycholinguistic perspective (e.g. Kornell, Hays and Bjork 2009, Pastötter, Schicker, Niedernhuber and Bäuml 2011, Wissman, Rawson and Pyc 2011, Grimaldi and Karpicke 2012 a.o.). These studies suggest that unsuccessful retrieval attempts "can enhance learning by improving the encoding of unrecalled items during subsequent study trials" (Grimaldi and Karpicke 2012:505). A common design of a study that evokes vocabulary retrieval is one in which the learner is given a cue word and she has to guess the target word (e.g. beach - ?) (Terrace 1963, Slamecka and

Fevreiski 1983; Squires, Hunkin and Parkin 1997, see also Grimaldi and Karpicke 2012 and references therein).

Under the search set model (Kornell et al. 2009), the presentation of the cue word is supposed to bias the activation of a set of related candidates. These candidates become readily available for subsequent encoding once the target word appears in the input. Thus, in the above example, words like 'ocean', 'wave', 'surf' would supposedly form a set of candidates that are promoted to a pending-status, in the sense that certain probes in memory are becoming sensitive to related linguistic material in the upcoming input. Moreover, false alarm during an unsuccessful retrieval of a word seems to exert a substantial impact on the semantic mediators in cue-target trials. Semantic mediators are the words that associate a cue to a target. Recall the example we used above: if the word 'tide' is the target for the cue word 'beach' the words 'surf', 'ocean', 'wave' could be the semantic mediators. Semantic mediators have shown protracted activation on final recognition tests as the learners were likely to reach the target word through the mediators (e.g. to retrieve 'tide' from 'ocean') (Carpenter 2011).

The effects of failed word recovery are also compatible with accounts that attribute vocabulary learning to error correction adjustments made by the subject when a learning event is unsuccessful (Rumelhart, Hinton and McClelland 1986, Kang, Pashler, Cepeda, Rohrer, Carpenter and Mozer 2011). Learning events in which the L2er experiences lack of knowledge trigger general correction mechanisms as a strategy that sheds light on the roots of such errors. In this error-correction framework, commission errors (as opposed to errors of omission or error-less learning) are supposed to benefit later retention through the implication of explicit rather than implicit memory (Mozer, Howe and Pashler 2004). The former is activated in the course of fact learning which takes place when the errors are the result of the learner's responses (i.e. commission errors). This is a critical condition for the learner "to obtain a more meaningful corrective […] signal" (Kang et al. 2011:50).

Along the above lines, it would be safe to say that the accommodation of the linguistic input would benefit from a preceding stage that fosters the learner's uncertainty and, therefore, anticipation about the upcoming input. Following Litman (2008), the learner becomes uncertain and therefore 'curious' (Litman 2008:1586) when is being deprived of the relevant information. Kang et al. (2011) suggest a positive correlation between curiosity and memory performance while Gruber, Gelman and Ranganath (2014) report on such correlation between curiosity and the neural structure that is supposed to enhance the process of learning.

Summing up, a wealth of evidence has led to strong advocacy for the link between noticing and learning exposure to subsequent relevant input. The next section reviews relevant work with respect to our study and justifies the design we followed in our experiment (Section 4). The results and the discussion follow in Sections 4 and 5 respectively.

3. Drawing attention on vocabulary

Ellis and He (1999) were the first to explore the role of output in promoting awareness of vocabulary limitations. In their study, besides the variable of the tasks the researchers investigated the effects of the peer-interaction during the tasks. Two groups of L2ers were treated input and output tasks in different order. The results of the post tests indicated that the students who received output tasks before taking the input task and

had the chance to interact were better at recalling and naming new words than the students in each of the rest conditions.

Kwon (2006) is also concerned with the role of output tasks as a source for developing both receptive and productive aspects of L2 vocabulary. At the same time, she is interested in the different "task performance" (Kwon 2006:18) that is designed to host the task (interaction vs. non-interaction). The researcher designed input and output tasks and manipulated the sequence of the tasks (i.e. input-before-output vs. output-before-input) across different modes of interaction.[1] The data suggested that output-before-input condition along with an interactive classroom environment set the optimal conditions for vocabulary development. Participants who dealt with output tasks before taking the input tasks and were encouraged to negotiate instructions and content were significantly better at naming pictures in the post tests. Although the above studies highlight the complementary effects between output tasks and negotiation processes, they do not tackle the issue of individual differences that might interfere. In other words, the basis upon which interaction was defined and measured is not clear. Moreover, the analysis does not seem to track interaction at the individual's level meaning that learners who did not interact during the process (a common experience in the L2 classroom) were described by the same mean of the respective group.

Turning to the studies that focused on the noticing function as a result of a sole factor, namely the specific type of the tasks, we will find two of them in the domain of vocabulary learning. The first one (Mahmoudabadi, Soleimani, Jafarigohar and Iravani 2015) comprised an input task in which the participants had to match words to pictures and an output task in which the participants had to name pictures. The design of the study was similar to the Kwon's study and resulted in two conditions, input-before-output (control group) and output-before-input (experimental group). The researchers delivered data suggesting high correlation between the sequence of the tasks and the noticing function. The latter was validated by qualitative data reporting that the students that were treated output tasks before input tasks had experienced gaps in their vocabulary. This experience was reflected in a recalling-words post test as significantly better performance by the experimental group.

In a recent study, de Vos et al. (2018) performed a speaking elicitation task in which the experimental group compared the prices of certain objects that should be named while the controls would only inspect the objects and think silently about pricing. Interviews were additionally administered as to ensure the successful induction of noticing through the above procedure. The elicitation task was followed by exposure to spoken input in which a native speaker provided a full description of the objects and their prices. This session was common for both groups and allowed the researchers to link the function of noticing with focal attention to subsequent input. The results of this study revealed a significant effect of the output task on noticing. The researchers acknowledge that noticing is only the first step towards the vocabulary development and postulate the necessity of respective input that will satisfy the need for 'eliminating gaps' (de Vos et al. 2018:2).

In conclusion, the above studies provide evidence underpinning the facilitative effects of output tasks on noticing function. Arguments have also emerged suggesting an interactive classroom setting, even though more refined analyses are required for the latter to be standardized (see Uggen 2012 for triangulation of measures). The design of this study associates Output with Input Hypothesis with precise implications for L2

[1] The task battery comprised input tasks (reading comprehension task, answering questions based on a pictorial scenario, word recognition task) and output tasks (writing and speaking elicitation, fill-in-the-blank task (see de Vos, Schriefers and Lemhöfer 2018 for more details).

vocabulary learning. Our study is a conceptual replication of de Vos et al.'s work as it will be shown in the following section.

4. The present study

The aim of this study is to investigate the role of the output tasks in vocabulary learning and retention. We assume that learning and retention of new words are associated with the function of noticing that is triggered by the output task. To test this assumption, we collected quantitative and qualitative data in a sequential explanatory design.

4.1 Participants

Eighteen learners of Greek as a L2 at the American College of Thessaloniki and fourteen learners at the Hellenic University of Thessaloniki were randomly assigned to the experimental (henceforth EG) and the control group (henceforth CG) respectively.[2] The learners in both groups enrolled in the B1 class and were following the same syllabus.[3] None of the participants reported Greek origins or exposure to Greek while growing up. Greek was the third language for the majority of the participants with some learners reporting Greek as fourth language. However, despite the proficiency level in other languages none of the learners was balanced bilingual. The L1 background of the participants is as follows: English (14), Turkish (5), Russian (5), Spanish (3), Mandarin (3) and German (2). Table 1 summarizes the profile of the participants.

Group	Male Learners	Female Learners	Age[4] of Learners	Mean Score (final exam)	Greek as L2	Greek as L3	Greek as L4
EG	8	10	18-22	84.7/100	3	11	4
CG	6	8	24-30	88..2/100	0	9	5

Table 1 | Participant profile

4.2 Method

We set two steps of noticing investigation. In the first step, we manipulated the type of the tasks to track any effect on noticing gaps in vocabulary. The second step exposed all the participants to input with which we measured noticing as a function of focal attention to the input. The experiment was inserted as warm-up activity during the first hour (out of four) of the unit "Sports and Life Style" and the participants were not aware of the study.

[2] The term ("L2") is used conventionally here. See details bellow in this section.

[3] The proficiency level (B1) was determined by the course completed in the previous academic year. This means that all participants had successfully passed the final exam of the A2 level (mean score 86,4/100/SD: not available). The two groups followed the same syllabus (based on the criteria that are set by the Center of Greek Language) which was also supported by the same course book.

[4] The two institutions did not allow access to personal data of their students. Instead, we were given the range of age of the enrolled students as well as the mean score of the final exam for the A2 level (at class level). Thus, SDs could not be estimated for age and mean score.

4.2.1 Design and procedure

Step 1. To test the assumption that noticing enhances awareness of limitations in vocabulary we asked the EG to describe a number of pictures in cards. The EG performed this output task in pairs using two blocks of cards. We expected that the experience of not knowing a word would set an alert for the respective gap in the vocabulary. Post-experiment interviews were administered to confirm the above assumption (see Appendix 1). The CG was instructed to inspect carefully a number of cards as they were supposed to use them in a follow-up speaking task. This means that the CG worked on the same pictures but the participants remained silent. Therefore, we did not use any kind of input task.[5] Instead, we had the controls engaged by asking them to sort the cards according to certain properties of the sport (indoors vs. outdoors, team vs. solo, water sports). Post-experiment interviews tested the assumption that the controls had not been induced to noticing gaps in their vocabulary.
Step 2. Both groups were exposed to a main session in which spoken input was provided by a male native speaker of Greek who described the same cards that were previously encountered by the L2ers. The input was displayed twice in video format and the participants were instructed to watch it carefully. They were told that the video was based on the cards they had already seen. To test the assumption that the EG would be more effective in eliminating gaps in vocabulary we conducted a post exposure test (picture-naming task). The two groups should name in written form pictures that were projected on a wall screen. The pictures appeared in randomized order and the time window for each picture was fifteen seconds based on the pilot study we run. The task lasted ten minutes including a two-minutes-break.

The same task was carried out sixty minutes later (at the end of the next hour) and tested the assumption that the new words would be more stable in the L2ers' lexicon when a relative gap was pre-activated.

4.2.2 Materials

Ten critical items intermingled with thirty fillers formed the two blocks (twenty trials per block) of the cards that were used in the experiment. The critical items were controlled for not being cognates as well as for their length. The fillers included words of the same semantic category with no restrictions about their cognate status. The target items were also used in context (see bellow) that was assessed for the level of proficiency.[6]

Target items	
2-syllable items	**4-syllable items**
πάλη (wrestling)	ιππασία (equestrian)
μήκος (long jump)	ξιφασκία (fencing)
κρίκοι (rings)	πυγμαχία (wrestling free style)

[5] This is a critical decision for the design as it eliminates potential advantages for the group that is treated input tasks. Input tasks, like the ones that were used in the studies reviewed in the previous section, are suspicious (de Vos et al. 2018:4) of setting a different threshold for the participants who received such tasks. For example, a matching-words-to-the-pictures-task allows for speculations over the meaning of unknown words. This process can accumulatively facilitate word retrieval through the subsequent task (whether an output task or not) (Barcroft 2007), thus providing advantages for those participants.
[6] The context was assessed by the readability certification software:
(http://www.greek-language.gr/certification/readability/index.html)

σκάκι (chess)	ακόντιο (javelin throw)
τόξο (archery)	κατάδυση (diving)

Table 2 | Target items

The above target words and the fillers were depicted by pictograms in forty cards, sized 8cm X 5cm each. The pictograms were taken from the official site of the Olympic Games and were located at the center of the card (see Appendix 2).

The same pictograms were used in the video that was used as the main source of input in the second stage. The experimenter added narration in a series of slides which included all the target items along with twelve fillers.[7] Each item was mentioned twice in a context that provided key information about the sport (whether Olympic sport or not, year of introduction, record and record holders) (see Appendix 3). The video was edited in QuickTime Player (10.3) and lasted eight minutes.

4.2.3 Performance measurement

In the picture-naming task the L2ers should write the word for the pictogram on the screen. Each pictogram appeared in randomized order for 20 seconds. The participants were given a blank numbered list in a sheet of paper (A4 size). They were instructed to feel free to use any written form (even phonemic representation) of the word as they would not be corrected for spelling errors. The task was performed right after the input session to measure immediate learning gains and sixty minutes later to measure sustained learning gains.[8] Once the second session was completed, we recorded each participant reading out the target words in the two lists.[9] A native speaker listened to the recordings and judged the items. An item was target-like on the condition that the listener could understand the word.

4.3 Results[10]

We analysed the quantitative data of the picture naming task by using Mixed ANOVA with group (EG vs. CG) as between-subjects factor and learning gains in time (immediate vs. sustained) as within-subjects factor. The statistical analysis has been conducted on the participants' raw scores and allowed us to determine whether any change in these scores (the dependent variable) was the result of the interaction between the two factors (group and time). There was a significant main effect of group ($F_{(1,317}$=12.732, p<.001, η^2=0.416) and a marginal main effect of time ($F_{(1,317)}$= 7.126, p=.0584, η^2=0.137) on the participants scores. The interaction between the two factors was also marginally significant ($F_{(1,634}$=9.125, p=.0442, η^2=0.176). Table 2 includes the mean scores and the standard deviations (SD) for each condition.

[7] We used twelve fillers due to time limitations as we had to run the experiment within 50min and before the break. This adjustment, however, does not exert any negative effect on the design of the experiment because there were no features of the critical items that should not be notified by the participants. In this view, we did not expect any biases to occur due to increased density of the critical items. Rather, the critical input is enhanced and this, in cognitive terms, could attenuate the processing load.

[8] At the end of the next hour, before the break.

[9] Each participant should pass by the teacher's office to get the score. This was a pretense to have them alone in order to record the items.

[10] Size limitations in this volume have restricted the range of the analyses to those touching upon the main assumptions. Secondary analyses (e.g. length-of-the-word variable, number of L2s) could not be included in this paper.

Word Learning (dependent variable)						
Conditions	**Immediate Learning Gains (ILG)**			**Sustained Learning Games (SLG)**		
Group	Means	SD	N	Means	SD	N
EG	0.32 (57/178)	2.9	18	0.25 (45/178)	2.1	18
CG	0.17 (24/139)	1.1	14	0.08 (11/139)	.3	14

Table 3 | Mean scores across conditions in the picture naming task

Descriptive statistics (see Table 3) can explain the main effect of group. Participants in the EG outperformed the controls across the two levels of time in which the post test was administered. In particular, when the post test took place after the exposure to input, 3.2 out of ten new words were successfully named by each participant (mean score) in the EG. Independent samples *t*-Test indicated that this is by far more accurate performance relative to the one by the CG (1.7 words per participant) ($t_{(81)}$=9.182, p<.001). The difference remained highly significant sixty minutes later ($t_{(56)}$=14.129, p<.001). These results confirm our second assumption as the L2ers that were treated an output task (EG) were more effective at exploiting the input.

Turning to the testing moment, pairwise comparisons attribute its marginal effect on word-learning to the significant decline ($t_{(35)}$=17.035, p<.001) in word retention only in the case of the CG. Sixty minutes after exposure to input the controls retained less than half of the new words. The latter do not decay at the same rates in the EG ($t_{(102)}$=6.152, p=.059) that exhibits significant resistance to time effects. The results are congruent with our assumption that new words would be more stable in the L2ers' lexicon when a relative gap was pre-activated.

The performance we described above is displayed in Figure 1.

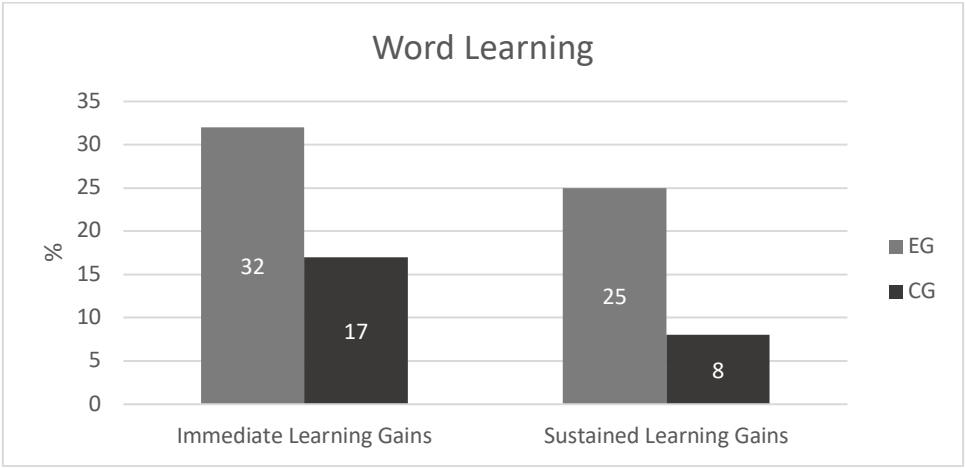

Figure 1 | Word learning across two conditions

Post-experiment interviews verified the first assumption of the study regarding the noticing function that was elicited by the task design. The majority of the participants in the EG reported experience of unknown words. However, an unexpected finding in those interviews has to do with participants in CG that also reported the same experience.

Self-reported experience of Noticing Function			
	Yes	**No**	**Not sure**
EG	12/18 (69%)	4/18 (18%)	2/18 (13%)

CG	4/14 (31%)	8/14 (54%)	2/14 (15%)

Table 4 | Post-experiment interviews report

As depicted in Table 4, four (out of 14) controls told us that they had noticed gaps while two controls did not exclude such a case. Even though these learners were not supposed to notice gaps, it seems that noticing was induced by their sub-vocal intention to name words. The fact that the interviews were administered at the end of the experiment did not allow us to reassign those participants to a separate group in the analysis. We consider this to be a limitation of the design of this study. In the following section, we discuss these results and explain the reason we decided to place the interviews at the end of the experiment.

5. Discussion

In this study two groups of L2ers engaged in different task settings and were treated different opportunities for language production. We aimed at testing Swain's functions of noticing postulating that output tasks can lead learners to notice "a gap between what they want to say and what they can say" (Swain 1995:125). To this end we contrasted the effects of a speaking task to those of an observation task in which the participants remained silent. This setting turned out to be a contrast between vocal and sub-vocal experience of noticing gaps as there were controls reporting experience of gaps in their vocabulary too. The finding here is consistent with de Vos et al.'s (2018) reports on gap consciousness that was, against predictions, triggered by sub-vocal inspection. Those learners were reassigned in a third group in the above study. Unlike de Vos et al.'s study the fact that we conducted our experiment in class settings imposed limitations with respect to collecting qualitative data immediately after the tasks.[11] Therefore, though the assumption of the output impact was confirmed quantitatively and qualitatively, further research should address the issue of sub-vocal intention to speaking.

An important direction for future research would also be the triangulation of the results. As discussed in Section 4.2.1 (see also footnote 5) the output task in this study cannot be directly comparable to input tasks. From a design point of view, the effect of the input tasks should be measured against an inspection task like the one used in this study. This way, safer conclusions could be drawn on whether output tasks better facilitate gap awareness and hence attention to linguistic input than input task conditions. In relation to this, the specific magnitude of different tasks within the above distinction would also be a fruitful ground for future research.

Despite the above unresolved controversies, in the current study there is a certain appeal to interpreting the EG's behaviour as "unsatisfied need-like state that energizes specific exploration aimed at solving problems" (Litman 2008:1595). The results of the above exploration seem to benefit both, vocabulary learning and retention.

[11] de Vos et al. (2018) conducted their experiment in a laboratory and the participants were tested one-by-one. They interviewed each participant right after the task and therefore they could reform the initial setting. This, in turn, required more participants to enroll something that was not an option in our study. Moreover, the addition of a third group in the statistical analysis of our data was not possible as the sample would be too weak even for non-parametric standards.

References

Barcroft, Joe. 2007. "Effects of opportunities for word retrieval during second language vocabulary learning." *Language Learning* 57:35–56. doi:10.1111/j.1467-9922.2007.00398.x

Carpenter, K. Shana. 2011. "Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37:1547–1552. doi:10.1037/a0024140

de Vos, F. Johanna, Herbert Schriefers, and Kristin Lemhöfer. 2018. "Noticing vocabulary holes aids incidental second language word learning: An experimental study." *Bilingualism: Language and Cognition*, 22:500-515. doi:10.1017/S1366728918000019

Ellis, Rod, and He Xien. 1999. "The roles of modified of input and output in the incidental acquisition of word meanings." *Studies in Second Language Acquisition* 21:280-301.

Grimaldi, Phillip J., and Jeffrey D. Karpicke. 2012. "When and why do retrieval attempts enhance subsequent encoding?" *Memory and Cognition* 40:505–513. doi:10.3758/s13421-011-0174-0

Gruber, Matthias J., Bernard D. Gelman, and Charan Ranganath. 2014. "States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit." *Neuron* 84:486–496. doi:10.1016/j.neuron.2014.08.060

Kang, Sean H. K., Harold Pashler, Nicholas J. Cepeda, Doug Rohrer, Shana K. Carpenter and Michael C. Mozer. 2011. "Does incorrect guessing impair fact learning?" *Journal of Educational Psychology* 103:48–59. doi:10.1037/a0021977

Kornell, Nate, Matthew J. Hays, and Robert A. Bjork. 2009. "Unsuccessful retrieval attempts enhance subsequent learning." Journal of Experimental Psychology. *Learning, Memory, and Cognition* 35:989–998. doi:10.1037/a0015729

Litman, Jordan A. 2008. "Interest and deprivation factors of epistemic curiosity." *Personality and Individual Differences* 44:1585–1595. doi:10.1016/j.paid.2008.01.014

Mahmoudabadi, Zahra, Hassan Soleimani, Manoochehr Jafarigohar, and Hasan Iravani. 2015. "The effect of sequence of output tasks on noticing vocabulary items and developing vocabulary knowledge of Iranian EFL learners." *International Journal of Asian Social Science* 5:18–30. doi:10.18488/journal.1/2015.5.1/1.1.18.30

Mozer, Michasel C., Michael Howe, and Harold Pashler. 2004. "Using testing to enhance learning: A comparison of two hypotheses." In *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society,* edited by Kenneth Forbus, Dedre Gentner, and Terry Regier, 975–980. Mahwah, New Jersey: Erlbaum.

Pastötter, Bernhard, Sabine Schicker, Julia Niedernhuber, and Karl-Heinz T. Bäuml. 2011. "Retrieval during learning facilitates subsequent memory encoding." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37:287–297. doi:10.1037/a0021801

Rumelhart, David, Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning representations by back-propagating errors." *Nature* 323:533–536. doi:10.1038/323533a0

Schmidt, Richard W. 1990. "The role of consciousness in second language learning." *Applied Linguistics* 11:129–158. doi:10.1093/applin/11.2.129.

Schmidt, Richard W. 2001. "Attention." In *Cognition and Second Language Instruction,* edited by Peter Robinson, 3–32. Cambridge, UK: Cambridge University Press. doi:10.1017/ CBO9781139524780.003

Slamecka, Norman. J, and Jacobo Fevreiski. 1983. "The generation effect when generation fails." *Journal of Verbal Learning and Verbal Behavior* 22:153–163. doi:10.1016/S0022-5371(83)90112-3

Squires, Ella J., Nicola M. Hunkin, and Alan J. Parkin. 1997. "Errorless learning of novel associations in amnesia." *Neuropsychologia* 35:1103–1111. doi:10.1016/S0028-3932(97)00039-0

Swain, Merrill. 1993. "The Output Hypothesis: Just speaking and writing aren't enough." *The Canadian Modern Language Review* 50:158–164.

Swain, Merrill. 1995. "Three functions of output in second language learning." In *Principles and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson*, edited by Guy Cook and Barbara Seidlhofer, 125–144. Oxford: Oxford University Press.

Terrace, Herbert S. 1963. "Discrimination learning with and without 'errors'." *Journal of the Experimental Analysis of Behavior* 6:1–27. doi:10.1901/ jeab.1963.6-1

Uggen, Maren S. 2012. "Reinvestigating the noticing function of output." *Language Learning* 62:506–540. doi:10.1111/j.1467-9922.2012.00693.x

Wissman, Kathryn T., Katherine A. Rawson, and Mary A. Pyc. 2011. "The interim test effect: Testing prior material can facilitate the learning of new material." *Psychonomic Bulletin & Review* 18:1140–1147.

Appendix 1

Post experiment interviews protocol. The interviewer makes the following questions and takes notes.

1. Have you noticed that you participated in a study in the first hour? [YES NO]
2. Could you rate the following? (scale 1 – 5)
    - Vocabulary is very important for me
    - I try to practice new words in my daily routine
    - In the class I use various techniques to memorize new words (etc. guessing)
3. Did you notice gaps in your vocabulary when you tried to describe the cards in the "Sports task"?
4. In cases of unknown words, did you try to explain the sport in a different way?

Appendix 2

Example of a target and a filler item used in the output and the inspection (sorting) tasks

Target item

Filler

Appendix 3

Example of an item that was used in the input session (video).

Ιππασία