Revisiting the Role of Awareness in Implicit Language Learning: An Extension of Williams (2005)

Florian Riedel (Florian.Riedel@rwth-aachen.de)[1]

E. Kerz (Kerz@Anglistik.rwth-aachen.de)[1] **D.Wiechmann (d.wiechmann@uva.nl)[2]**

[1]Departement of English Linguistics, RWTH Aachen, Kármánstraße 17/19, Aachen, 52072, Germany
[2]Institute for Logic, Language and Computation, University of Amsterdam, Science Park , Amsterdam, 1098 XG, Netherlands

Abstract

Recent years have witnessed an increased interest in the investigation of implicit learning in second language acquisition. A key challenge in this area of research concerns the identification of reliable indicators of incidental learning and the development of experimental paradigms that induce scenarios that afford learning without awareness. Building on design components described in prior research, we present a forced-choice experiment targeting the incidental learning of form-meaning alignments that alleviates some of the tendencies of alternative designs to falsely include episodes of learning in which some degree of awareness must be presumed to be present. Nevertheless, our results suggest the presence of a weak but statistically significant of incidental learning effect.

Keywords: implicit learning; incidental learning; awareness; second language learning; semi-artificial language; generalized additive models; crowdsourcing

Introduction

Implicit learning – "the unselective and passive aggregation of information about the co-occurrence of environmental events and features" (Hayes and Broadbent, 1988, p. 251) – is considered to be fundamental to human cognition (cf. Cleeremans, Destrebecqz, & Boyer, 1998; Shanks, 2005; Perruchet, 2008, for overviews). This domain-general mechanism elicited in incidental learning conditions has attracted a lot of attention in many research areas, including motor learning, object knowledge formation and language acquisition (Perruchet & Pacton, 2006).

There is a long-standing interest in investigating the role of implicit and explicit learning in the area of second language acquisition (cf. Ellis 1994; DeKeyser, 2003; Hulstijn & Ellis, 2005; Williams, 2005; Rebuschat & Williams, 2012, inter alia). This interest was kindled by Krashen's acquisition/learning divide (cf. acquisition-learning hypothesis, 1981, 1994). The question of differences between implicit and explicit learning and their interaction and the role they play in L2 learning is "fundamental in that it determines how one believes second language are learned and whether there is any role of instruction" (Ellis, 2011, p. 35).

While many studies have focused on incidental learning especially in the domain of L2 vocabulary acquisition (see Hulstijn, 2003, for a review) without attempts to account for the implicitness of acquired knowledge as well as on the effectiveness of instruction intervention (cf. Norris & Ortega 2009; Spada & Tomita 2010), there is still relatively little work on the nature of implicit learning mechanisms and knowledge in L2 acquisition (Rebuschat, 2013a).

In the L2 acquisition context, definitions and characterizations of implicit learning typically involve descriptions such as "learning without awareness", "unaware learning" or "incidental learning". Hence, one of the key challenges has been on how to assess whether a learner was aware or not of the statistical regularities in the stimuli s/he was exposed to, if the goal is to determine whether language can be acquired without awareness and whether and the extent to which language can be learned without awareness and, conversely, whether and the degree to which consciousness awareness is required and may facilitate language acquisition.

According to Williams (2009, p. 327) "implicitness can be operationalized only through assessments of subjective mental states, that is, through measurements of awareness." Three types of measures widely used in the implicit L2 literature to determine the conscious or unconscious status of knowledge include retrospective verbal reports, direct and indirect tests and subject measures (cf. Rebuschat et al. 2013b for a review).

A great deal of research on the implicit statistical L2 learning employing artificial grammar learning or serial reaction time paradigms has primarily focused on the abstraction of form-form regularities without considering meaning or function. However, for the acquisition of natural language learning it is great importance to also investigate implicit learning of form-meaning alignments, which has been addressed in more recent research (cf. Williams 2009 for an overview).

DeKeyser (1995) tested for learning without awareness (LwoA) by exposing subjects to a miniature artificial language with rich inflectional morphology used to signal biological gender, number and thematic role. In the implicit learning conditions, subjects were presented auditory stimuli in the new language and were at the same time exposed to aligned visual stimuli (sentences incorporating regularities). After an exposure period of over eight hours, subjects did not perform above chance-level in the test phase, even in cases where the relevant regularities were also present in their respective native language (e.g. subject-verb agreement).

While in one of his earlier studies Robinson (1996) provided some evidence for the detection of the highly constrained pseudo-cleft structure under implicit learning conditions, no evidence for LwoA was found in a more recent study investigating the L2 acquisition of Samoan morphosyntax.

In light of these mixed results Williams (2004, 2005) attempt to clarify the role of awareness in L2 acquisition received much attention. He investigated whether a form-meaning alignment (determiner-animacy mapping) can be learnt when the subjects' attention is directed to a different form-meaning alignment (determiner-distance mapping). During the training phase, subjects were exposed to auditory stimuli which contained noun phrases like "gi dog" which translates to 'the near dog'. In the testing phase the subjects' knowledge of the artificial determiner system was assessed via a two alternative forced choice (2AFC) completion task. Subjects were then asked to report the reasons for their judgments. If the subjects mentioned nothing related to animacy they were classified as being unaware of the animacy form-meaning mapping. Despite this, those participants performed significantly above chance level with an accuracy of about 60% while judging new items during the 2AFC-Task.

A more recent extension by Hama and Leow (2010) produced somewhat different results. They improved the measurements by prompting the participants to verbalize their thoughts while performing the tasks. Their results indicated a strong learning effect for participants who were classified as aware of the animacy rule but no learning effect in the unaware group.

Rebuschat et al. (2013, 2015) argued that this difference might stem from using awareness measurements relying on the verbalization of knowledge by the subjects. In Rebuschat et al. (2013) they reported the results of a close replication of Williams (2005) in which they additionally had the participants choose a source for each judgment and rate their confidence in their judgment. These additional measurements were meant to remove the need for verbalization in the participant's minds and thus to measure awareness with increased sensitivity. Analyzing the retrospective reports, their findings replicated the results of Hama and Leow (2010), however, the subjective measurements showed that low confidence ratings still correlated with above average accurate judgments and that even for the categories guess and intuition the participants performed at about 70% accuracy. They concluded from this that the participants develop at least some unconscious structural knowledge.

This paper contributes to the ongoing debate on implicit statistical language learning of form meaning alignments through the further refinement of the methodological instruments geared to the identification of learning without awareness. Building on design components described in prior research, the goal of this study is to minimize the chance of attributing learning without awareness effects to contexts in which some degree awareness must be assumed to be present.

Method

The present study was designed both to further bias the experiment against the possibility of learning with awareness and filter out participants who did appear to develop some degree of awareness. To this end, we modified the setup of the original experiment to the design sketched in Figure 1.

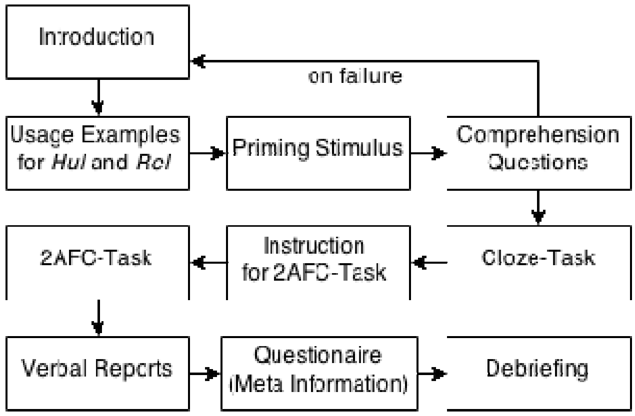

Figure 1: Design of the Experiment

This setup allowed us to bin the participants across three groups depending on their performance in the Cloze task and on basis of their retrospective comments. Participants that used modified forms of the artificial determiners in the Cloze task were tagged as "Cloze-Aware" and those who expressed knowledge regarding the animacy rule in the retrospective comments were tagged as "2AFC-Aware". Participants who did not use modified forms in the Cloze task and expressed no knowledge of the animacy rule were tagged as "Unaware".

Participants

The experiment included a total of 201 participants, 112 women and 89 men with a mean age of 39.1 years. The participants come from Australia (5%), Great Britain (28%) and from the USA (67%). 182 are native speaker of English and 41 speak a second language. 32% of the participants have graduated college, 25% have an associate degree, 40% have a high school diploma and 3% have neither. The participants completed the task on average in 10.26 minutes. We excluded results from participants whose behavior suggested that they did not understand the nature of the tasks or intentionally ignored it, which was indicated by consistent use of only one version of the artificial determiners throughout both tasks (49 participants) and by strong underperformance during the Cloze task on the distance dimension of the determiners (28 participants).

Material and Procedure

The artificial determiner system used in this experiment is inspired by Williams (2005). The system consists of two artificial determiners *hul* and *rel* which encode distance and are used to describe inanimate objects. These items were modified by adding the morpheme 't' at the end of the word when the referent object was animate. *Hul* and *hult* are thus used to refer to nearby objects while *rel* and *relt* are used for nouns that refer to distant objects. We decided to mark the animacy contrast by just one letter in word-final positions in order to minimize the perceptual salience the form signaling animacy. In contrast to the other extensions of Williams (2005) the stimuli were embedded into a larger natural language context to further decrease the salience of the animate forms.

The experiment was deployed using Crowdflower and offered to participants using their subcontractors. The trials were constructed manually using the Crowdflower API using HTML, JQuery and the Crowdflower Markup. The experiment was presented to the participants as survey with the goal to test their ability to learn two artificial determiners encoding distance of the object to the subject.

The subpages of the experiment were presented on white background with black text in Verdana 16 points. After each step was completed the previous step was hidden and the next step displayed instead. The participants were required to navigate through the steps using one button on the lower right corner.

The experiment started with a written text introducing the artificial determiners in the non-animate form and explaining how they encode distance. This was followed by two explained example sentences with an illustrating picture for both determiners.

Next the priming stimulus, a text containing both artificial determiners in their non-animate and animate forms two times each for a total of eight items, was introduced. The natural language context was the first part of a children story written for this experiment. The participants received the task to read the text in detail and were told that content-related questions would be asked.

Once they navigated to the next part they were asked three multiple choice questions on the content. If they did not pass, they were sent back to step 1 with instructions to reread the text. Otherwise the participants were shown the Cloze task.

The Cloze task was presented as text containing eight small textboxes as gaps for the missing determiners. The instructions for the task were to "fill in the correct form of the artificial determiners". The task description was kept intentionally unspecific in order to avoid emphasis on the animate forms and to prevent the participants from including the animate forms. The participants could not submit this task unless they filled out all text boxes using only the four forms of the artificial determiner.

Next a short introduction to the second task was shown. The participants were informed of a hidden rule and asked to choose the correct item in the following task

Then they were presented a two-alternative forced-choice (2AFC) task similar to the one used in Williams (2005). It consisted of a test sentence with a gap containing three question marks for the determiner, two buttons with the animate and inanimate form of the determiner, two five point scales for intuition and certainty levels and a submit button. Clicking one of the two buttons would replace the gap with the determiner displayed on it. Submitting was only possible after choosing a determiner and completing the scales. Both the sequence of items and the position of the determiner-buttons were randomized. For each form of the two determiners four items, a total of sixteen, were displayed and the participants were asked to provide a source attribution and choose a certainty level.

Finally the participants were asked to report whether they recognized the "hidden rule" governing usage of the four forms of the artificial determiner and were asked to describe the rule(s), if they found any.

Results

Like in Williams (2005), performance on the 2AFC-Task served as the measure for learning. In order to prevent memorization all nouns used in combination with a stimulus were unique. We also conceived the natural language context for the stimuli in a way to clearly imply the distance to the object.

Due to the strong bias against aware learning only 14 of the valid 124 participants fit the criteria for the Cloze-Aware group. As expected this group performed both strongly and significantly above chance level with an average accuracy of 79.6% (SD=17.6%). Five of those participants had wrong assumptions on the nature of the rule and performed only slightly above average with 58.8% (SD=13.0%). The other participants performed very high with an average of 90.0% (SD=6.7%).

Of the participants that did were not flagged aware in the Cloze Task 19 expressed some kind of knowledge related to the animacy rule and were flagged as 2AFC-Aware and 28 participants expressed assumptions regarding rules describing other theories. Among those were plurality, ownership of the object, and tense of the predicate.

The 2AFC-Aware group performed at a high accuracy as well. They averaged at 87.5% (SD=12.8%) accuracy and, as displayed in Figure 2, showed a strong learning effect.

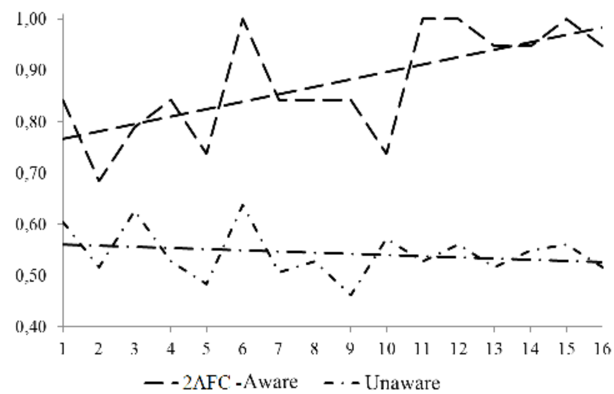

Figure 2: Performance during the 2AFC-Task

Due to the randomization we can rule out any influence from the nature of the question.

Interestingly this does not reflect in the confidence ratings at all. In most cases we find a linear development and in some cases confidence runs even counter to accuracy. This suggests that there are two groups within the 2AFC-Aware group:

One group is already aware of the animacy rule before they start the 2AFC-Task but they did not use the modified forms in the Cloze task due to the ambiguous instructions. Therefore they have consistently high confidence ratings as they have already developed meta-knowledge concerning the use of the rule, whether they apply it consciously or intuitively. That causes this group to display only little development, as they already have a high accuracy.

The other group appears to learn the animacy rule during the 2AFC-Task and therefore has low confidence ratings. This shows in the steeper learning curve. Still, even early on this group performs at a higher accuracy than the Unaware group. This suggests that they also have some awareness of the rule from the very beginning.

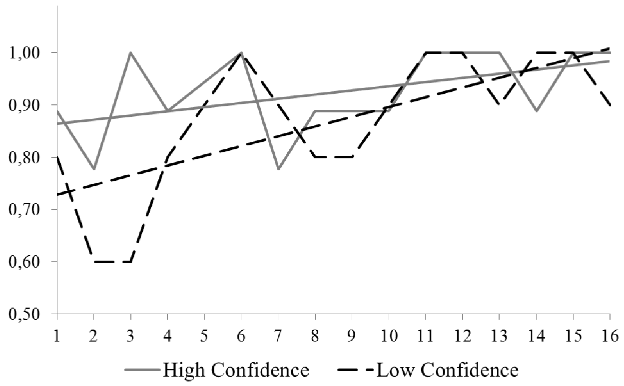

Figure 3: Performance per Item for the 2AFC-Group

However, the sample for these groups is very small and while the difference for the first four items is significant with $p < 0.05$, a larger sample would be necessary to provide convincing evidence.

The largest group is the Unaware group with 91 participants. They judged the items with an average accuracy of 54.3% (SD=13.5%) which is slightly, but significantly ($p < .005$) above chance. As can be seen in Figure 2 they display no signs of learning during the task. Surprisingly the subgroup, which had some wrong assumptions on the rule behind the modification of the artificial determiners, still scored slightly above the subgroup of the completely unaware participants with 55.6%, yet did not do so significantly.

Analysis of the confidence ratings across the Unaware group yielded no significant results. The difference between high confidence judgments (55.7% [SD=14.3%]) and low confidence judgments (51.4% [SD=13.0%]) proved minimal and was not statistically significant (Table 3). The same is true for the source attributions (see Table 3).

Non-native English speakers seem to consistently outperform native speakers. However, as the data pool for this is rather small it is hard to draw any conclusions from this. The same is true for English speakers that are able to speak another language.

Analysis

It is also interesting to have a closer look at the accuracy for each item in the 2AFC-Task: Analysis of the results has shown that the performance across the different items of the task was highly depending on the determiner and in one case on the question.

Table 1: Average Performance for Each Determiner

	Rel	Hul	Relt	Hult
Cloze-Aware	73%	93%	82%	70%
2AFC-Aware	89%	91%	84%	86%
Unaware	59%	70%	48%	40%

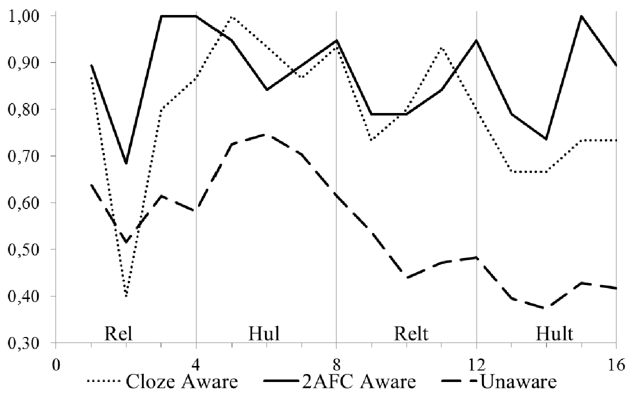

Figure 4: Performance across 2AFC-Task Items

These results indicate that the participants performed significantly better for the introduced determiners *hul* and *rel* (with the exception of item two, which is discussed below) and below chance level for the hidden modifications of the determiners *hult* and *relt*.

This can be interpreted in two ways: first that the participants tend to overuse *hul* and *rel* as they have been introduced and explained and are thus much more salient than *hult* and *relt*. Especially the participants in the Unaware group, who did not notice those two of modifiers in the priming stimulus, are very reluctant to use the new forms and prefer to use the more familiar forms.

Table 2: Overuse of Inanimate Forms

	Cloze-Aware	2AFC-Aware	Unaware
Overuse	7,5%	5,3%	19,9%

On the other hand this is further evidence against the existence of LwoA, as we can only find an above average performance for those items that had a strong focus due to being introduced as being relevant for the study, while the modified forms had less focus and the accuracy is even below chance level for those.

The item-specific accuracy analysis also allows some conclusions concerning which items are considered difficult to learn. Two items are of special significance: item 2 and item 9. Item 2 is "warren" and the most uncommon word in the experiment. This seems to confuse the participants as the error rate is significantly higher compared to the other items associated with rel. This implies that uncommon words are easier associated with new or uncomfortable forms. This is only a single occurrence, but this might justify further analysis.

The second item to cause unusual effect, item 9 is "salmon", which can be interpreted as a mass noun and as such is consistent with some incorrect hypotheses about the for-meaning alignment reported by some participants. A number of comments mention number as rule even though the study only uses nouns in singular but the one in item 9. This explains why there is an above average accuracy for item 9 by the Unaware-group. This is likely due to the highly salient nature of the number attribute and due to the fact that English speakers are used to synthetic modification of words in plural.

Focusing on the "unaware" group, we further analyzed the data using Generalized Additive Models (GAM), which were used to avoid the problematic steps of a priori estimation of response curve shape or a specific parametric response function (Hastie & Tibshirani 1990).

Table 3: Anova for Parametric Effects

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
item	3.00	52.34	17.45	17.24	0.00	***
position	1.00	0.15	0.15	0.15	0.70	
s(rule.scale)	1.00	0.49	0.49	0.48	0.49	
s(certainty)	1.00	2.94	2.94	2.91	0.09	.
Residuals	995.00	1006.86	1.01			

While we found a pronounced effect of ITEM (as described above) - task perfomance was clearly worse for the "t"-items, we found no evidence for effects of POSITION, suggesting that participants did not learn during from trials in the task. Similarly, we found no effects related to the subjective estimate of a participant as to whether their choice was based purely on intuition (low values of RULE.SCALE) and subjective certainty measure.

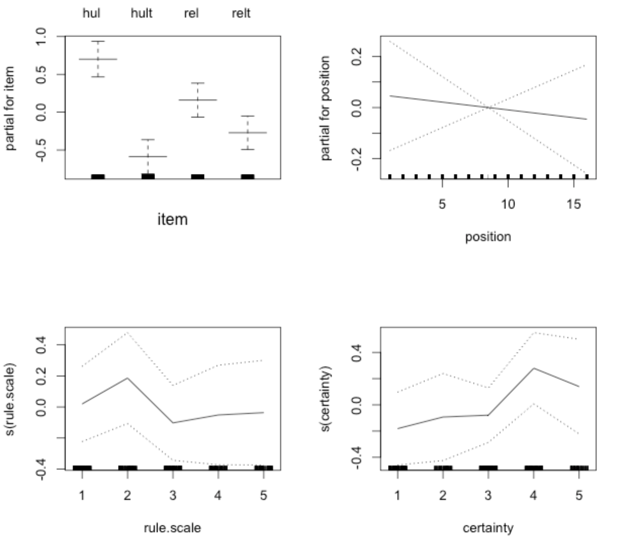

Figure 5: Effect plots from GAM: ITEM, POSITION, RULE.SCALE, and CERTAINTY

Discussion

This experiment sought to contribute to the ongoing debate on implicit statistical language learning without awareness. Specifically, we set out to examine if evidence for learning without awareness could still be found in a design that is strongly biased against the possibility of wrongly including episodes of learning with awareness. In contrast to previous research (Williams, 2005; Hama and Leow, 2010; Faretta-Stutenberg and Morgan-Short, 2011; Rebuschat, 2013; Rebuschat, 2015), this experiment included a new gating mechanism in order to test for aware participants before subjected them to the 2AFC-Task and employed three independent measures of awareness (indirect production test, subjective measures of awareness and retrospective reports). Due to our design we were able to address some issues brought up in Rebuschat et al. (2013).

In light of the considerably larger sample size of our study, the fact that we found a statistically significant but weak learning effect for the Unaware group, seems to suggest Hama and Leow (2010) overestimated the influence of explicit learning and Williams (2005) underestimated it.

Furthermore, our results suggest at a strong learning effect in the 2AFC-Aware group. This leads to conclude that participants with some unconscious rule knowledge quickly developed conscious knowledge when confronted with the necessity to employ it. However, due to the bias against this type of learning the results are not significant.

References

Cleeremans, A., Destrebecqz, & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences, 2,* 406-417.

Ellis, N. (Ed.) (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N. (2011). Implicit and explicit SLA and their interface. In C. Sanz & R.P. Leow (Eds.), *Implicit and explicit language learning: conditions, processes, and knowledge in SLA and bilingualism* (pp. 35-49). Georgetown: Georgetown University Press.

DeKeyser, R. M. (1995). Learning second language grammar rules: an experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17,* 379-410.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M.H. Long (Eds.), *Handbook of second language acquisition* (pp. 313-348). Oxford: Blackwell.

Hama, M. & Leow, R.P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition, 32,* 465-491.

Hastrie, T. & Tibshirani, R. (1990) *Generalized additive models*. Chapman and Hall.

Hayes, N.A. & Broadbent, D.E. (1988). Two modes of learning for interactive tasks. *Cognition, 28,* 249-276.

Hulstijn, J.H. & Ellis, R. (Eds.). (2005). Implicit and explicit second-language learning. [Special Issue]. *Studies in Second Language Acquisition, 27(2).*

Krashen, S.D. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.

Krashen, S.D. (1985). *The input hypothesis.* London: Longman.

Norris, J. & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning 51,* 157–213

Perruchet, P. (2008). Implicit learning. In H. Roediger (Ed.), *Cognitive psychology of memory. Vol. 2 of learning and memory: A comprehensive reference* (pp. 597-621). Oxford: Elsevier.

Perruchet, P. & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Science, 10(5),* 233-238.

Rebuschat, P. & Williams, J.N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33,* 829-856.

Rebuschat, P. (2013a). Implicit learning. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 298-302). London: Routledge.

Rebuschat, P. (2013b). Measuring implicit and explicit knowledge in second language research. *Language Learning 63(3),* 595-626.

Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2013). Implicit and explicit knowledge of form-meaning connections: Evidence from subjective measures of awareness. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka, (Eds.), Noticing and second language acquisition: Studies in honor of Richard Schmidt (pp. 255–275). Honolulu: University of Hawai‘i, National Foreign Language Resource Center.

Rebuschat, P., Hamrick, P, Sachs, R., Riestenberg, K. & Ziegler, N. (2015). Triangulating measures of awareness: a contribution to the debate on learning without awareness. *Studies in Second Language Acquisition.*

Robinson. P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition, 19,* 27-67.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London: Sage.

Spada, N. & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning 60(2),* 263–308.

Williams, J. N. (2004). Implicit learning of form-meaning connections. In B. VanPatten, J. Williams, S. Rott, & M. Overstreet (Eds.), *Form meaning connections in second language acquisition* (pp. 203–218). Mahwah, NJ: Lawrence Erlbaum Associates.

Williams, J.N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27(2),* 269-304.

Williams, J.N. (2009). Implicit learning in second language acquisition In William C. Ritchie and Tej K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319-353), Emerald Group Publishing Limited.