Applied Psycholinguistics (2021), **42**, 703–737
doi:10.1017/S0142716421000023

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

The role of modality and awareness in language learning

Chuqi Zhao[1], Judit Kormos[2,*], Patrick Rebuschat[3] and Shungo Suzuki[3]

[1]Lancaster University, [2]Lancaster University and University of Vienna and [3]Lancaster University
*Corresponding author: Judit Kormos, Deparment of Linguistics and the English Language, Lancaster University, County South, Lancaster, United Kingdom LA1 4YL, UK. E-mail:j.kormos@lancaster.ac.uk

(Received 05 June 2019; revised 09 October 2020; accepted 03 December 2020; first published online 03 March 2021)

Abstract

There is conflicting empirical evidence regarding the role of awareness in second language learning. Possible explanations for the contradictory results include the modality in the exposure and assessment phases of previous experiments. Our study investigated the acquisition of a novel determiner system under incidental exposure conditions and examined the effect of modality in both exposure and assessment phases. Animacy served as a hidden regularity in the determiners, which were embedded in sentences and presented to Chinese speakers of English either in auditory or in visual mode. Learning was assessed by a two-alternative forced-choice test either auditorily or in writing. Implicit and explicit knowledge were measured using retrospective verbal reports and source judgements. Bayesian analysis provided moderate evidence for above chance level learning. Significant learning effects were observed regardless of whether participants based their accuracy judgements on explicit or implicit knowledge. Bayesian analysis showed moderate evidence for above chance learning effects for aware participants. Generalized linear mixed-effects modeling revealed a small-size significant benefit of the auditory exposure modality over the written modality but indicated no significant effect of the modality of assessment or awareness. Our research underscores the importance of considering the role of modality of exposure in incidental second language learning contexts.

Keywords: awareness; implicit/explicit knowledge; incidental learning; language learning; modality effects

There are differing views and conflicting empirical evidence regarding the role of awareness in second language (L2) acquisition. On the one hand, the view that awareness is necessary for L2 acquisition in adulthood (Schmidt, 1990) is supported by a number of empirical studies (e.g., Leow, 1997, 2000). On the other hand, Williams (2005) proposed that learning a novel language in adulthood can occur without awareness, such as the acquisition of form–meaning connections. Several studies have adapted Williams's (2005) paradigm in order to contribute to the debate on the role of

© The Author(s), 2021. Published by Cambridge University Press.. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

awareness in language learning (e.g., Chen et al., 2011; Faretta-Stuttenberg & Morgan-Short, 2011; Hama & Leow, 2010; Kerz, Wiechmann, & Riedel, 2017; Rebuschat, Hamrick, Riestenberg, Sachs, & Ziegler, 2015; Rebuschat, Hamrick, Sachs, Riestenberg, & Ziegler, 2013), but results have been contradictory, with no clear evidence supporting either position.

In part, the conflicting findings in these studies can be explained by important conceptual differences. Williams (2005), following the tradition in cognitive psychology (Reber, 1967), focused on the implicitness of the learning *product*, that is, what was learned by participants, while Leow (1997, 2000; Hama & Leow, 2010), approaching the topic from the perspective of the field of second language acquisition (SLA), focused on the implicitness of the learning *process*, that is, how learning took place. Given their respective orientations, the authors measured different, albeit interrelated, aspects of language learning. Williams (2005) used retrospective verbal reports to establish if learners could account for their performance at the end of the experiment. In contrast, Leow (1997, 2000) pioneered the use of concurrent verbal reports to assess awareness at the time of encoding (see Rebuschat, 2013, for review).

Another potential explanation for the conflicting findings might be the changes in modality between exposure and testing phases in the studies (cf. Hama & Leow, 2010). Williams (2005) and Faretta-Stutenberg and Morgan-Short (2011) trained participants on auditory sentences but tested them on written items. In contrast, Hama and Leow (2010) and Rebuschat et al. (2013, 2015) kept the modality constant. However, Hama and Leow trained and tested participants on auditory items, while Rebuschat et al. did so with written items (for a recent overview, see Kim & Godfroid, 2019). It is well known that modality makes a difference in memory recall (Shanks & St. John, 1994), and there is recent evidence suggesting that modality also plays a key role in implicit learning (Frost, Armstrong, Siegelman, & Christiansen, 2015). However, there is a scarcity of studies that directly investigate how input modality might affect the learning of a novel language in adulthood. Recently, Kim and Godfroid (2019) addressed this important question, but their participants read out the written sentences, which meant that in the written condition they received dual mode input (visual and auditory).

In the present study, we address this gap by investigating how modality affects the acquisition of artificial determiners by Chinese speaking learners of English. We also examined the effect of modality on incidental learning in the exposure and testing phases. In this paper, we first define key concepts and then provide an overview of prior research that has examined the role of awareness in language learning and the impact of modality in incidental learning. Next, we describe the procedures of our research and present our findings relating to the incidental acquisition of novel form–meaning relationships through visual and auditory input. Our results are discussed with reference to previous research and cognitive theories of L2 learning. The paper concludes with suggestions for future research.

Implicit learning, explicit learning, and the concept of awareness

Our ability to unconsciously derive information from the environment is a fundamental aspect of human cognition. The concept of *implicit learning*, that is, learning that occurs without awareness of what was learned, is typically associated with Reber

(1967), though interest in the topic precedes Reber's work (e.g., Thorndike & Rock, 1934). The topic has been the focus of extensive theoretical discussion and empirical research in both cognitive psychology (Dienes, 2012; Perruchet, 2008) and SLA research (e.g., Williams, 2005; see contributions in Ellis, 1994; Rebuschat, 2015). Awareness plays a key role in distinguishing implicit from explicit learning. Implicit learning is typically observed under incidental exposure conditions, where participants are not informed about the learning target or that they will be tested. This type of experimental manipulation is generally assumed to result primarily in implicit (unconscious) knowledge, that is, knowledge that participants are unable to verbalize. In contrast, explicit learning is usually associated with intentional exposure conditions, which might involve "conscious comparisons between current and previous instances of input and the formation and testing of hypotheses" (Williams, 2005, p. 271). Explicit learning involves some level of awareness depending on how deeply one engages in processing novel L2 forms (Leow, 2018). L2 learners can have lower levels of awareness without being able to verbalize L2 rules or linguistic regularities. For example, when processing L2 input, learners might be aware of the occurrence of verb forms with past tense morphology (*walked*, *jumped*, *played*, etc.) but not fully grasp the underlying rule that generates the regular English past tense (adding an *–ed* to the verb stem). In the former case, Schmidt (1990) would argue that learners display awareness at the level of noticing, whereas in the latter case they display awareness at the level of understanding.

In the field of SLA, Schmidt's noticing hypothesis (1990, 1995, 2010) distinguishes between a narrow and a broad sense of awareness. In the narrow sense, awareness as *noticing* can be considered a surface level phenomenon, namely, "conscious registration of the occurrence of certain events" (Schmidt, 1995, p.19). This is similar to Tomlin and Villa's (1994) conceptualization of low level of awareness, which they define as "a particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus" (p. 193). In the broad sense, awareness as *understanding* involves a deeper level of processing that results in "a state of mind in which one has become cognizant of the regularities underlying the data" (Schacter, 1989, p. 377). Leow (2018, p. 773) further suggests that awareness should be operationalized at three levels: "behavioral or cognitive change (at the level of noticing), a meta-report of the experience but without any metalinguistic description of a targeted underlying rule (at the level of reporting), or a metalinguistic description of a targeted underlying rule (at the level of understanding)." The graded nature of awareness underlies Hulstijn's (2015) argument that explicit and implicit learning should not be seen as a dichotomy but rather a continuum. A key question for SLA research has been to what extent incidental learning can successfully contribute to novel knowledge representations and what depth of processing is needed for learning to take place (see Chan & Leung, 2014; Chen et al., 2011, Leow, 2015, Leung & Williams, 2011, 2012, 2014).

In his seminal study, Williams (2005) presented initial evidence of learning without awareness, which challenged the view that awareness could be a prerequisite for language learning in adulthood (Schmidt, 1990). Williams employed an artificial determiner system (*gi*, *ro*, *ul*, *ne*) that encoded both distance (near vs. far) and animacy (animate vs. inanimate) to investigate the acquisition of form–meaning connections under incidental exposure conditions. Participants were informed that the four determiners' functions were similar to the English word *the*, except that meaning of distance was also encoded: *gi* and *ro* preceded near objects, whereas *ul* and *ne* preceded far objects. However, they were not told that animacy also played a role in determiner usage, and animacy thus served as a hidden regularity. Participants were exposed to this artificial determiner system and later tested only on the hidden regularity. Immediately after the test, participants provided retrospective verbal reports to determine whether they were aware or unaware of the role that animacy played in determiner choice. Those who "were able to state the specific relationships between the determiners and noun animacy" (Williams, 2005, p. 283) were classified as being aware, those who were not able to provide a report were classified as being unaware.

Williams's (2005) results showed that 80% of participants reported having no awareness of animacy after their first exposure to the test sentences, but their performance was still significantly above chance (61% accuracy on generalization items). These unaware participants were then exposed to more training items and asked to consciously identify the hidden regularity. Despite these instructions, 50% of participants still failed to verbalize the hidden regularity, that is, they remained unaware even after an additional round of exposure and under intentional exposure conditions. Interestingly, their performance was still significantly above chance (58% accuracy). The results were interpreted as evidence that learning of form–meaning connections can occur without awareness.

Several studies have adapted Williams's (2005) paradigm to further investigate whether learning without awareness is possible (e.g., Chen et al., 2011; Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Rebuschat et al., 2013, 2015). Hama and Leow (2010) extended Williams's (2005) study by adding (a) another measure of awareness (concurrent verbal reports, or think-aloud protocols) and (b) a production task in addition to a multiple-choice comprehension test and by keeping (c) the same modality for the exposure and testing phases (auditory). The results of Hama and Leow (2010) showed that unaware participants performed significantly above chance on both the production and multiple-choice assessment tasks. However, analyses of the types of responses revealed that participants relied significantly more on distance bias than on animacy bias in both the multiple-choice and production tasks. These results were interpreted as a lack of evidence for significant learning of the animacy regularity among unaware participants. Faretta-Stutenberg and Morgan-Short (2011) also extended Williams's (2005) study, and their results showed that the performance of unaware participants on generalization test-items was not significantly above chance. This supported the findings of Hama and Leow's (2010) research.

Chen et al. (2011) adapted Williams's (2005) design to native Chinese speakers by using Chinese low-frequency characters as determiners and frequent Chinese characters as nouns. Their series of experiments conceptually replicated Williams's (2005) study, showing that participants could learn the novel form–meaning mappings when the hidden regularity was based on a feature that occurs in natural languages. They also showed that the acquired knowledge could be implicit (unconscious) as evidenced by trial-by-trial subjective measures of awareness (Dienes, 2004, 2012). In this method, participants report, for every test judgment, what the basis of their decision was: guess, intuition, recollection, or rule knowledge. The former two categories are associated with implicit knowledge and the latter two with explicit knowledge.

Like Chen et al. (2011), Rebuschat et al. (2013) also employed subjective measures of awareness in their extension of Williams's (2005) study. However, they also added retrospective verbal reports (the awareness measure used by Williams, 2005) to directly compare both measures of awareness. This type of comparison is important, given that measures are differentially sensitive. After classifying participants as aware and unaware based on their retrospective verbal reports, Rebuschat et al. (2013) found that only aware participants (those who were able to report the hidden regularity) performed significantly above chance on all types of test items, whereas unaware participants did not perform significantly above chance on any types of test items. This would suggest that learning was linked to awareness, supporting Hama and Leow's (2010) findings. However, the analyses of the subjective measures of awareness revealed that the picture was more complex; their participants had acquired both explicit (conscious) and implicit (unconscious) knowledge. This result was interpreted as support for Williams (2005). Rebuschat et al. (2015) further expanded this work by comparing three measures of awareness (retrospective verbal reports, subjective measures of awareness, and concurrent verbal reports). They found again evidence of implicit knowledge (as evidenced by subjective measures of awareness) and that adding concurrent verbal reports to the experimental design impacted learning in the group that had to think aloud while completing the training task. While the different groups showed a clear learning effect, in the case of the think-aloud group, learning appeared limited to recognition of items repeated from training during the test phase, that is, these participants seemed unable to generalize the acquired knowledge to novel items. Taken together, studies by Williams (2005), Chen et al. (2011), and Rebuschat et al. (2013, 2015) suggest that adults are able to acquire novel form–meaning connections without intending to and without awareness of the acquired knowledge (the product of learning). In contrast, Hama and Leow's (2010) and Faretta-Stutenberg and Morgan-Short's (2011) analyses of unaware participants' performance indicated that adults do not acquire novel form–meaning connections without awareness.

The differences between the results of Williams (2005) and subsequent extension studies might be due to two important factors: the research methodology and the modality of stimuli presentation. In terms of methodology, there were important differences in how awareness was measured. The awareness measures employed by Williams (2005), Faretta-Stutenberg and Morgan-Short (2011), and Rebuschat et al. (2013) focused on the awareness of learning product, whereas Hama and Leow (2010) and Rebuschat et al. (2015) assessed awareness of learning as both a product and a process (see Leow, 2015, for an overview). Williams (2005), Hama and Leow (2010), and Faretta-Stutenberg and Morgan-Short (2011) did not include subjective measures to examine awareness of learning that were employed by Rebuschat et al. (2013, 2015). In addition, the stimulus presentation and the time permitted to enter a response (i.e., *response window*) in the exposure and testing phases of these studies may have varied. With the exception of Chen et al.'s (2011) study, none of the authors of other studies clearly reported whether their multiple-choice tasks were timed or untimed, which makes it impossible to investigate the role of timing in these studies. Chen et al. (2011) reported that participants had to respond within 20 s during the training task, otherwise the next test item was presented. However, there was no information about timing during testing, which is potentially problematic, given that timing during test can affect the type of knowledge that is being assessed (Ellis, 2005; Godfroid et al., 2015; though see Vafaee, Suzuki, & Kachinske, 2017).

Finally, given the stimuli specificity of implicit statistical learning (ISL; Conway & Christensen, 2006; Frost et al., 2015), the inconsistent use of modalities (auditory vs visual) in the exposure and test phases in Williams (2005) and its extension studies (Hama & Leow, 2010; Rebuschat et al., 2013, 2015) may be one reason for the conflicting results (though see Plonsky, Marsden, Crowther, Gass, & Spinner, 2019). Rebuschat et al. (2013) called for empirical research on a comparison between the effect of auditory and visual modality in implicit L2 learning.

The effect of modality in implicit and incidental learning

Recent research has demonstrated different modality effects in ISL, and Frost et al. (2015) have argued for the coexistence of domain generality and modality specificity in their novel ISL framework. For example, Conway and Christiansen (2006) found parallel effects of visual statistical learning and auditory statistical learning, and Silva, Folia, Inácio, Castro, and Petersson's study (2018) demonstrated similar behavioural outcomes across visual and auditory modalities. In contrast, two artificial grammar learning studies, Conway and Christiansen (2005, 2009), revealed that the auditory modality was superior to the visual modality. Modality effects were also found to depend on the presentation speed of stimuli, showing that audition is superior to vision at fast rates, whereas vision is superior to audition at slow rates (Collier & Logan, 2000; Emberson, Conway, & Christiansen, 2011). Qualitatively, Conway and Christiansen (2009) found a visual advantage for a primacy effect (better recall of first elements in a list) and an auditory advantage for a recency effect (better recall of final elements in a list) in an ISL task, which is consistent with Beaman's (2002) results regarding serial recall. ISL studies also suggest that auditory and visual input may be processed by separate mechanisms in the brain (Conway & Christiansen, 2006; Siegelman & Frost, 2015), which is supported by Silva et al.'s (2018) investigation of event-related potentials in response to grammar violations. It has also been found that ISL of artificial grammars can be transferred across different modalities (Altmann, Dienes, & Goode, 1995). In addition, differences in the developmental trajectory of auditory and visual ISL abilities have been observed in children (Raviv & Arnon, 2017). Taken together, most ISL studies have revealed multidimensional modality-specific effects, which are worth further investigation.

Previous SLA research has also compared the effect of exposure modality (auditory vs. visual). For example, Wong (2001) found better comprehension of written than auditory input. Leow's (1993, 1995) studies of L2 learners of Spanish suggest that learners' intake of both salient forms and less salient forms was equally successful from written input, whereas the intake of salient form was significantly more effective from auditory input. Cintrón-Valentín and Ellis (2016) found that visual input directed learners' attention to morphological cues more successfully than did aural input. More recently, Kim and Godfroid (2019) adapted the semiartificial language developed by Rebuschat and Williams (2012) to examine differences in the incidental learning of German word order rules in auditory and visual input conditions. Participants were exposed to three verb placement rules in written and auditory modality and received implicit feedback in the exposure phase. In the written mode, participants had to read out the target sentences, whereas in the auditory mode they listened and repeated the sentences. Implicit feedback meant that the participants were asked to repeat the sentences up to three times if they did not produce the correct word order. Similar learning effects were observed regardless of the modality of the input, but students developed implicit knowledge only in the visual condition.

Previous research on the effect of assessment modality has compared grammaticality judgment tasks (GJT) in auditory and written modalities. Most studies have found learners' performance is more accurate in written than in auditory grammaticality judgments (Johnson, 1992; Kim & Nam, 2017; Murphy, 1997; Shui, Yalçın, & Spada, 2018; Spada, Shui, & Tomita, 2015). Kim and Godfroid's (2019) recent study found that participants scored better in the auditory GJT compared to a written GJT that was presented word by word, but they did not differ when the written GJT was shown in whole sentence format. Plonsky et al.'s (2019) recent meta-analysis suggests that modality effects in grammaticality judgements are neither strong nor stable. In contrast, their meta-analysis revealed large and reliable differences between untimed and timed grammaticality judgment tests, with participants scoring higher in untimed than timed conditions. Apart from GJT studies, there are very few studies on the effect of assessment modality in the field of SLA. Of interest, Ziegler's (2016) meta-analysis of synchronous computer-mediated communication (SCMC) and interaction revealed that the same modality of training and testing contributes significantly to learning gains. This lends support to models of transfer-appropriate processing, which argue that learning is more effective if the cognitive processes employed in the encoding of new information are similar to those that are used for the retrieval (and assessment) of newly learned material (Roediger & Guynn, 1996).

Research questions and predictions

As highlighted above, the possibility of learning without awareness is still subject to debate in the domain of adult SLA. Furthermore, there is a scarcity of studies on the effect of exposure and assessment modality under incidental exposure conditions, and a potentially critical variable, namely, the response window during exposure and test tasks, might not have been controlled rigorously in previous studies. The present study aimed to directly address these gaps. In doing so, we also intended to contribute to the debate on learning without awareness by further extending Williams's (2005) study. To overcome the methodological shortcomings of previous research, we controlled the response window in both exposure and testing phases carefully. We also wanted to examine the effectiveness of incidental learning in a situation where the novel form–meaning connection is embedded in an L2 context. Our research addressed the following questions:

1. What characterizes the knowledge that Chinese L2 users develop when exposed to novel determiners under incidental exposure conditions?
2. How does modality of exposure and test phases (auditory vs. written) influence the acquisition of novel determiners under incidental exposure conditions?

As regards RQ1, based on Rebuschat et al.'s (2013, 2015) and Leung and Williams's (2014) studies, we hypothesized that Chinese L2 users would be able to establish new form–meaning connections in an incidental learning context, that is, without the intention to learn the hidden regularity and without advance knowledge of a test. Based on Hama and Leow (2010), Faretta-Stutenberg and Morgan-Short (2011), and Rebuschat et al. (2015), we expected that participants who demonstrated some level of awareness would display a greater learning effect than those participants who did not. In line with Rebuschat et al. (2013) and Rebuschat et al. (2015), we expected to find evidence for both implicit (unconscious) knowledge and explicit (conscious) knowledge in the test phase. For RQ2, we predicted a facilitative effect of the auditory presentation mode based on the results of artificial grammar learning studies in the field of cognitive psychology (cf. Conway & Christiansen, 2005, 2009). Based on research findings on the role of modality in accuracy in GJT, we hypothesized that participants' response would be more accurate in the written mode. However, as a result of transfer appropriate processing (Roediger & Guynn, 1996), participants in the auditory exposure condition can be hypothesized to be more likely to give correct responses in the auditory testing mode.

Method

Participants

Eighty-eight students (42 women; $M_{age} = 23.42$) at a UK university volunteered to participate in the experiment. After providing informed consent, participants were randomly assigned to one of four experimental conditions ($n = 22$ each). Participants in the *auditory–visual* (AV) condition received auditory stimuli in the exposure phase and written test-items in the test phase. Participants in the *visual–auditory* (VA) condition were exposed to written stimuli and auditory test-items in the test phase, while those in the *auditory–auditory* (AA) condition received auditory stimuli in the exposure phase and the test phase. Finally, participants in the *visual–visual* (VV) condition received written stimuli in both the exposure phase and the test phase.

All participants were native speakers of Chinese (Mandarin, $n = 67$; Cantonese, $n = 19$) who were learning English as an L2. Their self-reported IELTS scores ranged from 6.0 to 7.5, which is equivalent to B2 to C1 level in the Common European Framework of Reference (CEFR) (Council of Europe, 2001). Apart from English, 26 participants reported studying the following additional languages: Mandarin ($n = 9$), Japanese ($n = 6$), Spanish ($n = 6$), French ($n = 6$), German ($n = 5$), Korean ($n = 3$), Russian ($n = 1$), Minnan ($n = 1$), and Malay ($n = 1$). Seventeen participants (5 in the AV, 2 in the VA, 7 in the AA, and 3 in the VV conditions) reported that they had taken courses in linguistics before the experiment. A series of Chi-square tests revealed that the groups did not differ significantly in terms of the distribution of age, gender, number of L2s, and participants' language background.

Stimulus materials

The training and test items were adapted from Rebuschat et al. (2013, 2015). The artificial determiner system developed by Williams (2005) was used in this study.

Table 1. Design of training items

Each set	Noun	Determiner–noun combinations	Singular nouns	Plural nouns
24 determiner–noun combinations	12 animate nouns	6 near-animate: *gi* + noun	3	3
		6 far-animate: *ul* + noun	3	3
	12 inanimate nouns	6 near-inanimate: *ro* + noun	3	3
		6 far-inanimate: *ne* + noun	3	3

The system consists of four artificial determiners (*gi*, *ro*, *ul*, and *ne*), whose usage depends on the distance and animacy of the associated noun. The determiners *gi* and *ro* are used with nouns that refer to near objects, and *ul* and *ne* refer to distant objects. The determiners *gi* and *ul* refer to animate (living) entities, whereas *ro* and *ne* refer to inanimate (nonliving) entities. In line with previous studies, participants were instructed about the role of distance in determiner usage, but the animacy regularity was not revealed to them. Animacy thus served as a hidden regularity.

Training items

The training items consisted of two training sets. Each set included 24 determiner–noun combinations that were repeated three times over the exposure phase. These noun phrases (NPs) were constructed with 12 animate and 12 inanimate nouns. Half of the NPs in each category were presented in near-context sentences and the other half in far-context sentences. Half of the nouns in each category were in singular form and the other half in plural form. The order of the total of 144 sentences was randomized for each set, and plurality was counterbalanced within each set (see Table 1). That is, if *gi rats* was presented in Set 1, then *gi rat* was presented in a different context in Set 2; then Set 1 appeared again, and so on. Following Williams (2005, Experiment 2) and Rebuschat et al. (2013, 2015), each noun was matched with only one determiner to guarantee that learning occurred as a result of establishing form–meaning connections rather than form–form associations between determiners. For example, there were *gi rat/s* (near-animate) but not *ul rat/s* (far-animate), so that it could be concluded that *gi* is used with animate nouns, but not that any noun that follows gi can also be preceded by *ul*. The items used in training and testing are reproduced in Appendix A. Noun phrases are reproduced in Appendix B.

Test items

Thirty-six new-context sentences were used to assess all participants. The sentences had not appeared during the exposure phase and were presented in the same random order. The sentences consisted of three types of NPs: trained, partially trained, and untrained. The NPs in the trained items were those that appeared in exactly the same form in the exposure phase (e.g., *gi rats*). The NPs in the partially trained items had appeared during exposure but were changed into opposite near-far determiners.

For example, *ul monkeys* (far monkeys) that occurred in training was partially changed into *gi monkeys* (near monkeys) in the test. Novel nouns that had not occurred in the exposure phase were used to construct untrained items (e.g., *gi goat*). The animacy and inanimacy of NPs was counterbalanced, and items were presented to participants in random sequences (see Appendix C for a list of the nouns in the testing phase).

Given that the participants in the present study were nonnative English speakers, we wanted to ensure that acquisition of the four artificial determiners would not be hindered by unknown or unfamiliar English words in the stimuli. Therefore, we made adaptations to the original training and test items from Rebuschat et al. (2013, 2015). All English words and phrases in the original items were checked on the website English Vocabulary Profile (http://vocabulary.englishprofile.org). English words that were above B1 level on the CEFR were either replaced by synonyms or other words, or deleted.

The training and test items used in the two modalities during the exposure and testing phases were identical, except that the order of training items within each set in the exposure phase in each modality was independently randomized for each participant. In the visual modality, written sentences were presented. In the auditory modality, audio recordings of the sentences read by a female native speaker of British English were played. To ensure learning and assessment of the artificial determiners would not be hindered by the English proficiency level of the participants, the speaker used a delivery rate of 100–120 words per minute, which is lower than the normal speaking rate (Crystal & House, 1990).

For each training and test item, we calculated a time range during which participants could enter their responses. This response window was calculated based on the results of a preliminary pilot study (cf. procedures in Shiu et al., 2018). First, words per second for the shortest and longest trials of the two modalities from the pilot results were calculated. Then, these figures were used to divide the number of words of the shortest (6 words) and longest items (20 words) in the present study. That is, the response window took into account the time needed for reading or listening to the input stimulus, that is, it included the presentation of the stimulus item as well. For the written modality, the response window ranged from 3 to 15 s depending on the length of the sentence, and the response window in the auditory modality ranged from 5 to 12 s, again depending on the length of the stimulus sentence (see Appendix D for more details).

Procedure

All participants completed (a) vocabulary pretraining of the four artificial determiners for the distance regularity, (b) an exposure phase, (c) a testing phase, (d) a retrospective verbal report interview, and (e) a debriefing questionnaire. All participants met the experimenter individually in a quiet laboratory. The vocabulary pretraining, exposure, and testing phases were delivered via Microsoft PowerPoint. All participants' responses in the exposure and testing phases were screen recorded to ensure they had followed the instructions. The retrospective verbal report interviews were audio recorded. The debriefing questionnaires were administered via Qualtrics on a mobile device. The entire procedure took approximately 1 hr.

Vocabulary pretraining

The original material from Rebuschat et al. (2013, 2015) was adapted for the present study. Instructions were translated into Chinese, and the pronunciation of each artificial determiner was recorded by the same female native speaker of British English. The modality of this session was mainly written (except the audio recordings). The purpose of the pretraining was to introduce the four artificial determiners and to instruct participants on the role of distance in determiner selection. Participants were informed that they would be tested after the pretraining and that they could complete the pretraining at their own pace. Participants were first presented with a list of the four new words and their meanings in English, together with the spoken form of the words. Then, they completed a first practice task by saying aloud the missing determiner that matched the English translation on the list of four determiners. Next, they carried out a second practice by saying aloud the English translations of the new words. The practice tasks gave participants feedback by presenting correct answers after each response, and they exposed participants to 12 written repetitions of each artificial determiner. Participants were allowed to repeat this session as many times as they wished. Finally, the original quiz from Rebuschat et al. (2013, 2015) was used to test the participants on the distance meanings of the determiners on the online testing website ClassMarker. Following previous studies, participants were not allowed to continue to the exposure phase until they scored 90% or higher on the quiz. Most of them were able to do so on their first attempt, except one participant from the written–auditory group who was only able to do this on second attempt.

Exposure phase

Visual exposure

Written instructions adapted from Williams (2005, pp. 281–282) were translated into Chinese to explain that the general purpose of the experiment was to investigate whether people think differently in different languages, using the four artificial determiners. The participants were informed that the functions of the four novel determiners were similar to the English word *the*, but they also included distance meanings of objects. An example sentence ("The little boy patted *gi* tiger in the zoo"), which did not appear in the following sessions, was used to illustrate the function of this determiner in context.

Then, the participants were instructed that they would see written sentences that included the novel determiners they had learned from vocabulary pretraining. Their task was to read each sentence and judge, as quickly as possible, whether the determiner indicated far or near. They could respond by clicking F (far) or N (near) buttons located in the bottom left and right corners of the screen. They were informed that they would have only a few seconds to decide. The sentence then disappeared and the participants were required to form a mental image of the sentence, which was not to be timed. The previous sample sentence was used as an illustration (i.e., the participants would imagine that a boy was patting a tiger that was near them). A short practice session with four sentences, which did not appear in subsequent sessions, was given to all participants before the exposure.

Auditory exposure

The auditory exposure was identical to the written exposure except that the stimuli were delivered in auditory mode, which required participants to listen to the whole sentence before giving a response. This was designed differently from the written exposure to ensure participants did not skip any sentences. For each item, the response window was different from that in the written exposure.

Testing phase

The learning of the hidden animacy was measured by means of a two-alternative forced-choice (2AFC) test.

Written 2AFC test

Each sentence was displayed in the center of the computer screen (e.g., "The lady is eating pasta with ___ fork.") with two choices of artificial determiners, which differed only in animacy meaning but were the same in distance meaning (e.g., *gi* and *ro*). The two choices of determiners were located in the bottom left and right corners of the screen. The correct determiners appeared as the left and right choice with equal frequency. Participants were informed in written Chinese to choose the determiner that "seems more familiar, better, or more appropriate" (Williams, 2005, p. 282) and that this session was timed. They gave their responses by clicking one of the two determiners. Before the test, a short practice session with four sentences, which did not appear in the test, was given to all participants, but no feedback was given.

Source attributions were used as a subjective measure of awareness to determine whether participants had acquired unconscious (implicit) or conscious (conscious) structural knowledge (Dienes & Scott, 2005). After entering their 2AFC response, participants were asked to report what their decision was based on. Here, they could select one of four options: *guess*, *intuition*, *recollection*, or *rule knowledge*. Participants were instructed in written Chinese to select *guess* only when they made their judgment by purely guessing (i.e., similar to flipping a coin); to select *intuition* when they made their judgment based on a gut feeling that they were correct but could not explain why; to select *recollection* when their choice was based on recollection of a similar sentence (or part of it) in the exposure phase; to select *rule knowledge* when their decision was based on a rule that they could verbalize after the test. Subjective measures have been introduced in implicit learning research by Dienes and Scott (2005) as more sensitive measures than retrospective verbal reports. In line with Dienes and Scott (2005), judgments attributed to guessing and intuition were taken as indications of unconscious structural knowledge, while judgments attributed to recollection and rule knowledge were assumed to reflect participants' conscious structural knowledge (see also Chen et al., 2011).

Auditory 2AFC test

The auditory 2AFC test was identical to the written 2AFC test, except that it was delivered in the auditory mode and the response window for each test sentence was different from that in the written 2AFC test. The blank space in the written 2AFC test was replaced by a beep sound that lasted 30 ms.

Retrospective verbal reports

All participants were interviewed by the experimenter after the test. The interviews were conducted in the participants' native language, that is, Mandarin or Cantonese. Following Rebuschat et al. (2015), the participants were first asked how they made their choices. If they mentioned any distinction related to living or nonliving, they were asked when they had become aware of this distinction (exposure or test phase) and whether they had chosen rule knowledge as the source of their choices. If so, they were required to describe the rule or regularity. If participants had not chosen rule knowledge in the source attributions, they were asked about the other sources of their decisions. If participants did not make any reference related to animacy or if they did not use rule knowledge as the source of their judgments, they were told that there was a rule to distinguish the two near/far determiners (i.e., *gi* and *ro*; *ul* and *ne*), and they were encouraged to guess the rule. If participants still did not refer to animacy, the experimenter presented the determiner system and then asked if they had been aware of the hidden regularity at any point during the experiment. The interview questions were designed to move from implicit to explicit knowledge gradually to investigate participants' awareness of the hidden rule, so that the effect of the questions on participants' awareness of the hidden regularity would be minimized. Different from Rebuschat et al.'s (2015) study, participants were also required to report their awareness of animacy in both phases of exposure and assessment, respectively.

Following Rebuschat et al.'s (2015) study, participants' verbal reports were coded conservatively. Participants were considered to be aware of the hidden regularity even if they displayed only fleeting or partial awareness of the regularity in the verbal reports. To be considered unaware, participants could not make a mention to the animacy regularity, even after extensive and careful probing by the experimenter. The coding scheme to categorize aware and unaware participants can be found in Appendix E.

Debriefing session

After the interviews, participants completed a brief questionnaire about their age, gender, field of study, native language(s), previous experience in linguistics courses, foreign language(s) studied, contexts of instruction, and length of study.

Analysis

In order to answer RQ1, we conducted both frequentist and Bayesian one-sample *t* tests to investigate whether performance was significantly above chance and whether the data provides statistical support for learning effects in performance (the alternative hypothesis or H_1). Common in both frequentist and Bayesian one-sample *t* tests, our alternative hypothesis predicted that participants' performance would be better than by chance, and we thus adopted one-tailed *t* tests ($H_1 > H_0$) for all the analyses. Our rationale for reporting both frequentist and Bayesian statistical tests is to ensure the comparability of the current results with previous studies. However, one striking disadvantage of the frequentist approach over the Bayesian approach is that nonsignificant results can indicate either the lack of target effects or the lack of sensitivity to detect the effects (e.g., underpowered design). This problem of logical status of nonsignificant results was solved by the Bayesian analysis, which can distinguish the support for the null hypothesis from the lack of sensitivity (Keysers, Gazzola, & Wagenmakers, 2020). In addition, due to the use of *p* value in hypothesis testing in a frequentist approach, multiple comparisons would increase the rate of Type I errors, while with an adjustment of alpha level (e.g., the Bonferroni correction), the rate of Type II errors also increases. As we conducted a total of 30 one-sample *t* tests, particularly in our research context, a Bayesian approach, which does not rely on *p* value, may provide robust results (for the comparisons of statistical results between raw and adjusted *p* values of the frequentist approach and the Bayesian approach, see Appendix F).

Regarding the Bayesian analysis, we calculated a Bayes factor (BF_{10}), which refers to the relative predictive performance of H_1 to H_0, using the JASP software (JASP Team, 2020). As recommended by Keysers et al. (2020), we used a default prior distribution for *t* tests (the Cauchy distribution with a scale parameter of $r =$ .707) to calculate $BF_{10.}$ Following Keysers et al. (2020) and Kim and Godfroid (2019), we set $BF_{10} > 3.0$ as the criterion for the evidence of learning effects. The 3.0 value of BF_{10} means that the given data was predicted three times better by H_1 than H_0. We also set $BF_{10} < 0.333$ (the inversed value of 3.0) as the absence of learning effects and the BF_{10} between 3.0 and 0.333 as the absence of evidence (i.e., insufficient evidence to draw a conclusion; see Keysers et al., 2020). In addition to BF_{10}, we also reported the median and 95% credible intervals of the posterior effect size δ, which is not simply equivalent to the Cohen's *d* value, because this effect size index reflects both the prior distribution and the data observed (Keysers et al., 2020).

In order to address RQ2, we performed the generalized (logistic) linear mixed-effects modeling (GLMM) in R (R Core Team, 2018) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) to predict the accuracy of 2AFC items. We entered the binary variable representative of the correctness of answers (either correct [1] or incorrect [0]) as the dependent variable and the following variables as fixed effects (i.e., independent predictor variables): exposure (auditory vs. written), assessment (auditory vs. written), training (trained, partially trained, and untrained items), knowledge (conscious vs. unconscious structural knowledge, based on source attributions), and awareness (aware vs. unaware based on retrospective interviews). In order to test the significance of a particular predictor variable while controlling for the effects of the other predictor variables, we adopted contrast coding, as opposed to dummy coding. For the variables with two levels (exposure, assessment, and knowledge), simple contrast coding was applied (–0.5 or 0.5). Meanwhile, regarding the training variable, which had three levels, we regarded untrained items as the reference level and applied simple contrast coding to test the contrasts between untrained and partially trained items and between untrained and trained items. The same model building procedure was repeated for two different measures of awareness (knowledge vs. awareness). We entered the correctness of answers as the outcome variable and exposure, assessment, training, and either knowledge or awareness as the fixed effects predictor variables.

Results

The role of awareness in learning novel determiners (RQ1)

To determine if Chinese L2 learners could acquire the hidden regularity and to assess the role played by awareness (RQ1), we first analyzed overall performance on the 2AFC test (measure of learning), followed by closer analysis of the retrospective verbal reports and the source attributions provided by participants (measures of awareness).

Performance on the 2AFC test

To determine if learning took place, we analyzed performance on the 2AFCT test across all participants, independent of condition. As can be seen in Table 2, participants' mean accuracy on the test was 0.542 ($SD = 0.160$), which was significantly above chance, $t\ (87) = 2.45$, $p = .008$, $d = 0.26$, thus indicating that learning of the hidden regularity occurred under incidental exposure conditions. The Bayesian analysis provided moderate evidence for learning effects above chance level, $BF_{10} = 3.91$, median posterior $\delta = 0.253$, 95% confidence interval (CI) [0.057, 0.462]. Furthermore, one-sample t tests indicated that participants performed significantly above chance on each test-item type: for the trained items, $t\ (87) = 4.04$, $p < .001$, $d = 0.43$; for the partially trained items, $t\ (87) = 2.41$, $p = .009$, $d = 0.26$; for the untrained items, $t\ (87) = 2.27$, $p = .013$, $d = 0.24$. Using Plonsky and Oswald's (2014) guidelines on the interpretation of Cohen's d values, all these effects are considered small. The Bayesian analysis showed extremely strong support for above chance learning effects on trained items ($BF_{10} = 340.89$, median posterior $\delta = 0.417$, 95% CI = [0.202, 0.634]) and moderate support for partially trained items ($BF_{10} = 3.58$, median posterior $\delta = 0.249$, 95% CI = [0.054, 0.458]). The results of the Bayesian analysis regarding untrained items were inconclusive ($BF_{10} = 2.65$, median posterior $\delta = 0.233$, 95% CI = [0.025, 0.443]).

Subjective measures of awareness

The analyses of the item-by-item source attributions provided during the completion of the 2AFC test showed that participants (across the four conditions) scored 0.576 ($SD = 0.494$) when basing their decisions on the explicit response categories (recollection and rule knowledge) and 0.550 ($SD = 0.498$) when basing decisions on the implicit response categories (guessing and intuition). The mean accuracy was significantly above chance in either case: explicit categories, $t\ (1281) = 5.47$, $p < .001$ $d = 0.15$; implicit categories, $t\ (1776) = 4.26$, $p < .001$ $d = 0.10$. The Bayesian analysis provided extremely strong support for above chance learning effects in both implicit and explicit response categories (cf. Table 3).

Table 2. Frequentist and Bayesian analysis of overall response accuracy and response accuracy for trained, partially trained and untrained items

	Descriptive statistics				Frequentist one-sample t test			Bayesian one-sample t test		
Condition	N	M	SD	95% CI	t	p	d	BF_{10}	Posterior δ	95% CI
Overall	88	0.542	0.160	[0.508, 0.576]	2.451	.008	0.261	3.91	.253	[0.057, 0.462]
Trained item	88	0.583	0.193	[0.542, 0.624]	4.042	<.001	0.431	340.89	.417	[0.202, 0.634]
Partially trained item	88	0.548	0.187	[0.508, 0.588]	2.411	.009	0.257	3.58	.249	[0.054, 0.458]
Untrained item	88	0.547	0.193	[0.506, 0.588]	2.272	.013	0.242	2.65	.233	[0.025, 0.443]

Table 3. Frequentist and Bayesian analysis of overall response accuracy and response accuracy for items based on subjective measures

	Descriptive statistics				Frequentist one-sample t test			Bayesian one-sample t test		
Condition	N	M	SD	95% CI	t	p	d	BF_{10}	Posterior δ	95% CI
Explicit knowledge	1282	0.576	0.494	[0.549, 0.603]	5.479	< 001	0.153	169102.18	0.153	[0.098, 0.208]
Implicit knowledge	1777	0.550	0.498	[0.527, 0.574]	4.267	<.001	0.101	451.67	0.101	[0.054, 0.148]

Retrospective verbal reports

Our analysis of the retrospective verbal reports revealed that 52 participants reported knowledge of the animacy regularity, whereas 36 participants did not. Based on their verbal reports (or lack thereof), we categorized the former participants as "aware" and the latter as "unaware" of the hidden regularity (see Table 2, and Appendix E for coding scheme). The aware participants correctly identified 0.562 ($SD = 0.194$) of the test items and the unaware participants 0.512 ($SD = 0.083$). Analyses showed that aware participants scored significantly above chance in the 2AFC test, $t\ (51) = 2.30$, $p = .013$, $d = 0.32$, and that unaware participants did not, $t\ (35) = 0.89$, $p = .190$, $d = 0.15$. This suggests that, according to verbal reports, learning was limited to those participants who displayed at least some awareness, if only minimal, of the hidden regularity in their retrospective verbal reports. The Bayesian analysis showed moderate evidence for above chance learning effects for aware participants ($\mathrm{BF}_{10} = 3.32$, median posterior $\delta = 0.304$, 95% CI = [0.058, 0.577]), whereas the results of the Bayesian analysis regarding overall learning effects in the unaware group were inconclusive ($\mathrm{BF}_{10} = 0.41$, median posterior $\delta = 0.176$, 95% CI = [0.011, 0.467]).

As a next step, we analyzed the performance of aware and unaware participants across the three test-item types. The results show that aware participants performed above chance only on trained items, t (51) = 3.82, $p < .001$, $d = 0.53$, and that there was an extremely strong evidence for learning effects using the Bayesian analysis ($\text{BF}_{10} = 141.30$, median posterior $\delta = 0.504$, 95% CI = [0.220, 0.794]). In contrast, the unaware participants did not perform above chance on any test-item types. The Bayesian analysis indicates that there is insufficient evidence for the validity of either the null or the alternative hypothesis for any of the item types in the unaware group and for the partially trained and untrained items in the aware group (see Table 4). In other words, learning appeared limited to those participants who displayed at least some conscious knowledge of the hidden regularity in the verbal reports. Even in those cases, the learning effect was driven by above chance performance on trained items, and there was no sufficient evidence of generalization.

The effect of exposure and assessment modality (RQ2)

To examine whether the modality used in the exposure and testing phases affected the acquisition of novel determiners, we analyzed performance on the 2AFC test across groups. We found that the AA group scored highest ($M = 0.614$, $SD = 0.155$), followed by the AV group ($M = 0.588$, $SD = 0.146$), the VA group ($M = 0.524$, $SD = 0.100$) and the VV group ($M = 0.471$, $SD = 0.197$). One-sample t tests on the groups' mean accuracy scores revealed that only the AA group performed above chance, t (21) = 3.45, $p = .001$, $d = 0.74$. Using Plonsky and Oswald's (2014) guidelines on the interpretation of Cohen's d values, this effect is considered moderate, and the Bayesian analysis also showed very strong evidence for the existence of the overall learning effect in this group. Regarding the AV group, the results of one-sample t tests in frequentist and Bayesian approaches did not concur, t (21) = 1.87, $p = .038$, $d = 0.40$, $\text{BF}_{10} = 1.85$, median posterior $\delta = .364$, 95% CI = [0.043, 0.780]). Considering the robustness of Bayesian analysis, we regarded the leaning gains in the AV group as the evidence of absence rather than the lack of learning effects. (i.e., inconclusive evidence). The other groups' performance did not differ significantly from chance: the VA group, t (21) = 1.12, $p = .134$, $d = 0.24$, the VV group, t (21) = –0.69, $p = .751$, $d = -0.15$. The Bayesian analysis showed inconclusive evidence for overall learning effects in the VA groups and revealed moderate evidence for the null hypothesis in the VV group ($\text{BF}_{10} = 0.14$, median posterior $\delta = 0.096$, 95% CI = [0.004, 0.367]). We then investigated whether performance differed across test-item types. One-sample t tests showed that only the AV group performed above chance on all test-item types, and the Bayesian analyses suggested moderate evidence for these learning effects. The AA group performed above chance on trained test items and partially trained test items, and the evidence for the significance of these learning effects was extremely strong for trained items ($\text{BF}_{10} = 136.65$, median posterior $\delta = 0.804$, 95% CI = [0.322, 1.305]) and strong for untrained items ($\text{BF}_{10} = 10.36$, median posterior $\delta = 0.548$, 95% CI = [0.133, 1.001]). The other groups did not score above chance on any test-item types. In the VA group the Bayesian analysis showed an absence of evidence for for the null or the alternative hypothesis for all item types, whereas in the VV group there was sufficiently strong evidence for the null hypothesis (see Table 5).

Table 4. Frequentist and Bayesian analysis of overall response accuracy and response accuracy for items based on retrospective measures

	Descriptive statistics				Frequentist one-sample *t* test			Bayesian one-sample *t* test		
Condition	*N*	*M*	*SD*	95% CI	*t*	*p*	*d*	BF_{10}	Posterior δ	95% CI
Aware (overall)	52	0.562	0.194	[0.508, 0.616]	2.304	.013	0.320	3.32	0.304	[0.058, 0.577]
Trained	52	0.614	0.215	[0.554, 0.674]	3.823	<.001	0.530	141.30	0.504	[0.220, 0.794]
Partially trained	52	0.560	0.221	[0.499, 0.622]	1.963	.028	0.272	1.72	0.262	[0.037, 0.530]
Untrained	52	0.559	0.225	[0.496, 0622]	1.892	.032	0.262	1.52	0.253	[0.034, 0.530]
Unaware (overall)	36	0.512	0.083	[0.484, 0.541]	0.889	.190	0.148	0.41	0.176	[0.011, 0.467]
Trained	36	0.538	0.146	[0.489, 0.587]	1.568	.063	0.261	1.02	0.254	[0.025, 0.569]
Partially trained	36	0.530	0.123	[0.489, 0.572]	1.490	.073	0.248	0.91	0.244	[0.023, .0556]
Untrained	36	0.529	0.133	[0.484,0.574]	1.291	.103	0.215	0.68	0.220	[0.018, 0.526]

To further investigate variances in each group's performance across test-item types, we constructed the parsimonious models of GLMM in two steps (for the full model, see Appendix G). In the first step of the analysis, we examined whether the model fit to the current data was improved by adding the fixed-effects predictors and their interactions, while keeping the random effects constant. In order to compare the simpler models with the more complex ones, we used the likelihood ratio test (LRT; Baayen, 2008) and evaluated whether the inclusion of fixed-effects predictors was justified. To this end, we started with a minimal model of the log odds response accuracy (Model 1), which included the random intercepts of participants and items. We compared the minimal model with the models with one of the fixed-effect factors: exposure (Model 2), assessment (Model 3), training (Model 4), or knowledge type (Model 5). The LRT revealed that only Model 2 improved the (exposure) model fit to the data, compared to the minimal model, indicating that only exposure was the significant predictor for response accuracy, $\chi^2(1) = 7.842$, $p = .0051$. Then, although we added each of the remaining three fixed-effects factors to Model 2, none of them further improved the model fit. In addition, there was no significant interaction effect between exposure and those nonsignificant fixed-effects factors. Therefore, we included one fixed-effects factor for the final model.

In the second step of the analysis, we examined whether to include the random slopes of participants and items in the slope of the fixed-effect factor (exposure). Note that we decided to include both random intercepts of participants and items considering the hierarchical structure of our data. The LRT tests confirmed that neither of random-slope of participants, $\chi^2(2) = 0.056$, $p = .972$, and items, $\chi^2(2) = 1.986$, $p = .370$, improved the model fit. Therefore, our finalized model was specified as follows:

$$Response\,Accuracy \sim Exposure + (1|Participant) + (1|Item).$$

Another similar model following the same procedures was run where knowledge was replaced by awareness (aware vs. unaware, based on retrospective verbal reports). The full model is reported in Appendix G. The full model does not include interactions and random slopes because that complex model made the estimation very unstable and failed to converge. Both of the GLMMs predicting the accuracy of 2AFC scores included a significant main effect of exposure. There were no other significant main effects or significant interactions between the predictor variables. None of the variables, except for exposure, improved the model fit compared to the baseline model, which included only random intercepts of participant and item. Therefore, we only report the final model, which includes exposure as the main effect in Table 6. The odds ratios for this model predicted participants to be 1.52 (1/0.66) times more likely to respond correctly in the auditory exposure condition than in the written exposure condition. Following Plonsky and Oswald (2014), we interpret R^2 with a value lower than .16 as small effects. The main effects model had a marginal R^2 of .012 and a conditional R^2 of .111, which indicates that exposure condition alone as well as the exposure condition and the random effects explain a relatively small amount of variance in the accuracy of responses.

Table 5. Frequentist and Bayesian analysis of overall response accuracy and response accuracy for items based in different exposure and assessment conditions

	Descriptive statistics				Frequentist one-sample *t* test			Bayesian one-sample *t* test		
Condition	*N*	*M*	*SD*	95% CI	*t*	*p*	*d*	BF_{10}	Posterior δ	95% CI
AV (Total)	22	0.558	0.146	[0.493, .0623]	1.866	.038	0.398	1.85	0.364	[0.043, 0.780]
Trained	22	0.597	0.185	[0.515, 0.679]	2.470	.011	0.527	5.07	0.474	[0.089, 0.913]
Partially trained	22	0.584	0.171	[0.509, 0.660]	2.315	.015	0.494	3.87	0.445	[0.075, 0.878]
Untrained	22	0.597	0.199	[0.508, 0.685]	2.277	.017	0.486	3.63	0.438	[0.071, 0.870]
VA (Total)	22	0.524	0.100	[0.480, 0.568]	1.124	.137	0.240	0.66	0.248	[0.018, 0.630]
Trained	22	0.553	0.148	[0.487, 0.619]	1.665	.055	0.355	1.37	0.330	[0.034, 0.738]
Partially trained	22	0.525	0.135	[0.465, 0.585]	0.869	.197	0.185	0.49	0.215	[0.013, 0.583]
Untrained	22	0.519	0.155	[0.450, 0.588]	0.574	.286	0.122	0.36	0.182	[0.010, 0.533]
AA (Total)	22	0.614	0.155	[0.545, 0.682]	3.449	.001	0.735	33.40	0.666	[0.217, 1.141]
Trained	22	0.669	0.192	[0.584, 0.754]	4.128	<.001	0.880	136.65	0.804	[0.322, 1.305]
Partially trained	22	0.612	0.184	[0.531, 0.694]	2.856	.005	0.609	10.36	0.548	[0.133, 1.001]
Untrained	22	0.568	0.172	[0.492, .0644]	1.861	.038	0.397	1.84	0.363	[0.043, 0.779]
VV (Total)	22	0.471	0.197	[0.384, 0.558]	−0.691	.751	−0.147	0.14	0.096	[0.004, 0.367]
Trained	22	0.513	0.216	[0.417, 0.609]	0.286	.389	0.061	0.28	0.155	[0.008, 0.488]
Partially trained	22	0.471	0.226	[0.371, 0.571]	−0.609	.726	−0.130	0.15	0.099	[0.004, .0375]
Untrained	22	0.503	0.234	[0.399, 0.606]	0.055	.478	0.012	0.23	0.137	[0.006, 0.455]

Discussion

Establishing form–meaning connections under incidental exposure conditions

In our first research question, we examined the acquisition of novel form–meaning connections by adult learners under incidental exposure conditions. In line with our predictions and previous research (e.g., Chen et al., 2011; Hama & Leow, 2010; Leung & Williams, 2014; Rebuschat et al., 2013, 2015; Williams, 2005), the findings confirm that adult learners can learn a hidden regularity that governs determiner selection in a miniature linguistic system. However, the overall learning effect was found to be relatively small, and the Bayesian analysis only showed moderate support for the alternative hypothesis. This could be explained by the fact that the determiners were processed at a relatively low level of depth of cognitive processing (cf. Leow, 2015). In our study, a novel structure was embedded in L2 sentences, and even though these sentences were simple and contained familiar vocabulary, our participants might not have had sufficient attentional resources for processing and encoding the new meaning of the determiners. The small learning effects might also be explained by the characteristics of the participants' first language (Mandarin and Cantonese) in which there are no classifiers that specifically distinguish between animate and inanimate nouns. Instead, the elaborate classifier system denotes specific characteristics of objects such as flatness or length. The participants' learned attention to these other types of semantic cues and subsequent processing bias that originates from their native language (cf. Ellis & Sagarra, 2010) might also explain the small and predominantly exemplar-based learning effects, for the learning of which we found very strong evidence using Bayesian analysis (cf. also Leung &Williams, 2014, for a similar argument).

In terms of our awareness measures, the analyses of the retrospective verbal reports showed that learning was limited to aware participants who were able to verbalize knowledge associated with the hidden regularity at the end of the experiment. The effect size of learning for the aware group was small, and the alternative hypothesis was only moderately supported by the Bayesian analysis. Significant learning was only observable for those test items that were repeated from the training phase (*trained items*) in the aware group. The very strong support for the above chance learning effects in the aware group on trained items from the Bayesian analysis also highlights that most learning in the experiment took place through memorization and was exemplar based. These results are consistent with those of Faretta-Stutenberg and Morgan-Short (2011), Hama and Leow (2010), and Rebuschat et al. (2015), but contradict the findings of Williams (2005), who showed that unaware participants can perform above chance level. Our Bayesian analysis also indicates that in the unaware group there is insufficient evidence to either confirm or reject the alternative hypothesis. The small and unstable effects across studies might be due to sensitivity to input frequency, insufficient statistical power, and variations in the sample and research methodology.

The analysis of participants' source attributions indicated the presence of both implicit (unconscious) and explicit (conscious) knowledge at test, as evidenced by above chance performance when 2AFC decisions were based on implicit categories (guessing or intuition) and explicit categories (recollection or rule knowledge). This confirms the observations made in previous studies that investigated the effect

Table 6. Summary of the final model

Fixed effects	Estimate	*SE*	Odds ratios	95% CI of odds ratios Lower	95% CI of odds ratios Upper	*Z* value	*p*
(Intercept)	0.27	0.08	1.31	1.12	1.52	3.46	<.001
Exposure	−0.42	0.15	0.66	0.49	0.88	−2.85	0.004
Random effects (intercepts)	Variance	*SD*					
Participant	0.35	0.59					
Item	0.02	0.13					
Information criterion	Estimate						
LogLikelihood	−2032.7						
DIC	4065.4						
AIC	4071.4						
BIC	4089.4						
R2	Estimate						
Marginal	0.012						
Conditional	0.111						

Note: Auditory exposure is the reference level for exposure.

of incidental exposure on the acquisition of form–meaning connections (e.g., Chen et al., 2011; Monaghan, Schoetensack, & Rebuschat, 2019; Rebuschat et al., 2015), L2 vocabulary (e.g., Hamrick & Rebuschat, 2012, 2014), L2 morphology (e.g., Grey et al., 2014), and L2 syntax (e.g., Kim & Godfroid, 2019) in adult learners. The results also confirm that reliance on retrospective verbal reports might not be sufficient to detect the presence of unconscious knowledge as these tests might not be sensitive enough (see Rebuschat et al., 2015, for discussion). The presence of implicit knowledge at test does not imply that learning necessarily occurred without awareness, given that subjective measures assess awareness of the learning product, not the learning process, but it suggests that the required level of awareness can be relatively low, for instance, at the level of Schmidt's (1990) noticing. It is also important to acknowledge that subjective measures of awareness (like most behavioral measures of awareness) are not without problems (for a thorough investigation of the limitations of source attributions, see, for example, Sachs, Hamrick, McCormick, & Leow, 2020). Clearly, more research is necessary to assess the internal validity of awareness measures.

The role of modality in incidental language learning

Our study also sought to investigate the effect of modality on the incidental learning of novel determiners in the exposure and assessment phases, respectively. The results of GLMM showed a significant benefit with a small effect size for auditory exposure, but the modality of assessment did not influence performance in the testing phase of the study. No significant main effects for awareness, knowledge-type, or training were detected either, and there was no interaction among any of these variables. Therefore, the final model of GLMM in which the item-level subjective measure of knowledge type was included as a predictor confirmed the results of both the frequentist and the Bayesian one-sample *t* tests. However, the GLMM with the retrospective measure of awareness as the predictor variable was not fully in line with the results of the frequentist or the Bayesian one-sample *t* tests, as it showed no effects of awareness or item-type, whereas above chance learning effects were observed for aware participants on trained items and in the auditory exposure conditions for trained and partially trained items. As above chance learning effects were relatively small in all conditions, except for the auditory exposure and auditory assessment group, it is not unexpected that the GLMM, which took into account random intercept variations at participant and item levels, did not detect these main effects.

Overall, our findings suggest that the success of incidental learning might depend on the modality of exposure. The beneficial effects of auditory input seem to apply in our study in comparison with Kim and Godfroid's (2019) research because our participants did not engage in additional processing such as repetition of the sentences, and they did not receive feedback. In this respect, our design is similar to artificial grammar learning studies in cognitive psychology, which found that the auditory modality is superior to the written modality in ISL (e.g., Conway & Christiansen, 2005, 2009; Drewnowski & Murdock, 1980; Qi, Araujo, Georgan, Gabrieli, & Arciuli 2018). Our research demonstrates that the incidental learning of novel form–meaning connections is assisted by auditory input regardless of the level of awareness or whether judgments were made based on conscious or unconscious knowledge. Furthermore, the auditory facilitation effect can be observed irrespective of whether items were trained, partially trained, or untrained. Therefore, our results might indicate that sequences are more successfully remembered in the auditory mode, which might support extraction of statistical regularities more effectively than visual input. This finding seems to provide indirect support for the nonunitary nature of statistical learning ability by showing superior auditory modality effects for linguistic stimuli (cf. Frost, Armstrong, & Christiansen, 2019). Our findings showing the superiority of the auditory exposure mode in incidental learning conditions are also in line with recent research in the field of first language literacy development, which has shown that auditory statistical learning ability is a better predictor of word and sentence reading fluency than visual statistical learning ability (Qi et al., 2018).

As regards the lack of a significant effect of the assessment modality, the findings seem to contrast with previous studies that found that learners' accuracy is higher in written forced-choice tasks than auditory ones (e.g., Johnson, 1992; Kim & Nam, 2017; Murphy, 1997; Shui et al., 2018; Spada et al., 2015). However, our results support the conclusions of Plonsky et al.'s (2019) meta-analysis, which revealed unstable effects of modality in grammaticality judgment performance. As discussed above, potential differences in the length of response window during the tasks in previous studies may explain the contradictory findings. In our research, we did not detect evidence of transfer appropriate processing either as no interaction between the input modality and assessment mode was found. Nonetheless, the frequentist and Bayesian one-sample *t* tests indicated that the strongest evidence for the alternate hypothesis can be detected in the auditory exposure and auditory assessment group, which might indicate some advantage for the auditory mode both in exposure and assessment.

Conclusion

The present study investigated the acquisition of form–meaning connections by adult participants under incidental exposure conditions. As an original contribution to the study of incidental language learning, we also examined the effect of modality in both the exposure and the test phases. The findings showed small, but significant, learning effects among participants supporting numerous previous studies that employed the paradigm developed by Williams (2005). However, in light of the Bayesian analysis, we interpret this finding cautiously as the evidence for the learning effect was overall only moderate and was extremely strong only for trained items (in contrast to Rebuschat et al., 2015). We demonstrate that incidental exposure can result in implicit and explicit knowledge but argue that some level of awareness is facilitative for the establishment of accurate form–meaning relationships, especially when a novel grammatical construction is embedded in a sentence-level L2 context and when the participants' native language may contribute to a processing bias.

It also needs to be noted that statistical analyses that group participants into aware and unaware categories based on retrospective reports have much lower statistical power due to the small sample size than subjective analyses that give information on the basis of participants' accuracy decisions for each item. Our Bayesian analyses indicated that there is no sufficient empirical basis for either the null or the alternative hypotheses except for the aware participants' overall scores and their scores on trained items. Therefore it would be important to conduct larger scale studies and a meta-analysis of existing research that would compare learning effects based on retrospective reports of awareness.

In terms of the effects of modality, our findings extend the results of previous studies that have demonstrated the benefits of the auditory exposure modality over the written modality (e.g., Conway & Christiansen, 2005, 2009; Drewnowski & Murdock, 1980; Qi et al., 2018). The lack of effect of the assessment modality is in line with the conclusions of a recent meta-analysis by Plonsky et al. (2019), who showed that, in forced-choice tasks, modality effects are small and prone to variation based on the research methodology.

One of the limitations of our study is that we did not employ a first language-speaking control group as a baseline for the performance of the experimental groups. Future research with this additional group and with participants whose first language denotes animacy in their classifier or determiner system could reveal to what extent the L2 sentence context and the learned attention of the Mandarin- and Cantonese-speaking participants influenced our findings. In our study, we did not assess participants' productive knowledge or the longer term duration of learning effects. Follow-up studies could be conducted that examine how incidental exposure conditions can promote the active recall of form–meaning links and how stable the representations of these links are over time. More research is also necessary to investigate the role of individual differences in incidental exposure conditions across different modalities.

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 899–912. doi: 10.1037/0278-7393.21.4.899

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*. doi: 10.18637/jss.v067.i01

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology, 55A*, 371–389. doi: 10.1080/02724980143000307

Chan, R. K., & Leung, J. H. (2014). Implicit learning of L2 word stress regularities. *Second Language Research*, **30**(4), 463–484. doi: 10.1177/0267658313510169

Chen, W., Guo, X., Tang, J., Zhu, L., Yang, Z., & Dienes, Z. (2011). Unconscious structural knowledge of form-meaning connections. *Consciousness and Cognition*, **20**, 1751–1760. doi: 10.1016/j.conco

Cintrón-Valentín, M. C., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers of Psychology*, **7**, 1–21. doi: 10.3389/fpsyg.2016.01284

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory and Cognition*, **28**, 529–538. doi: 10.3758/BF03201243

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**, 24–39. doi: 10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus specific representations. *Psychological Science*, **17**, 905–912. doi: 10.1111/j.1467-9280.2006.01801.x

Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, **21**, 561–580. doi: 10.1080/09541440802097951

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech, *Journal of the Acoustical Society of America*, **88**, 101–112. doi: 10.1121/1.399955

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies*, **11**, 25–45.

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition*. Berlin: Muton de Gruyter.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research*, **69**, 338–351. doi: 10.1007/s00426-004-0208-3

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory*, **6**, 319–332. doi: 10.1037/0278-7393.6.3.319

Ellis, N. C. (1994). Vocabulary acquisition: The implicit ins and outs of explicit cognitive mediation. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages*. (pp. 211–282). London: Academic Press.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, **27**, 141–172. doi: 10.1017/S0272263105050096

Ellis, N. C., & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning*, **60**, 85–108. doi: 10.1111/j.1467-9922.2010.00602.x

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation timing have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, **64**, 1021–1040. doi: 10.1080/17470218.2010.538972

Faretta-Stutenberg, M., & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams (2005). In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. Prieto Botana, & E. Rhoades (Eds.), *Selected proceedings of the 2010 Second Language Research Forum* (pp. 18–28). Somerville, MA: Cascadilla Proceedings Project.

Frost, R., Armstrong, B. C., & Christiansen, M. C. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, **145**, 1128–1153. doi: 10.1037/bul0000210

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, **19**, 117–125. doi: 10.1016/j.tics.2014.12.010

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S. M., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, **37**, 269–297. doi: 10.1017/S0272263114000850

Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, **36**(4), 611–645. doi: 10.1017/S0272263113000727

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition*, **32**, 465–491. doi: 10.1017/S0272263110000045

Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning? In P. Rebuschat & J. N. Williams (Eds.), *Statistical learning and language acquisition*. (pp. 365–382). Berlin: Mouton de Gruyter.

Hamrick, P., & Rebuschat, P. (2014). Frequency effects, learning conditions, and the development of implicit and explicit lexical knowledge. In J. Connor-Linton & L. Amoroso (Eds.), *Measured language: quantitative approaches to acquisition, assessment, processing and variation*. (pp. 125–139). Washington, DC: Georgetown University Press.

Hulstijn, J. H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25–46). Amsterdam: Benjamins.

JASP Team. (2020). JASP (Version 0.13.1) [Computer software]. Retrieved from https://jasp-stats.org/

Johnson, J. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, **42**, 217–248. doi: 10.1111/j.1467-1770.1992.tb00708.x

Kerz, E., Wiechmann, D., & Riedel, F. B. (2017). Implicit learning in the crowd: Investigating the role of awareness in the acquisition of L2 knowledge. *Studies in Second Language Acquisition*, **39**, 711–734. doi: 10.1017/S027226311700002X

Keysers, C., Gazzola, V., & Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, **23**, 788–799. doi: 10.1038/s41593-020-0660-4

Kim, J., & Nam, H. (2017). Measures of implicit knowledge revisited: Processing modes, time pressure, and modality. *Studies in Second Language Acquisition*, **39**, 431–457. doi: 10.1017/S0272263115000510

Kim, K. M., & Godfroid, A. (2019). Should we listen or read? Modality effects in implicit and explicit knowledge. *Modern Language Journal*, **103**, 648–664. doi: 10.1111/modl.12583

Leow, R. P. (1993). To simplify or not to simplify: A look at intake. *Studies in Second Language Acquisition*, **15**, 333–355. doi: 10.1017/S0272263100012146

Leow, R. P. (1995). Modality and intake in second language acquisition. *Studies in Second Language Acquisition*, **17**, 79–89. doi: 10.1017/S0272263100013784

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, **47**, 467–505. doi: 10.1111/0023-8333.00017

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, **22**, 557–584.

Leow, R. P. (2015). Implicit learning in SLA: Of processes and products. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 47–66). Amsterdam: Benjamins.

Leow, R. P. (2018). Explicit learning and depth of processing in the instructed setting: Theory, research, and practice. *Studies in English Education*, **23**(4), 769–801. doi: 10.22275/SEE.23.4.01

Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually-derived meanings. *Studies in Second Language Acquisition*, **33**, 33–55. doi: 10.1017/S0272263110000525

Leung, J. H. C., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, **62**, 634–662. doi: 10.1111/j.1467-9922.2011.00637.x

Leung, J. H., & Williams, J. N. (2014). Crosslinguistic differences in implicit language learning. *Studies in second language acquisition*, **36**, 733–755. doi: 1017/S0272263114000333

Monaghan, P., Schoetensack, C., & Rebuschat, P. (2019). A single paradigm for implicit and statistical learning. *Topics in Cognitive Science*, **11**(3), 536–554. doi: 10.1111/tops.12439

Murphy, V. A. (1997). The effect of modality on a grammaticality judgment task. *Second Language Research*, **13**, 34–65. doi: 10.1191/026765897671676818

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

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*. Advance online publication. doi: 10.1177/0267658319828413

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, **64**, 878–912. doi: 10.1111/lang.12079

Qi, Z., Araujo, Y. S., Georgan, W. C., Gabrieli, J. D. E., & Arciuli, J. (2018). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, **23**, 111–115. doi: 10.1080/10888438.2018.1485680.

Raviv, L., & Arnon, I. (2017). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developemental Science*, **21**, e12593. doi: 10.1111/desc.12593

R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, **6**, 317–327. doi: 10.1016/S0022-5371(67)80149-X

Rebuschat, P., Hamrick, P., Riestenberg, K., Sachs, R., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, **37**, 299–334. doi: 10.1017/S0272263115000145

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. Bergsleithner, S. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honor of Richard Schmidt* (pp. 249–269). Honolulu, HI: University of Hawai'i Press, National Foreign Language Resource Center.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, **33**(4), 829–856. doi: 10.1017/S0142716411000580

Roediger, I. H. L., & Guynn, M. J. (1996). Retrieval processes. In E. L. Bjork and R. A. Bjork (Eds.), *Memory* (pp. 197–236). New York: Academic Press.

Sachs, R., Hamrick, P., McCormick, T. J., & Leow, R. P. (2020). Exploring the veridaucality and reactivity of subjective measures of awareness: Is a "guess" really a guess? *Studies in Second Language Acquisition*. Advance online publication.

Schacter, D. L. (1989). On the relation between memory and consciousness: Dissociable interactions and conscious experience. In H. L. Roediger III & F. I. M. Craik (Eds.), (pp. 355–389). Lawrence Erlbaum Associates, Inc.

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

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1–63). Honolulu : University of Hawai'i Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010, Singapore, December 2–4* (pp. 721–737). Singapore: National University of Singapore, Centre for Language Studies.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, **17**, 367–395.

Shiu, L., Yalçın, S., & Spada, N. (2018). Exploring second language learners' grammaticality judgment performance in relation to task design features. *System*, **72**, 215–225. doi: 10.1016/j.system.2017.12.004

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, **81**, 105–120. doi: 10.1016/j.jml.2015.02.001

Silva, S., Folia, V., Inácio, F., Castro, S. L., & Petersson, K. M. (2018). Modality effects in implicit artificial grammar learning: An EEG study. *Brain Research*, **1687**, 50–59. doi: 10.1016/j.brainres.2018.02.020

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, **65**, 723–751. doi: 10.1111/lang.12129

Thorndike, E. L., & Rock, R. T. (1934). Learning without awareness of what is being learned or intent to learn it. *Journal of Experimental Psychology*, **17**, 1–19. doi: 10.1037/h0073815

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

Vafaee, P., Suzuki, Y., & Kachinske, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, **39**, 59–95. doi: 10.1017/S0272263115000455

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, **27**, 269–304. doi: 10.1017/S0272263105050138

Wong, W. (2001). Modality and attention to meaning and form in the input. *Studies in Second Language Acquisition*, **23**, 345–368. doi: 10.1017/S0272263101003023

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*, **38**, 553–586. doi: 10.1017/S027226311500025X

Appendix A
Stimuli used in the study

Training practice items

1. The farmer caught gi chicken near him.
2. The old man saw ul dolphins in the sea.
3. The students used ro pencils to draw pictures.
4. After the storm, ne roof was damaged by a tree.

Training items (Set 1)

1. I tried to play dead while gi bear saw me.
2. Yesterday the cook was bitten by gi rats in the kitchen.
3. The child screamed although ul bee was on the other side of the playground.
4. The tiger wanted to eat ul birds up in the tree.
5. I carefully packed my nephew’s present in ro box with newspaper.
6. After the waitress dropped ro cups on the floor she had to leave her job.
7. I couldn’t read the title of ne book that was on the top shelf.
8. From the kitchen I heard the sound of ne plate crashing to the floor.
9. The owner gave a snack to gi dog for his trick.
10. I held gi pig carefully so he would not escape.
11. On the journey to look at wild animals, the children were excited that gi lions were beside the car.
12. The teenagers pushed over gi cows in the field.
13. The old lady pointed at ul cat stuck at the top of the tree.
14. I could see ul fly on the television screen from the back of the classroom.
15. The sound of ul monkeys in the top of the tree was extremely loud.
16. I was frightened to see ul snakes on the other side of the garden.
17. The detective looked at ro picture carefully for information about the crime.

(*Continued*)

(*Continued*)

18. The workers carefully carried ro sofa up the stairs.
19. The children all sat down at ro tables to play board games.
20. The thief put ro televisions into his van.
21. I stood back to see how well ne cushion matched the couch.
22. I could not sleep because of the sound of ne clocks downstairs.
23. We hoped that ne cakes would be delivered in time for the party.
24. The thieves searched the living room for ne vases kept in the study.

Training items (Set 2)

1. The park officer told us that gi bears were not dangerous.
2. While I was sitting in the kitchen gi rat ran between my feet.
3. The paranoid man felt certain that there are ul bees in the box up in the tree.
4. The scientist tried to take a picture of ul bird from the jungle floor.
5. I stood on top of ro boxes to reach the light bulb.
6. I pushed ro cup accidentally and the coffee spilled on my book.
7. I asked the library staff to fetch ne books from the information desk.
8. The children threw baseballs at ne plates to win prizes at the fair.
9. We were attacked by gi dogs as we walked down the path.
10. The farmer could not stand the smell of gi pigs when he cleaned their place.
11. I was terrified when I turned around and saw gi lion right behind me.
12. The vet carefully checked gi cow at the farm.
13. In the night we heard ul cats fighting in the road outside.
14. We heard the sound of ul flies coming together in a large group around the dead animal.
15. The children threw sticks at ul monkey in the tree.
16. We could see ul snake eating a chicken from behind the trees.
17. The art museum staff checked ro pictures for signs of damage.
18. The guests sat down on ro sofas in the living room.
19. I spent an hour cleaning ro table before the dinner party.
20. The girl went to bed later than usual watching ro television by her bed.
21. I wanted to buy ne cushions from the store down the street.
22. I looked up at ne clock on the church and realized that I was late.
23. In the cafe I asked the waiter to get me ne cake.
24. I wanted to see ne vase at the far end of the museum.

Test practice items

1. The girl was scared because of gi spiders over on the wall.
2. Ul giraffe is walking fast outside the hotel window.

(*Continued*)

(*Continued*)

3. The lady is eating pasta with ro fork.

4. There are people in ne boats in the sea far from the island.

Test items

1. The babysitter poured juice into ro cups for the children.

2. I heard the sound of ul cat from the old building on the farm.

3. The boy played with gi monkeys in the rainforest.

4. I pushed ro television to one side to make room for the new speakers.

5. I was terrified when I heard ul bears making a loud noise in the distance.

6. As I was passing I knocked ro vase onto the floor.

7. I ordered ne book from the library.

8. The student was happy that gi goat was awake all night to stay with her so that she is not alone.

9. When I had a backache I asked my wife to fetch ne cushions from the bedroom.

10. On the bus, the passengers saw ul insects eating leaves on a rock.

11. The waitress was told to go and clear ne table at the end of the dinner.

12. The photographer took a very beautiful photo of ul camels across the sand hills.

13. The child wasn’t able to see ne box of biscuits hidden up on top of the fridge.

14. The farmer was kicked by gi cow when he tried to milk it.

15. The tourists had trouble reading ne clocks on the towers across the river.

16. The circus performer covered herself with gi snakes so that we could take a photo.

17. The hunter noticed ul lion behind the tree from a far distance.

18. The teacher made the bad students sit in ro desks right in front of her.

19. I was amazed when gi bird ate from my hand.

20. The children washed ro plates clean after the meal.

21. The smell of ul pig coming from the farm next door was disgusting.

22. I was surprised when ro phone rang in my hand.

23. From my apartment window, I could see ne candles lit in a nearby park.

24. The chef tried to hit gi rats with a stick.

25. She asked her husband to turn off ne lamp in the other room.

26. I couldn’t see ul horses in the field without glasses.

27. After the party we stored ro cakes in the fridge.

28. The child held gi rabbit at the petting zoo.

29. The researchers studied ul bees from a safe distance.

30. When I was out for a walk I touched gi dog and it bit me.

31. In the gallery we all admired ne pictures from the other side of the room.

32. I spent the night on ro sofa and let my guests sleep in the beds.

(*Continued*)

(*Continued*)

33. The man put away ro spoons in the drawer.
34. The kids at the zoo fed gi elephants peanuts right from their hands.
35. The man had to look carefully to see ul flies over on the wall.
36. The rich woman asked the pool boy to get ne towel from the chair.

Appendix B
Nouns phrases (n = 48) employed in the exposure phase

Training Set 1	Animate		Inanimate	
	Near	Far	Near	Far
	gi bear	*ul* bee	*ro* box	*ne* book
	gi pig	*ul* fly	*ro* sofa	*ne* plate
	gi dog	*ul* cat	*ro* picture	*ne* cushion
	gi rats	*ul* birds	*ro* cups	*ne* clocks
	gi lions	*ul* monkeys	*ro* tables	*ne* cakes
	gi cows	*ul* snakes	*ro* televisions	*ne* vases
Training Set 2	Animate		Inanimate	
	Near	Far	Near	Far
	gi bears	*ul* bees	*ro* boxes	*ne* books
	gi pigs	*ul* flies	*ro* sofas	*ne* plates
	gi dogs	*ul* cats	*ro* pictures	*ne* cushions
	gi rat	*ul* bird	*ro* cup	*ne* clock
	gi lion	*ul* monkey	*ro* table	*ne* cake
	gi cow	*ul* snake	*ro* television	*ne* vase

Appendix C
Noun phrases (n = 36) employed in the testing phase

	Animate		Inanimate	
	Near	Far	Near	Far
Trained	*gi* cow	*ul* cat	*ro* cups	*ne* book
Trained	*gi* rats	*ul* bees	*ro* television	*ne* cushions
Trained	*gi* dog	*ul* flies	*ro* sofa	*ne* clocks
Partially trained	*gi* monkeys	*ul* bears	*ro* vase	*ne* table
Partially trained	*gi* snakes	*ul* lion	*ro* plates	*ne* box
Partially trained	*gi* bird	*ul* pig	*ro* cakes	*ne* pictures
New	*gi* goat	*ul* insects	*ro* desks	*ne* candles
New	*gi* rabbit	*ul* camels	*ro* phone	*ne* lamp
New	*gi* elephants	*ul* horse	*ro* spoons	*ne* towel

Appendix D
Response Windows (RW) of the training and test items in the two modalities

Length of the item (words)	RW for written modality (second)	RW for auditory modality (second)
6	3 s	5 s
7	4 s	5 s
8	5 s	6 s
9	5 s	6 s
10	6 s	7 s
11	6 s	7 s
12	7 s	8 s
13	8 s	8 s
14	9 s	9 s
15	10 s	9 s
16	11 s	10 s
17	12 s	10 s
18	13 s	11 s
19	14 s	11 s
20	15 s	12 s

Appendix E
Coding scheme for categorizing awareness (adapted from Rebuschat et al. (2015)

Code	Description
Full-confident	Full and accurate characterization of the animacy regularity, with confidence in reporting it
Full-hesitant	Full and accurate characterization of the animacy regularity but with hesitation in reporting it (e.g., mentioned only after being prompted to speculate)
Partial	Participants referred to partial characterization of the animacy regularity, or they mentioned a relevant category of the hidden regularity, such as animals or objects, living or nonliving things. However, they could not further explain the animacy meanings of the determiners.
Minimal	Participants did not make any references related to animacy. However, after having been told the rule by the researcher, they indicated very brief or fleeting awareness of the hidden rule, which had occurred during the experiment.
Unaware	Participants did not make any references related to animacy and reported that they had not become aware of the hidden regularity (animacy) at any point during the experiment.

Appendix F
Summary of frequentist and Bayesian one-sample t-tests

	Frequentist one-sample t-test				Bayesian one-sample t-test		
Condition	*t*	*p*	*Adj. p*	*d*	BF_{10}	*Posterior δ*	*95%CI*
Overall	2.451	0.008	0.187	0.261	3.91	0.253	[.057, .462]
–Trained	4.042	< .001	0.002	0.431	340.89	0.417	[.202, .634]
–Partially-trained	2.411	0.009	0.198	0.257	3.58	0.249	[.054, .458]
–Untrained	2.272	0.013	0.243	0.242	2.65	0.233	[.025, .443]
Explicit knowledge	5.479	< .001	< .001	0.153	169102.18	0.153	[.098, .208]
Implicit knowledge	4.267	< .001	< .001	0.101	451.67	0.101	[.054, .148]
Aware (Total)	2.304	0.013	0.253	0.320	3.32	0.304	[.058, .577]
–Trained	3.823	< .001	0.005	0.530	141.30	0.504	[.220, .794]
–Partially-trained	1.963	0.028	0.441	0.272	1.72	0.262	[.037, .530]
–Untrained	1.892	0.032	0.481	0.262	1.52	0.253	[.034, .530]
Unaware (Total)	0.889	0.190	1.000	0.148	0.41	0.176	[.011, .467]
–Trained	1.568	0.063	0.692	0.261	1.02	0.254	[.025, .569]
–Partially-trained	1.490	0.073	0.726	0.248	0.91	0.244	[.023, .556]
–Untrained	1.291	0.103	0.923	0.215	0.68	0.220	[.018, .526]
AV (Total)	1.866	0.038	0.532	0.398	1.85	0.364	[.043, .780]

(*Continued*)

(*Continued*)

	Frequentist one-sample t-test				Bayesian one-sample t-test		
Condition	*t*	*p*	*Adj. p*	*d*	*BF*$_{10}$	*Posterior* δ	*95%CI*
–Trained	2.470	0.011	0.233	0.527	5.07	0.474	[.089, .913]
–Partially-trained	2.315	0.015	0.277	0.494	3.87	0.445	[.075, .878]
–Untrained	2.277	0.017	0.284	0.486	3.63	0.438	[.071, .870]
VA (Total)	1.124	0.137	1.000	0.240	0.66	0.248	[.018, .630]
–Trained	1.665	0.055	0.665	0.355	1.37	0.330	[.034, .738]
–Partially-trained	0.869	0.197	1.000	0.185	0.49	0.215	[.013, .583]
–Untrained	0.574	0.286	1.000	0.122	0.36	0.182	[.010, .533]
AA (Total)	3.449	0.001	0.030	0.735	33.40	0.666	[.217, 1.141]
–Trained	4.128	< .001	0.006	0.880	136.65	0.804	[.322, 1.305]
–Partially-trained	2.856	0.005	0.113	0.609	10.36	0.548	[.133, 1.001]
–Untrained	1.861	0.038	0.499	0.397	1.84	0.363	[.043, .779]
VV (Total)	−0.691	0.751	0.751	−0.147	0.14	0.096	[.004, .367]
–Trained	0.286	0.389	1.000	0.061	0.28	0.155	[.008, .488]
–Partially-trained	−0.609	0.726	1.000	−0.130	0.15	0.099	[.004, .375]
–Untrained	0.055	0.478	1.000	0.012	0.23	0.137	[.006, .455]

Appendix G
The full omnibus model including Knowledge

				95% CI of Odds Ratios			
Fixed effects	Estimate	SE	Odds Ratios	Lower	Upper	*z*-value	*p*
(Intercept)	0.27	0.08	1.31	1.25	2.29	3.58	< .001
Exposure	−0.42	0.15	0.66	0.49	0.88	−2.82	0.005
Assessment	−0.13	0.15	0.88	0.66	1.17	−0.87	0.384
Training: Partially trained	0.00	0.10	1.00	0.82	1.22	−0.02	0.986
Training: Trained	0.16	0.10	1.17	0.96	1.44	1.54	0.123
Knowledge	0.07	0.09	1.08	0.78	1.11	0.81	0.418
Random effects (intercepts)	Variance	SD					
Participants	0.35	0.59					
Item	0.01	0.11					
Information criterion	Estimate						
LogLikelihood	−2029.6						
DIC	4059.3						

(*Continued*)

(*Continued*)

Fixed effects	Estimate	SE	Odds Ratios	95% CI of Odds Ratios Lower	Upper	*z*-value	*p*
AIC	4075.3						
BIC	4123.5						
R2	Estimate						
Marginal	0.015						
Conditional	0.112						

Note. The reference level for Exposure is Auditory; the reference level for Assessment is Auditory; the reference level for Training is Untrained; the reference level for Knowledge is Unconscious.

Appendix H
The full omnibus model including Awareness

Fixed effects	Estimate	SE	Odds Ratios	95%CI of Odds Ratios Lower	Upper	*z*-value	*p*
(Intercept)	0.25	0.08	1.29	1.11	1.49	3.32	< .001
Exposure	−0.39	0.15	0.68	0.50	0.90	−2.63	0.008
Assessment	−0.13	0.15	0.87	0.66	1.17	−0.92	0.359
Training: Partially trained	0.01	0.10	1.01	0.82	1.23	0.05	0.958
Training: Trained	0.17	0.10	1.18	0.97	1.45	1.65	0.098
Awareness	0.16	0.15	1.17	0.87	1.57	1.03	0.302
Random effects (intercepts)	Variance	SD					
Participants	0.34	0.58					
Item	0.01	0.10					
Information criterion	Estimate						
LogLikelihood	−2029.4						
DIC	4058.8						
AIC	4074.8						
BIC	4123.1						
R2	Estimate						
Marginal	0.016						
Conditional	0.111						

Note. The reference level for Exposure is Auditory; the reference level for Assessment is Audiory; the reference level for Training is Untrained; the reference level for Awareness is Unaware.

Cite this article: Zhao, C., Kormos, J., Rebuschat, P., and Suzuki, S. (2021). The role of modality and awareness in language learning. *Applied Psycholinguistics* **42**, 703–737. https://doi.org/10.1017/S0142716421000023