CrossMark

RESEARCH ARTICLE

An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral and neural processing

Yuichi Suzuki[1]* (ID), Hyeonjeong Jeong[2] (ID), Haining Cui[2] (ID), Kiyo Okamoto[2], Ryuta Kawashima[2] and Motoaki Sugiura[2] (ID)

[1]Kanagawa University, Kanagawa, Japan; [2]Tohoku University, Sendai, Japan
*Corresponding author. E-mail: szky819@kanagawa-u.ac.jp

Abstract

In this study, neural representation of adult second language (L2) speakers' implicit grammatical knowledge was investigated. Advanced L2 speakers of Japanese living in Japan, as well as L1 Japanese speakers, performed a word-monitoring task (proposed as an implicit knowledge test) in the MRI scanner. Behavioral measures were obtained from aptitude tests for explicit (language analytic ability) and implicit (statistical learning ability) learning. Findings indicate that, although both L1 and L2 speakers recruited neural circuits associated with procedural memory during the word-monitoring task, different brain regions were activated: premotor cortex (L1 speakers) and left caudate (L2 speakers). The premotor cortex activation was weaker in L2 than L1 speakers but was positively correlated with the left caudate activation, suggesting that their grammatical knowledge, while less automatized, was still developing. Behavioral sensitivity to errors was predicted only by explicit language aptitude, which may play a key role in the automatization of grammatical knowledge.

Introduction

Explicit and implicit knowledge are key constructs in second language (L2) learning and instruction (e.g., the extent to which explicit instruction can facilitate the acquisition of L2 knowledge that can be used for fluent communication). The two types of knowledge are typically distinguished using the awareness criterion. Explicit knowledge is posited to involve awareness of linguistic exemplars and rules that are accessible to learner's consciousness, whereas implicit knowledge has no correlates with awareness (DeKeyser, 2009; Rebuschat, 2013; Williams, 2009). To advance the current understanding of the nature of explicit and implicit learning

and knowledge, two important research areas have emerged: (a) the validation of explicit−implicit knowledge tests and (b) interrelationships between knowledge and aptitude.

The validity of explicit and implicit knowledge tests has been extensively investigated in second language acquisition (SLA) research (see Isbell & Rogers, 2021 for a recent review). In particular, because designing adequate tests specifically targeting implicit knowledge is extremely challenging, a significant effort has been dedicated to developing reliable and valid tests of implicit knowledge (e.g., Ellis, 2005; Godfroid et al., 2015; Suzuki, 2017; Suzuki & DeKeyser, 2015; Vafaee et al., 2017).

In cognitive neuroscience, explicit and implicit knowledge are often discussed in relation to two long-term memory systems—declarative and procedural memory. It is stipulated that explicit knowledge is acquired and stored in declarative memory, whereas implicit knowledge is associated with procedural memory (Buffington et al., 2021; Ullman, 2020). Declarative memory "supports the acquisition of facts and personal experiences," whereas procedural memory is "one type of implicit learning and memory system that supports the acquisition of cognitive and motor skills and habits" (Buffington et al., 2021, p. 636). Moreover, as discussed in the Literature Review section, these memory systems are supported by different brain systems.

While explicit and implicit knowledge intersect with declarative and procedural memory, the declarative−procedural distinction does not completely parallel the explicit−implicit demarcation (e.g., DeKeyser, 2017; Paradis, 2009; Ullman, 2020). As shown in Figure 1, while explicit knowledge always resides in declarative memory, implicit knowledge can be acquired through multiple mechanisms, such as conditioning, reflex, priming, and procedural memory (Squire & Dede, 2015). Still, researchers focusing on the theoretical distinctions between declarative−procedural and explicit−implicit issues (DeKeyser, 2020; Paradis, 2009; Ullman, 2020) would likely concur that procedural memory plays an essential role in fluent comprehension and production of L2 grammar.[1] Hence, investigating neural underpinnings of grammatical
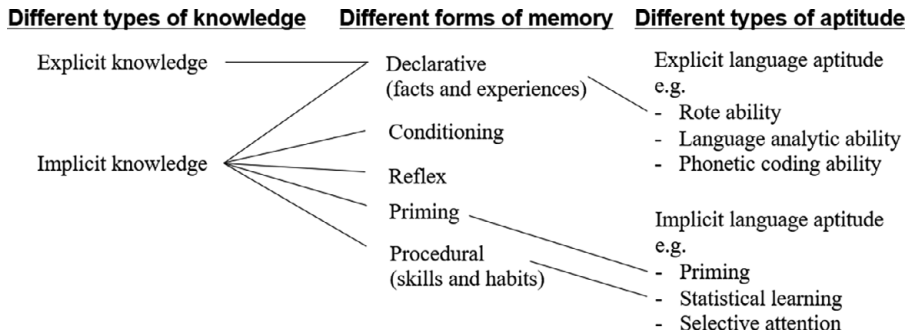

Figure 1. Conceptualizations of knowledge, memory, and aptitude in explicit−implicit and declarative−procedural domains.

---

[1]Declarative memory may underlie both implicit and explicit knowledge (see Ullman, 2020 for details), whereas procedural memory is always implicit in the sense of absence of awareness (Squire & Dede, 2015). Although the awareness criterion is used to distinguish explicit and implicit knowledge (DeKeyser, 2009; Rebuschat, 2013; Williams, 2009), it is extremely challenging to test awareness (DeKeyser, 2003) and applying

knowledge with a particular focus on procedural memory is highly informative (e.g., Paradis, 2009; Ullman, 2020; Yang & Li, 2012).

Recent theorization on aptitude has also provided important insights into the factors contributing to explicit and implicit knowledge and learning. Aptitude is a multicomponent construct comprising of "cognitive and perceptual abilities that predispose individuals to learn well or rapidly" (Granena, 2016, p. 577). Several SLA scholars argue that aptitude for implicit learning is distinct from that facilitating explicit learning (Granena, 2019; Li & DeKeyser, 2021; Linck et al., 2013). Explicit language aptitude is typically linked to the attention-driven processes such as associative (rote) and conscious analytic learning, whereas implicit language aptitude refers to the capacity for nonconscious, statistical sequence learning unintentionally through exposure (Granena, 2016, 2020; Li & DeKeyser, 2021). Investigating individual differences in explicit–implicit learning *aptitudes* in relation to explicit–implicit grammatical *knowledge* is a useful approach for advancing our understanding of the explicit and implicit *learning processes* (DeKeyser, 2012).

Explicit and implicit aptitude are also linked to declarative and procedural memory (see Figure 1).[2] Explicit aptitude is purported to encompass declarative memory as well as other attention-driven, conscious learning processes, such as language analytic ability and phonetic coding ability. Similarly, implicit aptitude has a broader scope than procedural memory, including priming and selective attention (Granena, 2016, 2020; Li & DeKeyser, 2021).

These transdisciplinary domains of explicit–implicit knowledge and aptitude lie at the crux of SLA research. However, there is paucity of neuroimaging studies focusing specifically on the complex relationships among these constructs. For instance, the role of procedural memory in naturalistic L2 acquisition has never been scrutinized by linking it to the implicit knowledge constructs from a neurocognitive perspective. To push the boundaries of this critical domain of SLA research, the brain responses of L2 Japanese speakers living in Japan were monitored as they completed a real-time grammar processing task. An individual difference approach was also adopted to investigate the relative importance of individuals' explicit and implicit aptitudes for the acquisition of L2 grammatical knowledge using both behavioral and neural measures. This study marks the first attempt at using fMRI findings to gain insight into the neural underpinnings of grammatical knowledge assessed by a word-monitoring task, proposed as a measure of implicit knowledge, as well as to elucidate the link between cognitive aptitudes and neural patterns elicited by such task.

Literature Review

Behavioral Measures of Implicit Knowledge in L2 Research

Explicit and implicit knowledge elicitation techniques are fundamental for advancing our understanding of explicit and implicit knowledge and learning. In one of the initial

---

the awareness criterion to probe implicit knowledge is beyond the scope of the current study (nonetheless, we do make an exploratory attempt in the last subsection of Discussion).

[2]Some researchers take a more focused approach to study individual differences in declarative and procedural memory (e.g., Buffington et al., 2021). In this article, following the tradition of language aptitude research (e.g., Carroll, 1981), we conceptualized individual differences in explicit and implicit language aptitude (e.g., Granena, 2020).

attempts to validate implicit knowledge tests of L2 grammar, Ellis (2005) proposed that imposing time pressure on a grammar task (e.g., timed GJT) can limit the use of explicit grammar knowledge, which would in turn elicit implicit knowledge. According to Paradis (2009), however, even some L2 adult learners that have attained high proficiency levels rely on the declarative memory, suggesting that advanced learners can use explicit knowledge rapidly. This possibility has been suggested by behavioral experiments showing that highly advanced L2 learners access their grammatical knowledge *consciously and quickly* even under time pressure (Suzuki & DeKeyser, 2015). Suzuki and DeKeyser termed this knowledge as *automatized (speeded-up) explicit knowledge*, which is defined as "a body of conscious linguistic knowledge including different levels of automatization" (Suzuki, 2017, p. 1230).[3]

As long as both explicit and implicit knowledge can be retrieved quickly by advanced L2 learners, it is extremely difficult to distinguish the two types of L2 grammatical knowledge employed at the behavioral level (DeKeyser, 2003). Nonetheless, researchers have started to utilize reaction-time psycholinguistic tasks to examine L2 learners' implicit knowledge that may be distinguishable from speeded-up explicit knowledge (Godfroid, 2016; Granena, 2013; Jiang, 2011; Suzuki, 2017; Suzuki & DeKeyser, 2015; Vafaee et al., 2017). One such task is a word-monitoring task, which can be administered to assess processing cost of specific grammatical errors relative to error-free sentences. In the word-monitoring task, participants are instructed to (a) listen for a monitoring word and react as soon as they hear it in an auditory sentence and (b) answer a comprehension question. The monitoring word is embedded in an auditory sentence and occurs right after the target grammatical structure. For instance, they could be presented with the following sentences:

Monitoring word: <u>to</u>
Grammatical sentence: John added a lot of milk <u>to</u> his tea.
Ungrammatical sentence: John added a lot of milks <u>to</u> his tea.

When participants listen for a monitoring word (e.g., to) in an ungrammatical sentence, if they can detect the error, they are likely to slow down to respond to the monitoring word compared to the one in the grammatical sentence. The reaction time (RT) difference between grammatical and ungrammatical items (defined as "grammaticality sensitivity index," or GSI) indicates the extent to which a processing slowdown is caused by the grammatical error (whether or not there is conscious awareness of the error).

While the word-monitoring task may be similar to timed GJT in terms of the rapid response requirement (cf., Godfroid et al., 2015; Vafaee et al., 2017), a potentially critical feature of word-monitoring task is the absence of explicit instructions to look for grammatical errors in the stimulus sentence. Participants are simply told to look for a monitoring word and answer a comprehension question at the end. The word-monitoring task can thus purportedly limit the use of (speeded-up) explicit knowledge. In addition, there is virtually no room to consciously apply explicit knowledge during real-time comprehension because the use of grammar knowledge is time-locked to hundreds of milliseconds. The word-monitoring task is thus

---

[3]Speeded-up explicit knowledge and automatized explicit knowledge refer to the same construct, and the term *speeded-up explicit knowledge* is used in this article.

arguably a purer measure of implicit knowledge than any types of GJT (Suzuki, 2017; Suzuki & DeKeyser, 2015; Vafaee et al., 2017).

As a case in point, factor-analytic research targeting advanced L2 learners (Suzuki, 2017; Vafaee et al., 2017) demonstrates that the word-monitoring task and other online comprehension tasks (self-paced reading and eye-tracking while listening task) scores load on the same axis and constitute a different latent factor (implicit knowledge) from the one underlying time-pressured GJTs (speeded-up explicit knowledge). While the word-monitoring may be a promising instrument for assessing implicit knowledge, in the extant research, the subtle difference (although potentially significant for the L2 theory construction) between speeded-up explicit knowledge and implicit knowledge was explored only through the behavioral, factor-analytic approach. Because these behavioral studies have left lingering ambiguities in part due to the lack of reliable method for assessing awareness (DeKeyser, 2003), we shift away from the criterion of awareness. In this study, a neuroimaging technique is adopted to directly examine the brain regions associated with declarative and/or procedural memory that can be linked to explicit and implicit knowledge (see the next section).

The Neural Basis of Declarative-Procedural Memory

Figure 2 illustrates the brain areas primarily associated with procedural and declarative memory systems. In contrast to declarative memory (rooted in hippocampus and medial temporal lobe structures), procedural memory is primarily associated with frontal cortical-basal ganglia regions (Squire, 2004). According to Ullman (2020), procedural memory is posited to account for specific stages of L2 learning. The basal ganglia (particularly the anterior caudate nucleus and putamen) is primarily recruited in the early phases of procedural learning. However, frontal regions, particularly in the premotor cortex (BA6) and the inferior frontal gyrus (IFG, BA44), can be more important for the later stage of proceduralization, that is, automatization.

This declarative-procedural distinction applies to language learning (Ullman, 2020) and is informed by two lines of fMRI research pertaining to (a) artificial linguistic system (ALS) learning and (b) first language (L1) syntactic processing. In the studies based on the ALS learning paradigm, participants are typically exposed to linguistic sequences (based on either artificial language or nonartificial language like miniature language) under different learning conditions such as intentional-explicit
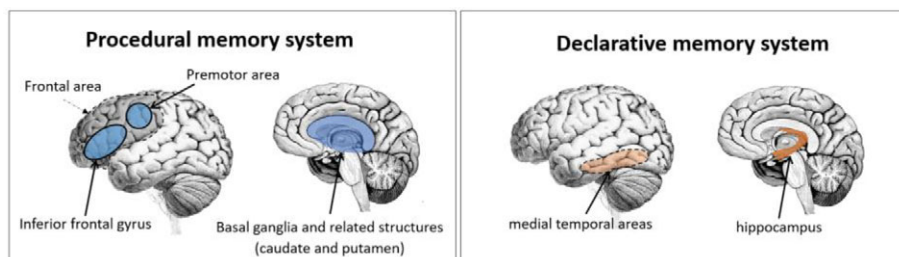

Figure 2. Brain areas primarily associated with procedural and declarative memory systems.

and incidental-implicit. After the exposure phase, a grammaticality judgment task (GJT) is typically administered as an outcome test to elucidate changes in the brain regions recruited for grammar processing. Recently, Tagarelli et al. (2019) conducted a meta-analysis of 24 fMRI studies focusing on adult grammar learning (including natural languages as well as ALS). To examine the neural correlates of declarative and procedural memory, the authors compared the findings yielded by two training conditions: (a) explicit grammar training condition that involved a type of explicit training such as explanation of grammatical rules (10 groups with 134 participants) and (b) implicit grammar training condition that required no attention to linguistic features of target ALS (14 groups with 195 participants). Their exploratory analyses revealed that hippocampal areas in the medial temporal lobe were significantly activated in the explicit training condition only. In contrast, the implicit training condition induced higher activation in frontal-basal ganglia circuits (e.g., basal ganglia [anterior caudate, putamen, and thalamus] as well as IFG [pars triangularis, pars opercularis]) without any hippocampus involvement. Furthermore, evidence yielded by the brain-lesion study conducted by Opitz and Kotz (2012) suggests that impairment in a frontal region associated with procedural memory (i.e., the ventral premotor region) impedes ALS learning.

Second, accumulating evidence yielded by neuroimaging research also suggests that the left prefrontal cortex is recruited for automatic syntactic processing by L1 speakers (e.g., Friederici et al., 2006; Hashimoto & Sakai, 2002; Sakai, 2005). For instance, Friederici et al. (2006) examined the neural processes of L1 German speakers by presenting them with grammatical and ungrammatical sentences (with word-order violations) as a part of the GJT. Their findings indicate that the left IFG, particularly in the pars opercularis, was selectively activated when participants were presented with ungrammatical sentences. Similarly, Hashimoto and Sakai (2002) demonstrated that L1 Japanese speakers recruit the left inferior frontal gyrus and the premotor cortex for making syntactic judgments pertaining to structure-dependent rules. Because L1 speakers presumably possess procedural knowledge that is highly automatized due to their extensive L1 use, the left prefrontal cortex seems to be implicated in the use of automatized grammatical knowledge. Hence, if L2 grammatical knowledge is highly automatized, the L2 cortical representation may potentially overlap with that of L1 (e.g., left IFG and premotor cortex). In the current study, the nature of implicit knowledge is scrutinized based the neurocognitive processes (e.g., automatization) that presumably involve procedural memory.

Behavioral Research on Individual Differences in L2 Grammar Acquisition in Naturalistic Contexts

Interest in cognitive aptitudes that explain individual variability in explicit and implicit learning has surged in recent years, based on the premise that such differences play a crucial role in L2 learning (Granena, 2019; Li & DeKeyser, 2021; Linck et al., 2013). Probing systematic relationships between aptitude and linguistic knowledge can shed light on the underlying learning processes by making inferences about cognitive processes that are facilitated or hindered by specific aptitude components (DeKeyser, 2012). For instance, a positive relation of a particular grammar test score with implicit aptitude would suggest that an implicit learning process is involved in the acquisition of knowledge tapped by that grammar test.

An emerging line of research in this domain has revealed that cognitive capacity for explicit and implicit learning can predict adult L2 learners' acquisition of implicit knowledge in naturalistic immersion contexts (Granena, 2013; Suzuki & DeKeyser, 2015, 2017). Two cross-sectional studies have been conducted targeting adult L2 learners living in naturalistic acquisition settings, yielding consistent behavioral evidence suggesting that implicit language aptitude, measured by the SRT task,[4] significantly predicts the attainment of real-time grammar processing ability, measured by the word-monitoring task (Granena, 2013; Suzuki & DeKeyser, 2015). In the study conducted by Granena (2013), adult advanced L2 Spanish learners with Chinese as their L1 were recruited in Spain. They had arrived in Spain after the age of 16 and had lived in Spain for at least 5 years (mean length of residence was 8.42 years). The authors found that their SRT scores were significantly correlated with the GSI from the word-monitoring task. Similarly, Suzuki and DeKeyser (2015) found a positive association between the SRT score and the GSI on five Japanese particles among advanced Japanese L2 learners with Chinese L1 who live in Japan. This positive relationship was found only among those whose duration of residence was relatively long (approximately 2.5 years), suggesting that it takes at least a few years of immersion experience (a proxy for enough L2 naturalistic exposure) for acquiring implicit knowledge, which was arguably measured by the word-monitoring task. However, Suzuki and DeKeyser (2015) found no significant relationship between metalinguistic knowledge task score and word-monitoring performance (GSI), regardless of lengths of residence. These outcomes suggest that the type of knowledge tapped by the word-monitoring and metalinguistic knowledge tasks is different.

Further advancement was made in this line of investigations by Suzuki and DeKeyser (2017) who examined the roles of both explicit and implicit aptitudes. As a part of their study, 100 advanced Japanese L2 learners with Chinese as their L1 were administered implicit knowledge tests (i.e., three real-time processing tasks, including the word-monitoring task), along with speeded-up explicit knowledge tests (i.e., form-focused task including time-pressured GJTs) as well as explicit and implicit aptitude (i.e., LLAMA_F and SRT task) tests. The findings yielded by structural equation modeling analysis showed that explicit aptitude significantly predicted the acquisition of speeded-up explicit knowledge, measured by time-pressured form-focused task (e.g., timed GJTs), which in turn significantly predicted implicit knowledge. In sum, while implicit learning aptitude may predict implicit knowledge in naturalistic L2 acquisition (Granena, 2013; Suzuki & DeKeyser, 2015; cf., Godfroid & Kim, 2021), explicit aptitude may have an indirect contribution to the acquisition of implicit knowledge, mediated by speeded-up explicit knowledge (Suzuki & DeKeyser, 2017). Further research from a neurocognitive perspective is thus needed to ascertain the extent to which explicit and implicit aptitudes play a facilitative role in the acquisition of implicit knowledge.

---

[4]Although the theoretical scope of explicit−implicit language aptitude and declarative−procedural memory research domains differs, the constructs and measurements employed sometimes overlap. For instance, serial-reaction time (SRT) task, which measures sequence learning ability, is one of the most frequently used cognitive tasks for assessing both implicit learning aptitude (Granena, 2020) and procedural memory (Buffington et al., 2021). Although this SRT task can be characterized as "implicit," "procedural," or "statistical" learning task, it is referred to as "implicit" aptitude test from the theoretical standpoint of this article.

fMRI Research on Individual Differences in ALS Learning

Adopting Ullman's declarative-procedural neurobiological model as the framework, Morgan-Short et al. (2015a) conducted a longitudinal neuroimaging experiment to elucidate the roles of individual differences in declarative and procedural memory ability. These authors trained 13 English native speakers on an ALS (i.e., meaning-bearing artificial language), which is consistent with natural language features, over 2 weeks for a total of four 3-hour training sessions. No explicit grammar rules or explanations were provided during the training phase. Changes in brain activation were assessed twice (after the first and fourth training session) by subjecting the participants to an MRI scan as they performed auditory GJT. In addition, individual differences in declarative memory (Modern Language Aptitude Test Part V and continuous visual memory task) and in procedural memory[5] (weather prediction and Tower of London tasks) were assessed to gain further insights into the two long-term memory systems.

Two key findings pertaining to individual differences in declarative and procedural memory ability emerged. First, somewhat surprisingly, the score on the procedural memory tasks was not positively associated with neural activity during the first or the second GJT performance. Second, the score on the declarative memory tasks was implicated in greater activation in the neural circuits associated with procedural memory (left IFG), as well as declarative memory (e.g., the insula and the right precuneus), during the first GJT performance. These results may be consistent with the notion that declarative memory was initially relied upon which facilitated procedural learning (DeKeyser, 2020).

While these systematic attempts to better understand the neural underpinnings of L2 grammar learning are clearly valuable, the methodologies adopted in the previous fMRI studies preclude in-depth understanding of L2 grammar acquisition. The ALS paradigm provides researchers with a methodological advantage, as learners can attain high levels of mastery in a carefully designed artificial language in a relatively short period of time in a laboratory setting (Morgan-Short et al., 2015a). However, this may also be considered a disadvantage in terms of ecological validity, given that in real-life contexts adult L2 learners often fail to reach high levels of mastery (e.g., nativelikeness attainment) despite extensive exposure even in immersion environments (e.g., Abrahamsson & Hyltenstam, 2009). With the aim of increasing the ecological validity of the reported findings, it is thus important to extend the research scope to a group of L2 learners in an immersion setting and examine neurocognitive individual differences.

The Current Study

The goal of the current study was to advance the current understanding of the nature of explicit and implicit learning and knowledge among adult L2 learners. Two related problems pertaining to these phenomena were investigated. First, as behavioral evidence from real-time grammar processing tasks is inevitably ambiguous, a neuroimaging technique was adopted to scrutinize the validity of the word-monitoring task as a measure of implicit knowledge from neural perspectives. Because implicit

---

[5]Although weather prediction and Tower of London tasks have been primarily used in research that focuses on procedural memory, they are also included as potential measurements of implicit language aptitude (e.g., Granena, 2020; Li & DeKeyser, 2021). If one takes a broader conceptualization of implicit language aptitude, these two tasks can be construed as measures of implicit language aptitude. Hence, Morgan-Short et al.'s (2015a) work was included in the Literature Review section.

knowledge is associated with procedural memory (see the Literature Review section), examining neural underpinnings of the word-monitoring task performance from a declarative-procedural memory perspective can shed light on the nature of L2 implicit knowledge.

Second, accumulating evidence indicates that explicit and implicit language aptitudes play a key role in L2 grammar knowledge attainment in naturalistic L2 acquisition settings (Granena, 2013; Suzuki & DeKeyser, 2015, 2017). However, the relationships of cognitive aptitude with the neural representations and processing of implicit grammatical knowledge, particularly among advanced L2 learners, remains insufficiently understood. Hence, a neurocognitive individual difference approach was taken here to explore the potential relationships between cognitive aptitude and the neural responses elicited by the word-monitoring task.

In this study, advanced Japanese learners (L2 speakers), as well as native Japanese speakers (L1 speakers), completed a word-monitoring task (targeting Japanese case-marking particles) inside an MRI scanner aiming to identify underlying neural circuits that support task performance. To explore individual differences, three predictors were derived, based respectively on the participants' performance on a linguistic task (metalinguistic knowledge task) and cognitive aptitude tests for explicit (LLAMA_F) and implicit (SRT task) learning, which were administered outside the MRI scanner. The following four research questions (RQs) were addressed:

1. What are the neural correlates of sensitivity to grammatical errors in the word-monitoring task?
2. To what extent do neural patterns of L2 and L1 speakers overlap?
3. What linguistic and cognitive aptitude factors predict behavioral sensitivity to grammatical errors in the word-monitoring task (GSI) among L2 speakers?
4. What linguistic and cognitive aptitudes predict the brain activations during the word-monitoring task among L2 speakers?

RQ1 was motivated by the predictions based on the neurobiological theories proposed by Paradis (2009) and Ullman (2020). Due to the word-monitoring task design features (i.e., the absence of explicit instructions to look for grammatical errors and assessment of online grammar processing), it was hypothesized that real-time processing of errors would preferentially recruit brain regions responsible for procedural memory (frontal-basal ganglia circuits), rather than those pertaining to declarative memory (hippocampus and medial temporal lobe).

RQ2 focused on the comparison between L2 and L1 speakers. Because L1 speakers' linguistic knowledge is presumably automatized, their brain imaging results were contrasted with those obtained for L2 speakers. In line with Ullman's model (2020), it was hypothesized that L1 speakers would activate the frontal region, particularly in the left IFG (BA44) and the premotor cortex (BA6), but would not rely on the basal ganglia when retrieving L1 knowledge because it is primarily recruited in the early phases of procedural learning. In contrast, as L2 speakers' knowledge is not fully automatized (Ullman, 2020), the basal ganglia might still remain active, but brain imaging results of some L2 speakers might show similar patterns to those noted for L1 speakers (i.e., activation of BA6 and BA44).

RQ3 and RQ4 are aimed at uncovering the potential links between the cognitive aptitudes and grammatical knowledge of adult naturalistic L2 learners. RQ3 was motivated by previous behavioral L2 studies that elucidated the roles of

metalinguistic knowledge and cognitive aptitudes for explicit and implicit learning in the acquisition of grammar knowledge assessed by the word-monitoring task in naturalistic L2 immersion contexts (Granena, 2013; Suzuki & DeKeyser, 2015, 2017). Based on these findings, it was hypothesized that implicit aptitude would significantly predict word-monitoring performance (GSI), whereas neither metalinguistic knowledge nor explicit aptitude would be a significant predictor of word-monitoring task performance. Regarding RQ4, it was hypothesized that implicit aptitude, rather than metalinguistic knowledge or explicit aptitude, would predict the activation of brain regions associated with procedural memory (Paradis, 2009; Ullman, 2020).

Method

Participants

Participants were recruited at a national university located in the northern part of Japan. Only the individuals that met the following inclusion criteria were invited to take part in the study: (a) Mandarin native speakers, (b) advanced Japanese proficiency equivalent to N1 in the standardized Japanese Language Proficiency Test (JLPT), which is the minimum requirement for acceptance into a regular college undergraduate/ graduate program in Japan, (c) arrived to Japan at the age of 17 or older, and (d) living in Japan for at least 12 months.

Thirty-two L2 Japanese learners meeting these stringent requirements were enrolled in this study. However, data pertaining to seven participants were subsequently excluded from the analyses, as four participants failed to attend all experimental sessions, one participant was removed due to the experimenter's error, and excessive motion (over 3 mm) within the scanner was detected in two cases. Data related to the remaining 25 participants (10 males, 15 females) was analyzed and is reported throughout this article. The participants' background information is presented in Table 1. In terms of their academic level, they were undergraduate ($n = 3$), research ($n = 4$), master's ($n = 16$), and doctoral ($n = 1$) students. More than half of participants ($n = 14$) obtained a bachelor's degree in Japanese as a major at a Chinese university, while other participants obtained a bachelor's degree in other fields (e.g., biology, engineering, food science, environment). In addition, four participants were pursuing or had obtained a master's degree in Japanese linguistics at a Japanese university.

To examine the common neural responses during the word-monitoring task between L1 and L2 speakers (i.e., RQ2), 21 native Japanese speakers were also recruited. They were undergraduate students recruited at the same university as L2 learners (14 males, 7 females; mean age = 21.57 years, $SD = 1.62$, range: 18−24).

Table 1. Background information for L2 learners

	M	*SD*	*Min*	*Max*
Age	24.25	1.70	20	27
Starting age of instruction	18.21	1.47	15	21
Length of instruction (months)	50.62	27.20	6	120
Age of arrival	21.46	1.62	18	24
Length of residence (months)	30.21	12.63	12	58

Note: One participant failed to complete the questionnaire.

All participants met the fMRI experiment requirements, as they were right-handed, of normal hearing, and had either normal or corrected-to-normal vision without neurological deficits or psychiatric disorders. This study was conducted with the approval of the Institutional Review Board of the university from which the study participants were recruited. Written informed consent was obtained from each participant prior to the experiment.

Target Structures

Four grammatical structures that do not exist in participants' L1 Chinese were used for this study: (a) case-marking particles *ga−o* for transitive-intransitive verb pairs, (b) case-marking particles *wa−ga* in adverbial clause, (c) case-marking particles *wa−ga* in relative clause, and (d) locative case-marking particles *ni−de*. These particles are basic grammatical structures in Japanese because they essentially convey the functions of arguments. In addition, these structures were previously used by Suzuki and DeKeyser (2015), and they are usually taught explicitly in Japanese classes. In the debriefing questionnaire, all learners reported to have studied about the transitive-intransitive verbs and *ni−de* in school and/or through self-study using grammar reference books. However, eight and four learners, respectively, indicated no recollection of having learned about *wa−ga* in adverbial clause and *wa−ga* in relative clause.

Particles o−ga *for transitive−intransitive verbs*

Sixteen transitive/intransitive verb pairs were chosen that share the stem and morphological markings that differentiate transitive from intransitive verbs. Example (1a) illustrates a sample grammatical and ungrammatical sentence with a transitive verb (*agkeru*, "open"). A theme (*mado*, "window") should be followed by the object-marking particle *o* rather than the subject-marking particle *ga*. In contrast, as shown in Example (1b), with an intransitive verb (*hajimaru*, "start"), the subject should be followed by the subject marking particle *ga* rather than *o*.

(1a) *Fuyu ni mado o/*ga   akeru to,   samui to   omou.*
Winter window−OBJECT   open if,   cold   that think.
I think it is cold if we open the window in winter.
(1b) *Natsuyasumi ga/*o   hajimaru to, gakusei wa   ureshii.*
Summer vacation−SUBJECT start if,   students−SUBJECT   happy.
When summer vacation starts, students become happy.

Particles wa−ga *in adverbial clause*

Topic-marking particle (*wa*) and subject-marking particle (*ga*) are often confusing for L2 Japanese learners. One of the distinctions made between the case-marking particles *wa* and *ga* is based on the location in the sentence structure. When the first adverbial clause contains *wa*, another subject is not expected in the main clause. As illustrated in Example (2), a new subject (i.e., "*otona*," adults), which was a monitoring word, is not expected when *wa* is used in the adverbial clause. In other words, a monitoring word (i.e., "*otona*") occurs at the exact point of ungrammaticality.

(2) *Chugakusei ga/*wa   tabako o sutteita ra, <u>otona wa</u>   okoru bekida.*
    JHS students—SUBJECT smoke   if, <u>adults—SUBJECT</u> scold   should.
    If junior high school students are seen smoking, adults should scold him/her.

Particles wa—ga *in relative clause*
In a similar vein, the case-marking particle *ga* should also be used (rather than *wa*) within the relative clause, as illustrated in Example (3). A monitoring word (i.e., "*manshon,*" mansion) occurred at the exact point of ungrammaticality.

(3) *Yumeijin ga/*wa   sumu <u>manshion</u> wa   takai   darou.*
    Celebrity—SUBJECT live   mansion—TOPIC expensive maybe
    The mansion in which a celebrity lives may be expensive.

Particles ni—de *indicating locations*
The locative case-marking particles *ni* and *de* are distinguished by the verb semantics. *De* should be used for indicating the place where an action takes place, while *ni* is mainly used for stative verbs (e.g., be, live). Example (4) illustrates this restriction with an action verb (*kaimonosuru* "do shopping").

(4) *Konbini de/*ni   <u>kaimonosuru</u> no wa   totemo benri   da.*
    Convenience store—LOCATION do shopping TOPIC very   convenient be
    It is convenient to do shopping at the convenience store.

Instruments

The participants completed the word-monitoring task in the 3T-MRI scanner, while the other tasks were administered outside the scanner in a quiet room. All materials are available in the IRIS Digital Repository (Marsden et al., 2016).

Word-monitoring task (fMRI)
Figure 3 illustrates the word-monitoring task procedure. In this task, participants (a) saw a monitoring word, (b) listened to a sentence for that monitoring word and pressed a button as soon as they identified it in the sentence, and (c) made a semantic plausibility judgment of the sentence.

An event-related design was employed for the fMRI word-monitoring task. Each trial started with the presentation of a fixation point (+) for one second, followed by a monitoring word. Two seconds later, the auditory sentence was played through the headphones. The monitoring word remained on the screen until the response was provided.

When responding to the monitoring word, participants were told to use their right index finger to press the blue button on the game pad. After the sentence ended, a yes/no plausibility judgment question appeared on the screen, which focused participants' attention on the meaning of the sentence. For instance, participants would be expected to respond "agree" (using the right index finger to press the blue button) to sentences such as "China is located near Russia," or "disagree" (using the right middle finger to press the yellow button) to sentences such as "We feel much better if we don't
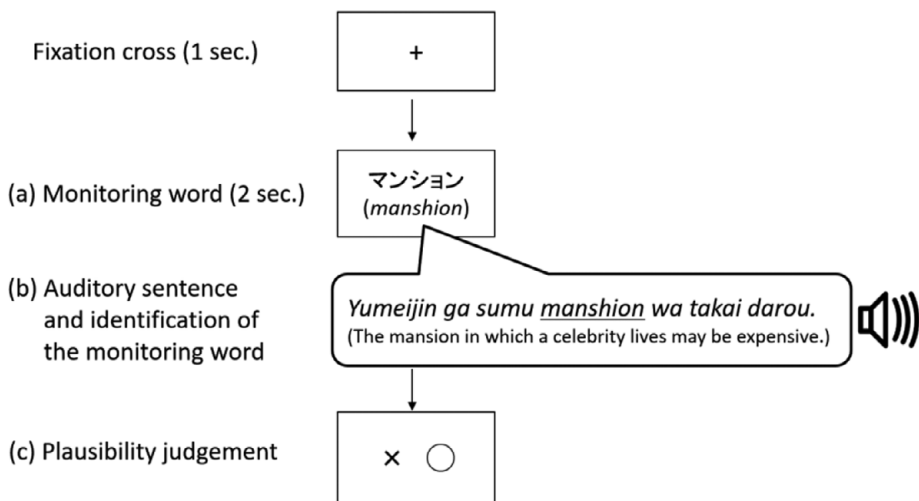
Figure 3. Word-monitoring task.

Short resting periods of 2—8 second duration were inserted between trials. These randomly determined between-trial intervals were included to increase the sensitivity of brain imaging for the critical cognitive process (e.g., detection of grammatical structures).

The word-monitoring test comprised 96 trials, 64 of which were critical trials (all sentences were plausible) and 32 were filler trials. The critical trials included 32 grammatical (8 sentences × 4 structures) and 32 ungrammatical sentences. The filler trials consisted of implausible sentences only (e.g., Monitoring word: *Basukettobooru*, *Basukettobooru o suru toki wa, ashi de booru o takusan keru*, "When playing basketball, we kick the ball a lot"). Two counterbalanced lists were created for the 64 critical trials. The 32 grammatical sentences in List 1 had corresponding ungrammatical sentences in List 2, and vice versa.

The timing of this experiment (word presentation, response time, and button press) was controlled and the responses were recorded using DMDX (Forster & Forster, 2003). Head movement was also restricted using a foam rubber pad and a head-restraining belt. All auditory stimuli, which were digitally recorded (44.1 kHz) by a native speaker of Japanese, were presented through MRI-compatible noise-canceling headphones (Optoacoustics Ltd., Israel), which reduced MRI scanning noise and projected auditory stimuli well. An intermission was provided in the middle of the word-monitoring task to reduce fatigue. It took about 40 minutes to complete the word-monitoring task.

All participants were given instructions for the word-monitoring task. In addition, to familiarize participants with the MRI task procedures, they first performed practice trials using a gamepad outside the scanner, after which they were presented with 10 practice items inside the MRI scanner. Participants were allowed to repeat the practice trials until they became comfortable with performing the task. They were also told to minimize head movement during MRI scanning and learned how to keep their heads still.

In the preliminary analysis, the accuracy scores on the plausibility judgment component were computed to check whether the participants were focusing on

meaning when performing the task. The mean accuracy score was 97.37% (SD = 3.31%) and 97.57% (SD = 2.65%) for the L1 and L2 groups, respectively. In the previous studies, the exclusion criterion was typically set at 75% accuracy (e.g., Suzuki, 2017). Because the lowest accuracy scores were above the criterion (85% and 88% for the L1 and L2 groups, respectively), all the participants' RT data related to the monitoring word were subjected to further analyses. To clean the RT data, outlying responses (those that fell outside the ±2.5 SD range around each participant's mean) were discarded. These procedures, along with display errors (i.e., a frame could not be moved into video memory by the specified time), eliminated 1.24% and 1.39% of L1 and L2 speakers' responses, respectively.

To compute GSI, RTs to the monitoring word in the critical sentences (all of which were plausible) were analyzed. The monitoring word was always a content word and underlined for the example sentences (1)−(4) for each grammatical structure described in the preceding text. GSI was computed by subtracting grammatical RT from ungrammatical RT, indicating the online sensitivity to grammatical errors (e.g., Granena, 2013; Suzuki & DeKeyser, 2015; Suzuki, 2017). Reliability indexed by Cronbach's alpha for the word-monitoring task was high for the two counterbalanced lists (List 1 = .93 and List 2 = .80 in the L1 group; List 1 = .96 and List 2 = .98 in the L2 group).

Metalinguistic knowledge task

After the word-monitoring task, participants took a paper-and-pencil metalinguistic knowledge task, which consisted of (a) a correction and (b) an explanation component. They were told that each sentence contained one grammatical error and were instructed to (a) underline the part where they believe the grammatical error exists and write down the correct Japanese term below, and (b) explain why the original was incorrect (either in Japanese or Chinese). The list presented to the participants contained 16 ungrammatical sentences (4 sentences × 4 target structures), all of which were extracted from the stimulus list for the word-monitoring task. No time limit was imposed for the completion of this task.

The responses were dichotomously scored as correct or incorrect for correction and explanation parts. A credit was given only when both the correction and the explanation were accurately provided for the target rule. A rubric for scoring the test-takers' explanation was developed for each target structure (see preceding text). Two native Japanese speakers used the rubric to independently score the explanation part, achieving 98.25% interrater reliability (any inconsistencies in scoring were resolved by a third coder). Reliability indexed by Cronbach's alpha was .86 for the L2 group.

SRT Task

A probabilistic SRT task was administered to measure sequence learning ability as a component of implicit language aptitude. It was adopted from Kaufman et al.'s (2010) study and has been used in previous L2 research on explicit and implicit knowledge and learning (Granena, 2013; Suzuki & DeKeyser, 2015, 2017; Yi, 2018). In this task, a dot was displayed at one of four locations on the computer screen and the participants were instructed to react to the stimulus as quickly and as accurately as possible by pressing the corresponding key. The sequence of dots was generated by two statistical rules that altered randomly unbeknownst to the participants: 85% of the sequences followed a more probable rule (the training condition), whereas the other 15% of the sequences was generated by a less probable rule (the control condition). The test comprised eight blocks, with 120 trials in each block. Task

performance was scored by subtracting the mean RTs in the training condition (Sequence A) from those in the control condition (Sequence B), which reflected the amount of learning. Reliability indexed by split-half reliability, corrected using Spearman–Brown formula, was .66 for the L2 group. This value is higher than the reliability (about .40–60) for statistical SRT tasks reported in previous L2 research (Suzuki & DeKeyser, 2015, 2017; Yi, 2018).

LLAMA_F

The LLAMA_F (Meara, 2005) was administered to measure language analytic ability as a component of explicit language aptitude (Granena, 2019). Participants were told that the test consisted of a 5-minute learning phase and a test phase. In the learning phase, participants were given 5 minutes to learn a new language by studying sentences matched with pictures. In the testing phase, the program displayed a picture and two sentences, one grammatical and the other ungrammatical, and their task was to choose the grammatical sentence. Ten additional items were added to the original 20 items to increase reliability (see Suzuki & DeKeyser, 2017). There was no time limit for completing the items, but participants were not allowed to return to the items they had already answered. Reliability indexed by Cronbach's alpha was .68 for the L2 group. This value is higher than the reliability (.60) reported in a recent large-scale validity study on the LLAMA test battery (Bokander & Bylund, 2019).

Procedure

Participants attended two test sessions in the laboratory. In the first session, they completed the word-monitoring, SRT, and LLAMA_F task, along with the background questionnaire. The metalinguistic knowledge task was administered during the second session. This order minimized the potential influence of taking the metalinguistic knowledge task on the more implicit word-monitoring task.

Brain Data Acquisition

Scanning was conducted using the Philips Achiva 3T MRI scanner (Eindhoven, the Netherlands). Blood oxygen level-dependent T2*-weighted MR signals were measured using a gradient echo-planar imaging (EPI) sequence. Thirty-two axial gradient-echo images (EPI) covering the entire brain were acquired during all sessions with the following parameters: repetition time = 2,000 ms, echo time = 30 ms, flip angle = 80°, slice thickness = 4 mm, no slice gap, field of view = 190 mm, matrix = 64 × 64, and voxel size = 3 × 3 × 4 mm. Additionally, T1-weighted anatomical images (thickness = 1 mm, field of view = 224 mm, 224 × 224 matrix, repetition time = 1,800 ms, echo time = 3.2 ms) were obtained from each participant to serve as a reference for anatomical correlates. The following preprocessing procedures were performed using Statistical Parametric Mapping (SPM12) software (Wellcome Department of Imaging Neuroscience, London, UK) and MATLAB (MathWorks, Natick, MA, USA): adjustment of acquisition timing across slices, correction for head motion, coregistration to the anatomical image, spatial normalization using the anatomical image and the Montreal Neurological Institute (MNI) template, and smoothing using a Gaussian kernel with a full-width at a half-maximum (FWHM) of 6 mm. Imaging data that showed more than 3 mm of excessive motion within the scanner and technical problems were excluded from the statistical analysis.

Statistical Analysis

Group-Level Analysis

Conventional first-level (within-subject) and second-level group (between-subjects) analyses were performed using SPM12 for event-related fMRI data. In the first-level analysis for word-monitoring, the functional imaging data from each subject was input into a general linear model to examine hemodynamic responses using a multisession design matrix pertaining to the three conditions (grammatical sentences, ungrammatical sentences, and fillers) as well as the trials in which wrong response to the plausibility judgment question was given. Six movement parameters (three translations, three rotations) were also included as regressors of no interest. A high-pass filter with a cutoff period of 128 seconds was used to eliminate an artifactual low-frequency trend. Each trial was modeled as an epoch for the duration of each auditory sentence for the word-monitoring task, during which targeted grammar processing occurs. Contrast images between conditions (ungrammatical sentences > grammatical sentences) were generated for each participant.

The second-level group analysis at the whole-brain level was conducted to investigate the neural correlates of sensitivity to grammatical errors in the word-monitoring task. A random effect one-sample *t*-test was performed using as data the contrast estimate (ungrammatical sentences > grammatical sentences) for each subject (RQ1).

To further investigate the commonalities and differences between the brain activation patterns of L1 and L2 groups, a joint group analysis was conducted (RQ2). At the whole brain level, a mixed ANOVA was conducted using SPM12 with groups (L1 versus L2) as a between-subject factor and grammaticality (grammatical vs. ungrammatical) as a within-subject factor. Region of interest (ROI) analysis was further conducted for the premotor cortex and the left caudate. The choice of these two brain regions was informed by prior ALS research (Tagarelli et al., 2019) and L1 syntactic processing studies (Friederici et al., 2006; Hashimoto & Sakai, 2002; Sakai, 2005), as well as declarative-procedural models proposed by Paradis (2009) and Ullman (2020). For the ROI analysis, a mixed ANOVA was conducted on the parameter estimates, with groups (L1 vs. L2) as a between-subject factor and brain areas (premotor cortex and head of left caudate) as a within-subject factor. Using the Marsbar toolbox, parameter estimates were extracted for each participant based on the ungrammatical−grammatical contrast in the premotor cortex and head of left caudate activation profiles (Brett et al., 2002).

In all analyses, the statistical threshold was set at $p < .05$ using multiple comparison correction with the cluster size (Slotnick, 2017). Monte Carlo simulation with 2,500 iterations was applied at the whole brain level ($64 \times 64 \times 32$) and 6-mm FWHM Gaussian kernel, yielding a voxel threshold of $p < .001$, corrected for multiple comparisons to $p < .05$ with a cluster extent threshold of 27 voxels. Only clusters that exceed this threshold were reported with the following detailed information: the coordinates (x, y, z) of the activation peak in the MNI space, peak T-value, and size of the activated cluster in number (k) of voxels ($2 \times 2 \times 2$ mm^3). Activation peak coordinates were reported in the MNI space and activated brain regions were identified using the SPM Anatomy Toolbox in SPM12 (Eickhoff et al., 2005).

Individual Difference Analysis

To examine the extent to which linguistic and cognitive aptitude measures account for the word-monitoring task behavioral performance in L2 speakers, multiple regression analysis was conducted on the GSI as a dependent variable (RQ3). Three predictors

were included in the model: metalinguistic knowledge task score and two aptitude measures—one for implicit (SRT) and another for explicit learning (LLAMA_F). All measured variables were normally distributed, and the multicollinearity assumption was met (VIF < 10, tolerance > .02).

Regarding RQ4, the multiple regression analyses were conducted on the contrast used for the whole-brain analysis (i.e., the contrast areas, denoted previously as [ungrammatical sentences > grammatical sentences]) with the same three predictors (metalinguistic knowledge, SRT, and LLAMA_F scores).

Results

Descriptive Statistics

The descriptive statistics related to the word-monitoring task performance, as well as three independent variables, are presented in Table 2. In the word-monitoring task, in the L2 group, the mean RT was 524 ($SD = 137$) and 527 ($SD = 118$) for the grammatical and ungrammatical sentences, respectively. The L1 group's mean RT was 376 ms ($SD = 69$) and 440 ms ($SD = 75$) for the grammatical and ungrammatical sentences, respectively. A mixed ANOVA revealed a significant interaction between group (L1 and L2) and grammaticality (grammatical and ungrammatical items), $F(1, 44) = 17.60$, $p < .001$, $\eta p^2 = 0.29$. L1 speakers' response to the monitoring word in ungrammatical items significantly slowed down compared to that in grammatical items, $t(20) = 7.25$, $p < .001$, $d = 1.58$, 95% confidence interval (CI) of d [0.93, 2.22]. In the L2 speakers' group, however, virtually no RT difference was noted between the ungrammatical and grammatical items, $t(24) = 0.31$, $p = .76$, $d = 0.06$, 95% CI of d [−0.33, 0.45]. Two sample t-test on GSI revealed a significant difference between the L1 (64 ms) and L2 (3 ms) groups, $t(44) = 4.20$, $p < .001$, $d = 1.24$, 95% CI of d [0.59, 1.85].

Group-Level Analysis

Comparisons between Grammatical and Ungrammatical Sentences (RQ1)

In the L1 speaker group, the left precentral gyrus (i.e., premotor area) was significantly more activated when presented with ungrammatical versus grammatical sentences during the word-monitoring task (cluster size = 34, MNI *x, y, z* coordinates = −40, −2, 32, $t = 4.36$; see Figure 4).

In the L2 speaker group, significantly greater activation was observed in the following two brain regions: left anterior caudate nucleus (cluster size = 30, MNI *x,*

Table 2. Descriptive statistics for L2 speakers

|	*M*	*SD*	*Min*	*Max*
Word-Monitoring Task				
Grammatical RT (ms)	524	137	332	873
Ungrammatical RT (ms)	527	118	351	765
GSI (ms)	3	55	−131	104
Metalinguistic Knowledge (%)	64.75	16.93	37.50	100
SRT (ms)	19.50	15.91	−4.54	51.59
LLAMA_F (raw accuracy score)	24.36	3.35	16.00	28.00

Note: GSI (grammaticality sensitivity index) was computed as follows: RT (ungrammatical sentences)−RT (grammatical sentences).
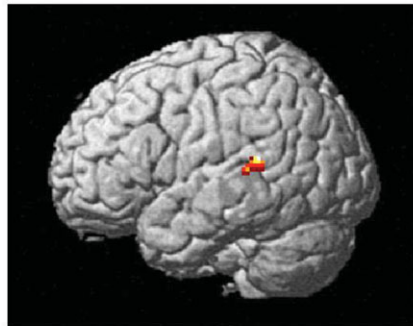
[L1 Group]
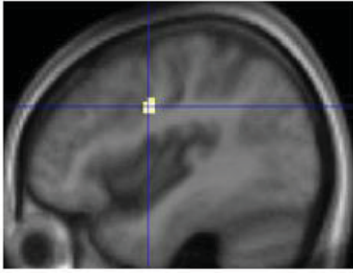
Left premotor area

[L2 Group]
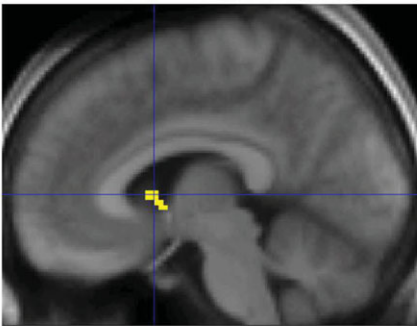
Left caudate nucleus   Left superior temporal gyrus

Figure 4. Brain areas showing greater activation in response to ungrammatical than grammatical sentences during the word-monitoring task *(L1 and L2 Groups)*.

y, z coordinates = −4, 10, 6, *t* = 5.28) and left superior temporal gyrus (cluster size = 45, MNI *x, y, z* coordinates = −60, −36, 10, *t* = 4.48), as illustrated in Figure 4.

Joint Analyses: Comparisons between L1 and L2 Groups (RQ2)
Mixed ANOVA was conducted using SPM12 to compare the brain activation patterns between L1 and L2 groups at the whole brain level. Although significant activation was not detected in any brain region under corrected statistical threshold (family-wise error correction, *p* < .05, cluster-level), for both L1 and L2 groups, the premotor area was more activated in response to ungrammatical sentences than when participants were presented with grammatical sentences under the liberal threshold (*p* < .005, uncorrected, cluster size = 50, MNI *x, y, z* coordinates = −38, −4, 30, *t* = 3.38).

To further clarify the activation patterns in the two groups, region of interest (ROI) analysis was performed targeting two brain areas (premotor cortex and head of left caudate). Mixed ANOVA revealed a significant interaction between group and brain areas, *F*(1.46, 4.05) = 5.05, *p* = .02, ηp^2 = 0.10 (see Appendix A in the Online Supplementary File). In the L1 group, the premotor cortex was activated more strongly than in the L2 group, *p* = .002, *d* = 0.97, 95% CI of *d* [0.34, 1.57]. In contrast, L2 group scans revealed a significantly higher activation in the left caudate compared to the L1 group, *p* = .001, *d* = 1.03, 95% CI of *d* [0.40, 1.63].

Table 3. Results of correlation and multiple regression analyses for L2 speakers

[Correlations]

	1	2	3	4
1. Word-Monitoring Task (GSI)	—	.15	−.12	.44*
2. Metalinguistic Knowledge		—	.16	−.03
3. SRT			—	.03
4. LLAMA_F				—

*p < .05

[Multiple Regression]

Predictor	B	95% CI	β	t	p
Constant	−183.89	[−348.29, −19.5]		−2.33	.03
Metaling. Knowledge	2.67	[−2.89, 8.24]	0.19	1.00	.33
SRT	−0.56	[−1.93, 0.81]	−0.16	−0.85	.41
LLAMA_F	7.38	[0.94, 13.83]	0.45	2.38	.03

Note: See Appendix B in the Online Supplementary File for the scatter plot between GSI and LLAMA_F.

Individual Difference Analysis

Behavioral Data (RQ3)

Table 3 shows the results of correlation and multiple regression analyses for L2 learners. GSI from the word-monitoring task was significantly correlated with LLAMA_F score ($r = .44$, $p = .03$). In the multiple regression results, LLAMA_F was a significant predictor of GSI ($\beta = 0.45$, $p = .03$), while metalinguistic knowledge and SRT scores were not. Although the omnibus model was not significant, $F(3, 21) = 2.31$ $p = .11$, $R^2 = 49.80\%$, Adjusted $R^2 = 24.77\%$, this was most likely due to the redundant predictors. The regression model based solely on LLAMA_F was significant and accounted for a similar amount of variance in the word-monitoring performance (GSI), $F(1, 23) = 5.56$ $p = .03$, $R^2 = 44.10\%$, Adjusted $R^2 = 19.50\%$.

Brain Data (RQ4)

The multiple regression analyses at the whole brain level for L2 speakers revealed that none of the activated brain regions were significantly predicted by any variables.

Discussion

Procedural Memory Activation during Word-Monitoring Task

The first RQ of this study probed into the neurocognitive underpinnings of grammar knowledge measured by a real-time grammar processing (word-monitoring) task. Based on the task design features, it was hypothesized that brain regions responsible for procedural memory, rather than those related to declarative memory, would be recruited more strongly. Consistent with this hypothesis, the whole-brain analysis revealed that one of the regions underlying procedural knowledge (i.e., left anterior caudate nucleus, which is a part of the basal ganglia) was significantly more activated among L2 speakers in response to ungrammatical compared to grammatical sentences in the word-monitoring task. This finding lends support to the claim that word-monitoring task is a fine-grained measure that can tap into implicit knowledge, in

the sense of recruiting procedural system for fluent comprehension of grammar (DeKeyser, 2020; Paradis, 2009; Ullman, 2020).

One brain region outside the basal ganglia—superior temporal gyrus—was also significantly more activated in response to ungrammatical compared to grammatical sentences among L2 speakers. Because this region is not associated with procedural memory system, this result was not expected. Superior temporal gyrus is considered to be implicated in auditory sentence processing (Hugdahl et al., 2003). Because L2 speakers were processing ungrammatical case-marking particles in the auditory sentence in the word-monitoring task, they might have become more alert to ungrammatical relative to grammatical sentences. However, this interpretation may not be tenable given the lack of behavioral sensitivity to errors in the word-monitoring task.

Furthermore, in line with the hypothesis, no systematic association was found between GSI and activation of brain regions associated with declarative memory (e.g., hippocampus, medial temporal lobe). Consistent with the brain-imaging data, no association between GSI and metalinguistic knowledge score was found at the behavioral level. In other words, real-time processing of errors did not seem to preferentially recruit L2 explicit knowledge. Taken together, these findings suggest that GSI may be a good indicator of implicit knowledge use for detecting grammatical errors (whether or not this involved awareness is, however, uncertain from the findings reported here) with limited influence from speeded-up explicit knowledge (Granena, 2013; Suzuki, 2017; Suzuki & DeKeyser, 2015).

The Role of Left Caudate and Premotor Area in Automatization of Grammatical Knowledge: Comparisons between L1 Speakers and L2 speakers

RQ2 focused on the comparison of neural patterns produced by L1 and L2 speakers. It was hypothesized that L1 speakers would activate the frontal region, particularly in the left IFG and the premotor cortex, whereas L2 speakers (whose knowledge is presumably less automatized) would not show the same level of activation in these regions. In contrast, it was expected that L1 knowledge retrieval would rely less on the basal ganglia than accessing L2 knowledge because the basal ganglia is more involved in the earlier phases of procedural learning (Ullman, 2020).

The current findings were in agreement with this contrasting neural pattern for the basal ganglia and the premotor cortex. In L1 speakers that took part in the present study, premotor area was more strongly activated when processing ungrammatical sentences than grammatical sentences in the word-monitoring task.[6] The premotor area was also activated in L2 speakers (with the liberal statistical significance threshold) but to a lesser degree than in L1 speakers. However, the significantly greater activation in the left anterior caudate nucleus (a part of the basal ganglia) was observed among L2 speakers than L1 speakers for contrast between the ungrammatical and the

---

[6]Rather unexpectedly, in the current word-monitoring task, no significant activation of the left IFG was found for the ungrammatical > grammatical contrast. Although no strong explanation for this finding can be provided based on the evidence obtained in the present study, it could be attributable to the fact that the word-monitoring task required no grammaticality judgment. Consequently, participants' attention was more directed to meaning, potentially resulting in a less pronounced role of left IFG in the word-monitoring task than in form-focused tasks such as GJT.
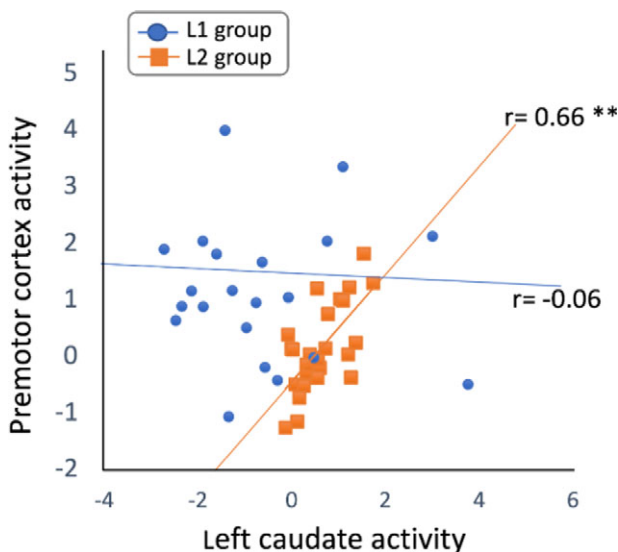
Figure 5. Correlations between left caudate and premotor activity in L1 and L2 groups.

grammatical sentences. This L1–L2 difference suggests that the current L2 speakers' grammatical knowledge was probably less automatized than that of L1 speakers'.

According to extant research on cognitive skill acquisition in general (Ashby & Crossley, 2012; Waldschmidt & Ashby, 2011), the basal ganglia (particularly, head of caudate) plays a major role in the earlier skill development stages. Once automaticity in a target skill has been developed, the basal ganglia is no longer activated, as cortico-cortical connections, including supplementary motor and premotor regions, have been established. Indeed, the L1 speakers that took part in current study might have already reached asymptotic state in terms of automatization, which would manifest as absence of significant left caudate activation, while L2 speakers are more likely to be still in the earlier skill development phase and have not yet reached the end stage of automatization.

To explore the potential link between left caudate and premotor cortex activation, a post-hoc correlation analysis was conducted on the activations of the two ROIs (i.e., head of left caudate and premotor cortex) obtained through the joint analysis of L2 and L1 speakers. Intriguingly, the findings revealed a significant positive relationship between the premotor cortex and the left caudate activation for the L2 group ($r = .66, p < .001$), but not for the L1 group ($r = -.06, p = .79$), as illustrated in Figure 5. This suggests that L2 speakers in whom the brain region primarily recruited in the earlier phases of procedural learning (left caudate) is more strongly activated are likely to recruit the region that is more important for the later stage (premotor cortex) in a more similar way to L1 speakers. In other words, the few L2 speakers who showed higher activation in both left caudate and premotor cortex might have automatized their grammatical knowledge to a greater extent than the rest of the L2 group. This positive association between left caudate and premotor activation may be consistent with the aforementioned cognitive neuroscience view of automaticity (Ashby & Crossley, 2012), suggesting that the basal ganglia (procedural memory) may serve as a mediating system

to establish the cortico-cortical representation (e.g., premotor cortex) of automaticity in L2 knowledge.

The Role of Explicit and Implicit Learning Aptitude in L2 Grammar Acquisition: Conflicting Evidence

In this work, an individual difference approach was taken to investigate the extent to which cognitive aptitude for explicit and implicit learning (LLAMA_F and SRT) predict sensitivity to grammatical errors in the word-monitoring task at the behavioral and neural levels among L2 speakers (RQs 3 and 4, respectively). Even though systematic relationship between GSI and implicit aptitude was hypothesized in the current study, explicit, rather than implicit, aptitude emerged as a significant predictor of word-monitoring task performance at the behavioral level.

The lack of association between GSI and implicit aptitude is inconsistent with the prior research findings. Specifically, both Granena (2013) and Suzuki and DeKeyser (2015) consistently demonstrated a significant relationship between GSI and SRT among adult naturalistic L2 learners.[7] The insubstantial role of implicit aptitude found in the present study may in part be due to shorter length of residence (LOR) or lesser amount of naturalistic L2 exposure compared to the participants in the aforementioned studies. The mean LOR of 30 months in the current study sample was considerably shorter than 101 months reported for adult L2 Spanish speakers that took part in Granena's (2013) ultimate-attainment study, and 55 months noted by Suzuki and DeKeyser (2015) for a subset of the L2 Japanese learner group (long-LOR) in their study. It can thus be speculated that, as their L2 exposure accumulates in this immersion context, the current study participants may start to develop their grammatical knowledge using implicit learning systems (DeKeyser, 2020; Paradis, 2009; Suzuki, 2017), which may result in a significant association between their GSI and SRT scores.

However, a systematic relationship between explicit aptitude and GSI was detected. Although unexpected, this finding may not be inconsistent with the neuroimaging study results reported by Morgan-Short et al. (2015a). According to these authors, declarative memory was implicated in significant activation of the brain region related to L1 processing (i.e., left IFG) in the earlier stages of grammar learning under the ASL paradigm. Both declarative and procedural model and skill acquisition theory posit that declarative memory/knowledge plays a crucial role in the initial stages of L2 acquisition, as well as its further proceduralization and automatization of L2 knowledge. Hence, greater language analytic ability might have allowed the current cohort of L2 learners to engage in a deliberate and systematic use of specific grammatical structures more effectively in naturalistic settings (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; Suzuki & DeKeyser, 2017).

Nevertheless, in contrast to the longitudinal intervention design employed by Morgan-Short et al. (2015a), the current cross-sectional design makes it difficult to identify *when* different types of aptitude are utilized for acquiring explicit and implicit knowledge. Because the current participants have already spent several years learning L2, when completing the word-monitoring task for this study, it is uncertain whether the knowledge they retrieved was identical to that they initially acquired by recruiting their explicit aptitude (cf., Suzuki & DeKeyser, 2017). Given that explicit language

---

[7]A recent study by Godfroid and Kim (2021) also reported a significant positive correlation between GSI and the alternating SRT task score.

aptitude may be instrumental in the earlier learning phases but implicit language aptitude may play a more important role for advanced learners (Li & DeKeyser, 2021), a longitudinal study is needed to shed light on the role of explicit and implicit aptitudes in different stages of learning in naturalistic L2 settings.

The significant role of explicit aptitude that emerged from the present study may also be attributable to the sample characteristics. Because the duration of learners' immersion experience (i.e., LOR) was shorter than that considered in the previous studies focusing on the acquisition of implicit knowledge (e.g., Granena, 2013; Suzuki & DeKeyser, 2015; Suzuki, 2017), this is likely to affect consistency in their implicit knowledge use. In addition, more than half the current sample held a bachelor's degree in Japanese and were thus probably more linguistically oriented than an average L2 learner. Thus, their background might have prevented them from reliably deploying implicit knowledge, possibly due to the competition between more robust explicit knowledge and still-developing implicit knowledge. This interpretation is plausible because the current L2 speakers failed to show sensitivity to grammatical errors (mean GSI = 3). These findings constitute conflicting evidence for the claim that implicit knowledge is accessed during the word-monitoring task. The word-monitoring task (itself) cannot be simply considered as an implicit or explicit knowledge test, as its completion is likely to recruit different types of knowledge depending on learners' proficiency and experience. In future research, administering the word-monitoring task to more advanced L2 learners with longer lengths of residence, as typically recruited in ultimate attainment research (e.g., Granena, 2013), may help resolve these conflicting findings.

It is also worth noting that none of the individual difference variables were significant predictors of the L2 brain imaging results. Because brain response is purported to be a more direct measure of cognitive processing than RT scores, it is puzzling that the role of L2 speakers' explicit aptitude was evident in the behavioral analysis, but not in the brain analysis. The whole-brain analysis revealed that the left caudate nucleus was more highly activated when L2 learners processed ungrammatical (as compared to grammatical) sentences in the word-monitoring task, indicating that procedural memory underlies the sensitivity to grammatical errors. This observation may indicate that the shift from reliance on the declarative system to the procedural system has already occurred in the brain (Paradis, 2009; Ullman, 2020). It is thus speculated that most of the L2 learners that took part in this study might have already transitioned to the procedural system for their L2 comprehension at the neural level, due to which no significant relationship was noted between explicit aptitude and declarative memory in the brain-level analysis. Nonetheless, their grammatical knowledge needs to be fine-tuned further through extensive L2 exposure and use. In the current L2 sample, this fine-tuning process (e.g., automatization, consolidation of implicit knowledge) might not have been sufficiently established to be observable in behavioral performance tests. As a result, the L2 learners might not have attained automaticity to the same degree as L1 speakers, as indicated by the weaker premotor cortex activation in this group.

Exploratory Analyses based on the Awareness Criterion: Insights from the Retrospective Questionnaire

Because the awareness criterion was not the focus of the present study, it is yet to be determined whether grammatical knowledge, measured by the word-monitoring task,

was indeed "implicit" in the strict sense of lack of awareness. In our view, it seems extremely difficult for any introspection method to sufficiently capture the state of awareness during word-monitoring task completion. For our exploratory attempt, however, a retrospective questionnaire was administered immediately after the word-monitoring task to examine the participants' noticing of any errors in the items presented to them. While all 21 L1 speakers noticed the ungrammaticality in the stimuli, only 52% of the L2 speakers (13/25) were aware of these errors.

Further exploratory analyses were conducted to compare both behavioral and neural responses between L2 speakers who reported noticing ($n = 13$) and those who did not ($n = 12$). Notable findings are highlighted here (see Appendix C in the Online Supplementary File for the full retrospective questionnaire results). At the behavioral level, GSI was significantly higher in the noticing group than in the nonnoticing group, $t(23) = 3.04$, $p = .006$, $d = 1.22$, 95% CI of d [0.33, 2.03]. At the neural level, the left caudate and the right hippocampus were significantly more activated in the noticing group than in the nonnoticing group with the liberal statistical threshold ($p < .005$, uncorrected). Taken together, these findings suggest that L2 speakers who noticed the errors showed higher sensitivity to the grammatical errors within a time-locked window (a few hundred milliseconds) than learners who did not report noticing errors. At the same time, they recruited both procedural and declarative memory more strongly than those who did not report noticing grammatical errors. It is difficult to discern when L2 speakers became aware of the grammatical errors. While error registration without awareness might have prompted conscious awareness *after* the point of ungrammaticality, explicit knowledge could have been accessed during the word-monitoring task. Given a small number of participants in each subgroup and, critically, an overly coarse retrospective questionnaire instrument, these interpretations are only speculative.

The results yielded by exploratory analyses using the awareness criterion may be crucial. That is, when completing the word-monitoring task, L2 learners recruited multiple processes that are not limited to the declarative and procedural memory systems (e.g., the right middle/inferior temporal cortex and the right fusiform gyrus, see Appendix C). These complex patterns indicate that the awareness criterion (at least that measured by a coarse retrospective method) might not be as useful as is generally assumed for distinguishing L2 knowledge. From a cognitive neuroscience perspective, consciousness is a poor criterion for distinguishing between declarative and procedural memory (Henke, 2010). Therefore, a more parsimonious and plausible explanation should also be sought for SLA research. If a goal of L2 research is to identify the nature of robust L2 knowledge that supports accurate and fluent use, the criterion of automaticity may be a more valuable operational definition of grammatical knowledge that can be linked to multiple memory systems (declarative-procedural and explicit-implicit) as well as multiple behavioral criteria (e.g., speed, stability, efficiency) that can be measured more comprehensively and straightforwardly.

Limitations and Suggestions for Future Research Directions

Based on the current study findings, as well as its inherent limitations, several suggestions for future research directions can be proposed. First, while the number of L2 learners recruited for the present study was relatively large compared to the samples employed in other L2 fMRI studies (e.g., Morgan-Short et al., 2015a), the sample size is still small for a behavioral study. Hence, in future research, a greater number of L2

learners with different backgrounds (e.g., varying lengths of residence and learning experience) should be studied to evaluate the generalizability of the current findings.

Second, while a word-monitoring task was adopted in the current study as a measure of real-time grammar processing, exposure to ungrammatical sentences could have raised participants' awareness of grammatical structures and could have possibly led some individuals to start ignoring ungrammaticality as the task proceeded. To eliminate these potential risks, employing a visual-world (eye-tracking while listening) task, which does not require any ungrammatical sentences to assess real-time grammar processing, may be more appropriate for this type of investigation (Suzuki, 2017).

Third, as the temporal resolution of the fMRI technique is poor, a different neural imaging method such as electroencephalography (EEG) can be adopted instead to investigate automatic and implicit L2 processing (Morgan-Short et al., 2015b). In extant studies employing fMRI and EEG data, form-focused tasks such as GJTs have been extensively used. While this is the first fMRI study involving word-monitoring task, EEG has never been applied to this real-time grammar processing task. For particularly ambitious investigations, fMRI and EEG can be combined to further scrutinize the nature of L2 knowledge and processing measured by various tasks including (timed) GJTs and word-monitoring tasks.

Last, as the aptitude measures (LLAMA_F and SRT) adopted in the present study were not particularly reliable, this might have attenuated the strength of associations between aptitude and linguistic knowledge. Furthermore, a set of cognitive aptitudes for explicit and implicit learning can be expanded in future research (e.g., Li & DeKeyser, 2021; cf., Perruchet, 2021). For instance, the long-term memory synonym test proposed by Granena (2019) can also be adopted as a potential measure of implicit language aptitude. It is therefore anticipated that further advancements in the understanding of the cognitive aptitude constructs, combined with greater instrument reliability, will impact the interpretations of the current findings, as well as those yielded by prior studies.

Conclusions

The goal of the current study was to shed light on the mechanisms underpinning explicit and implicit leaning and knowledge among adult L2 learners. For this purpose, two related issues were investigated. First, fMRI investigations were performed to scrutinize the validity of available evidence related to the types of grammar knowledge measured by a real-time grammar processing task. Neuroimaging results showed that, when detecting grammatical errors in auditory sentence in real time during the word-monitoring task, L2 speakers recruited basal ganglia (procedural memory), not hippocampus or medial temporal lobe structures (declarative memory), more strongly relative to the processing of grammatical sentences. Hence, the RT difference score (i.e., GSI) derived from the word-monitoring task arguably indicates implicit knowledge, rather than speeded-up explicit knowledge. However, the current L2 learners' grammatical knowledge may have been less consistent and automatized than that of L1 speakers, as indicated by the limited behavioral sensitivity to errors in the word-monitoring task and weaker activation of the premotor cortex in the former group. These neuroimaging findings compliment the interpretations of previous behavioral results offered by other authors (Granena, 2013; Suzuki, 2017; Suzuki & DeKeyser, 2015; Vafaee et al., 2017), suggesting that neuroimaging data is instrumental in elucidating the nature of L2 knowledge.

Second, to further probe the putative relationships between grammatical knowledge and explicit-implicit aptitudes, a neurocognitive individual difference approach was

employed in the present study. None of the individual difference variables were significant predictors of brain activation patterns. In contrast, behavioral data analysis indicated that explicit aptitude significantly predicted real-time sensitivity to errors (GSI) during the word-monitoring task. This may underscore the value of explicit analytic learning ability in using relevant declarative knowledge and initiating proce-duralization of L2 knowledge, which lays the foundation for further automatization in a naturalistic context. Nonetheless, the evidence provided here is insufficient for drawing any firm conclusions on L2 developmental processes. Clearly, additional longitudinal neuroimaging research, as well as replication of the current findings, is needed to resolve fundamental issues surrounding explicit and implicit learning and knowledge.

Supplementary Materials. To view supplementary material for this article, please visit http://doi.org/10.1017/S0272263122000043.

Data Availability Statement. The experiment in this article earned Open Materials and Open Data badges for transparent practices. The materials are available at https://www.iris-database.org/iris/app/home/detail?id=york%3a939985&ref=search, and the data are available at https://doi.org/10.7910/DVN/VPYYP9.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481–509. https://doi.org/10.1017/S027226310808073X

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, *59*, 249–306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Ashby, F. G., & Crossley, M. J. (2012). Automaticity and multiple memory systems. *WIREs Cognitive Science*, *3*, 363–376. https://doi.org/10.1002/wcs.1172

Bokander, L., & Bylund, E. (2019). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning, Advanced Access.* https://doi.org/10.1111/lang.12368

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the marsbar toolbox for SPM 99. *Neuroimage*, *16*, S497. https://doi.org/10.1016/S1053-8119(02)90013-3

Buffington, J., Demos, A. P., & Morgan-Short, K. (2021). The reliability and validity of procedural memory assessments used in second language acquisition research. *Studies in Second Language Acquisition*, *43*, 635–662. https://doi.org/10.1017/S0272263121000127

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Newbury House.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–534. https://doi.org/10.1017/S0272263100004022

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & H. M. Long (Eds.), *The handbook of second language acquisition* (pp. 312–348). Blackwell Publishers.

DeKeyser, R. M. (2009). Cognitive-psychological processes in second language learning. In H. M. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 119–138). Wiley-Blackwell.

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). Routledge.

DeKeyser, R. M. (2020). Skill acquisition theory. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 83–104). Routledge.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*, 1325–1335. https://doi.org/10.1016/j.neuroimage.2004.12.034

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

Forster, K. I., & Forster, J. C. (2003). Dmdx: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. https://doi.org/10.3758/BF03195503

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, *16*, 1709–1717. https://doi.org/10.1093/cercor/bhj106

Godfroid, A. (2016). The effects of implicit instruction on implicit and explicit knowledge development. *Studies in Second Language Acquisition*, *38*, 177–215. https://doi.org/10.1017/S0272263115000388

Godfroid, A., & Kim, K. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*, 606–634. https://doi.org/10.1017/S0272263121000085

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgements measure distinct types of knowledge. *Studies in Second Language Acquisition*, *37*, 269–297. https://doi.org/10.1017/S0272263114000850

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 665–703. https://doi.org/10.1111/lang.12018

Granena, G. (2016). Cognitive aptitudes for implicit and explicit learning and information-processing styles: An individual differences study. *Applied Psycholinguistics*, *37*, 577–600. https://doi.org/10.1017/S0142716415000120

Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, *41*, 313–336. https://doi.org/10.1017/S0272263118000256

Granena, G. (2020). *Implicit language aptitude.* Cambridge University Press.

Hashimoto, R., & Sakai, K. L. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron*, *35*, 589–597. https://doi.org/10.1016/S0896-6273(02)00788-2

Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. *Nature Reviews Neuroscience*, *11*, 523–532. https://doi.org/10.1038/nrn2850

Hugdahl, K., Thomsen, T., Ersland, L., Morten Rimol, L., & Niemi, J. (2003). The effects of attention on speech perception: An fMRI study. *Brain and Language*, *85*, 37–48. https://doi.org/10.1016/S0093-934X(02)00500-X

Isbell, D. R., & Rogers, J. (2021). Measuring implicit and explicit learning and knowledge. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 304–313). Routledge.

Jiang, N. (2011). *Conducting reaction time research in second language studies.* Routledge.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Li, S., & DeKeyser, R. M. (2021). Implicit language aptitude: Conceptualizing the construct, validating the measures, and examining the evidence: Introduction to the special issue. *Studies in Second Language Acquisition*, *43*, 473–497. https://doi.org/10.1017/S0272263121000024

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., … Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–566. https://doi.org/10.1111/lang.12011

Marsden, E., Mackey, A., & Plonsky, L. (2016). The IRIS repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1–21). Routledge.

Meara, P. M. (2005). LLAMA language aptitude tests: The manual. Lognostics. http://www.lognostics.co.uk/tools/llama/llama_manual.pdf

Morgan-Short, K., Deng, Z., Brill-Schuetz, K. A., Faretta-Stutenberg, M., Wong, P. C. M., & Wong, F. C. K. (2015a). A view of the neural representation of second language syntax through artificial language learning under implicit context of exposure. *Studies in Second Language Acquisition*, *37*, 383–419. https://doi.org/10.1017/S0272263115000030

Morgan-Short, K., Faretta-Stutenberg, M., & Bartlett-Hsu, L. (2015b). Contribution of event-related potential research into explicit and implicit second language acquisition. In P. Rebuschat (Ed.), *Investigating implicit and explicit language learning* (pp. 349–386). John Benjamins Publishing Company.

Opitz, B., & Kotz, S. A. (2012). Ventral premotor cortex lesions disrupt learning of sequential grammatical structures. *Cortex*, *48*, 664–673. https://doi.org/10.1016/j.cortex.2011.02.013

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins Publishing Company.

Perruchet, P. (2021). Why is the componential construct of implicit language aptitude so difficult to capture? A commentary on the special issue. *Studies in Second Language Acquisition*, *43*, 677–691. https://doi.org/10.1017/S027226312100019X

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595–626. https://doi.org/10.1111/lang.12010

Sakai, K. L. (2005). Language acquisition and brain development. *Science*, *310*, 815–819. https://doi.org/10.1126/science.1113530

Slotnick, S. D. (2017). Cluster success: fMRI inferences for spatial extent have acceptable false-positive rates. *Cognitive Neuroscience*, *8*, 150–155. https://doi.org/10.1080/17588928.2017.1319350

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology*, *7*, a021667.

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*, 1229–1261. https://doi.org/10.1017/S014271641700011X

Suzuki, Y., & DeKeyser, R. M. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*, 860–895. https://doi.org/10.1111/lang.12138

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*, 747–790. https://doi.org/10.1111/lang.12241

Tagarelli, K. M., Shattuck, K. F., Turkeltaub, P. E., & Ullman, M. T. (2019). Language learning in the adult brain: A neuroanatomical meta-analysis of lexical and grammatical learning. *Neuroimage*, *193*, 178–200. https://doi.org/10.1016/j.neuroimage.2019.02.061

Ullman, M. T. (2020). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed., pp. 128–161). Routledge.

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. https://doi.org/10.1017/S0272263115000455

Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage*, *56*, 1791–1802. https://doi.org/10.1016/j.neuroimage.2011.02.011

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

Yang, J., & Li, P. (2012). Brain networks of explicit and implicit learning. *PloS One*, *7*, e42993. https://doi.org/10.1371/journal.pone.0042993

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*, 831–856. https://doi.org/10.1017/S0272263118000141