



RESEARCH ARTICLE  

# Unpacking L2 explicit linguistic knowledge and online processing of the English modals *may* and *can*: A comparison of acceptability judgments and self-paced reading\*

Nadia Mifka-Profozic<sup>1</sup> , David O'Reilly<sup>1</sup>  and Leonarda Lovrovic<sup>2</sup><sup>1</sup>Centre for Advanced Studies in Language and Education (CASLE), Department of Education, University of York, United Kingdom; <sup>2</sup>University of Zadar, Croatia**Corresponding author:** Nadia Mifka-Profozic; Email: [nadia.mifka-profozic@york.ac.uk](mailto:nadia.mifka-profozic@york.ac.uk)

(Received 04 June 2022; Revised 03 July 2023; Accepted 21 August 2023)

## Abstract

The present study uses self-paced reading as a measure of online processing and an acceptability judgement task as a measure of offline explicit linguistic knowledge, to understand L2 learners' comprehension processes and their awareness of subtle differences between the modal auxiliaries *may* and *can*. Participants were two groups of university students: 42 native speakers of English and 41 native speakers of Croatian majoring in L2 English. The study is part of a larger project that has provided empirical evidence of the two modals, *may* and *can*, being mutually exclusive when denoting ability (*can*) and epistemic possibility (*may*) but equally acceptable in pragmatic choices expressing permission. The present results revealed that L1 and L2 speakers rated the acceptability of sentences in offline tasks similarly; however, L2 learners showed no sensitivity to verb–context mismatches in epistemic modality while demonstrating sensitivity when processing modals expressing ability. Implications for L2 acquisition of modals and future research are discussed.

## Introduction

Research into second language (L2) acquisition has long been debating the nature of the target language knowledge that L2 learners develop. Due to both its theoretical importance and the difficulty of knowledge measurement (Bowles, 2011; DeKeyser, 2003; Ellis, 2005; Norris & Ortega, 2013), this exploration has driven much L2 acquisition research. Although linguists differ in views on the sources and development of L2 knowledge, most would agree that explicit and implicit linguistic knowledge are

---

\*The online version of this article has been updated since original publication. A notice detailing the change has also been published

two separate constructs that develop in distinct parts of the brain and are accessed through different processes (Paradis, 2009). Explicit knowledge in L2 is described as conscious, imprecise, unstable, unsystematic, and as common in instructed L2 learners, whereas implicit knowledge is unconscious, intuitive, tacit, systematic, and easily accessible (Bowles, 2011; Ellis, 2005; Ellis et al., 2009).<sup>1</sup> Because implicit knowledge is required for linguistic competence (Norris & Ortega, 2013; Rebuschat, 2013), it is crucial to investigate this type of knowledge, both for its theoretical importance and for understanding L2 development and its implications for pedagogy.

In L2 research, the most widely used instruments to measure L2 linguistic knowledge have been grammaticality and acceptability judgment tasks (Plonsky et al., 2020). Judgment tasks without time constraints tend to be approximated with the measurement of explicit knowledge (Ellis, 2005; Ellis et al., 2009; Godfroid et al., 2015; Rebuschat, 2013; Zhang, 2015) where time is needed to retrieve an explicitly known, pedagogical rule and related information. The same tasks with time constraints are generally treated as tests of implicit knowledge where immediate, fast performance is demonstrated. However, the role of time pressure has been heavily debated. DeKeyser (2003) and Suzuki and DeKeyser (2017), for example, suggested that explicit knowledge can also be accessed rapidly after long practice, so time pressure itself cannot entirely block the influence of explicit knowledge during test performance. While maintaining the distinction between implicit knowledge and automatized explicit knowledge as two separate constructs, Suzuki and DeKeyser (2017) and Vafaei et al. (2017) showed timed grammaticality judgment tasks (GJTs) to be a rather crude measure involving focus on form and the addition of time pressure insufficient for engaging implicit knowledge. In a recent study, Maie and Godfroid (2022) reported the findings of an eye-tracking study that also questioned time-pressured AJT as a measure of automated processing tapping implicit knowledge. Their findings suggested that time pressure suppressed both the controlled and the automatic processes and that different L2 speakers were affected by time pressure in different ways.

As is increasingly recognized, more appropriate task types need to be used to measure real-time processing *during* comprehension and indirectly, implicit linguistic knowledge. Such tasks, along with other psycholinguistic methods, include self-paced reading (SPR), word monitoring, and visual-world task, which Suzuki (2017) proposed as measures of implicit knowledge and distinct from automatized explicit knowledge.

Along these lines, the present study employed SPR to examine whether upper-intermediate L2 users have developed mental representations of modals *may* and *can* similar to native speakers, which can be demonstrated only in online tasks, during real-time processing. So far, studies that have attempted to measure implicit knowledge or processing have typically focused on morphosyntactic features and to a lesser extent on lexical aspects of the target language. Processing of modals in L2 English has been a neglected area in research to date; thus, the present study set out to fill in this gap.

We adopted the psycholinguistic approach of sentence processing, which is based on the view that real-time sentence processing offers an insight into how grammar acquisition of an L2 occurs. In SPR studies, a baseline condition containing a set of well-formed stimuli is examined in relation to an experimental condition containing the same set of stimuli with an anomaly on the feature being tested. In the present study, this is the alternation of two modals, *may* and *can*, which are typically presented as a

<sup>1</sup>The long-lasting debate around the explicit/implicit interface and whether one type of knowledge can translate into the other is still ongoing, but it is outside the focus of the current study (for detailed accounts, see Ellis, 2005; Ellis et al., 2009; Suzuki & DeKeyser 2017; Zhang 2015).

pair in L2 English textbooks and are known as being especially challenging to master due to their multiple meanings and functions.

We compare L2 speakers' online processing of sentences containing the modals *may* and *can* with their acceptability ratings using the same sentences in a judgment task. The significance and originality of the present research lies in the fact that this is the first study to investigate L2 comprehension of the two modals by introducing modality as a grammatical category into the implicit/explicit L2 research paradigm.

## Background

### *Modality and modal auxiliaries*

Modality is a prime conceptual domain that enables people to comprehend and produce meanings that are not related to facts and reality, such as belief, imagination, possibility, necessity, inferred certainty, obligation, and permission (Coates, 2014; Leech, 2003; Traugott & Dasher, 2004). Modality exists in all natural languages, but different languages use different means to express modal meaning (Bybee et al., 1994). In English, modality is primarily expressed by the modal auxiliaries *can*, *may*, *will*, *shall*, *could*, *might*, *would*, *should*, and *must*. There are also lexical expressions such as adverbs (e.g., *possibly*, *probably*, *likely*, etc.), verbs (e.g., *believe*, *think*, etc.), and whole sentences—that is, conditional sentences (Stewart et al., 2009). English modal auxiliaries are well known for their peculiarities (Bybee et al., 1994; Coates, 2014; Leech, 2003; Palmer, 1986, 2003) such as their differences from English lexical verbs in terms of interrogative and negative formation, their lack of nonfinite forms (infinitive and participles), and their complexities in terms of semantic meaning and pragmatic interpretation. Most classifications of modals make the basic differentiation between epistemic and nonepistemic (root) modals where the same form can function differently in different contexts. The following example demonstrates the dual nature of the modal *may*:

- (1) *They may have arrived*—we do not know whether they have arrived or not: there is only a possibility that they have arrived (epistemic possibility).
- (2) *You may borrow this book*—meaning ‘you are allowed to borrow this book’ or ‘you are permitted to borrow this book’ (nonepistemic meaning of the same modal expressing permission).

However, it is also possible to use the modal *can* to convey the same meaning as in (2):

- (3) *You can borrow this book*.

The above examples demonstrate some of the “fuzziness” or “indeterminacy” (Coates, 2014) of modal usage in the English language, which presents significant challenges to learners of English as an L2. These challenges are deep rooted especially in the usage of two among the most frequently used modals, *may* and *can* (Biber et al., 1999; Brezina & Gablasova, 2015).<sup>2</sup> This suggests that they need to be examined closely,

<sup>2</sup>In the New General Service List (Brezina & Gablasova, 2015) of the 2,500 most frequent English words, *can* is ranked 41st and *may* 85th (*could*, ranked 52nd, is the only modal ranked higher than *may*). In Longman's Grammar (Biber et al., 1999), based on the Longman Spoken and Written English corpus (LSWE), the overall frequencies of *can* and *may* are, respectively, 2,500 and 1,000 occurrences per million words. *Can* is relatively common in all registers, whereas *may* is less common in spoken conversation but extremely common in academic prose. In academic prose, *can* has less than 500 occurrences per million words for expressing permission, slightly less than 1,500 occurrences for root possibility, and slightly more than 1,500

the main reason for making them the focus of our analysis. *May* and *can* are commonly paired together in textbooks for L2 English learners (Bolinger, 1989; Coates, 2014), and they are usually discussed as a pair in theoretical linguistics. The fact that they are interchangeable in some but not all situations is likely to make their usage additionally puzzling for L2 learners. In terms of research, these two modals present a dichotomy that allows for an examination at the level of sentence processing and detection of the effect that syntactic and semantic anomalies may have when they are used in mismatching contexts.

### *Syntactic, semantic, and pragmatic functions of may and can*

In our investigation we adopt Bybee's descriptive functionalist framework, which is based on a crosslinguistic and diachronic perspective focusing on the semantic content of grammatical categories. In describing the functions and sources of modals, Bybee et al. (1994) identified four types of modals according to their source of modality: (a) *agent oriented*, which includes obligation, necessity, ability, and desire; (b) *epistemic*, which is concerned with the speaker's knowledge or belief, expressing possibility, probability, and inferred certainty; (c) *speaker oriented*, which allows the speaker to impose certain conditions on the addressee, such as commands, demands, requests, permissions, etc.; and (d), subordinating moods, which involve the same forms as those used to express the above modalities to mark subordinate clause verbs.

We selected some of the most frequently used meanings (Biber et al., 1999; Coates, 2014) for agent-oriented, epistemic, and speaker-oriented modalities and omitted the subordinate moods, as the latter only appear in subordinate clauses and are always a secondary source of modality. For agent-oriented modality, which "reports the existence of internal and external conditions on an agent with respect to the completion of the action expressed in the main predicate" (Bybee et al., 1994, p. 177) *ability* (including the ability to sense) is selected, as in the examples (4) and (5):

- (4) *She can speak four languages.*
- (5) *I can smell something burning.*

For epistemic meaning, epistemic possibility is selected, as in (6), which indicates that the speaker is not entirely confident that a proposition is true.<sup>3</sup>

- (6) *He may have forgotten my address.*

For speaker-oriented modality, which includes the meanings of imperative, prohibitive, optative, hortative, admonitive, and permissive, and does "not report the existence of conditions on the agent, but rather allow(s) the speaker to impose such conditions on the addressee" (p. 179), asking and giving permission and offers are selected, as in (7) and (8):

- (7) *You may use only the main entrance to the building.*
- (8) *You may/can come with me this way.*

---

for ability. The frequency of *may* is less than 500 occurrences for permission but very high (almost 3,000 occurrences per million words) when expressing epistemic possibility.

<sup>3</sup>Epistemic possibility, which is framed internally or subjectively, should be kept distinct from root possibility that is regulated by external circumstances (Bybee et al., 1994; Coates, 2014; Lyons, 1977).

Matching the modals *may* and *can* with the meanings of ability, epistemic possibility, and permission in relevant contexts allows us to clearly identify and differentiate the semantic, syntactic, and pragmatic functions of the two modals. In a previous SPR experiment with English native speakers (Mifka-Profozic et al., 2020), it was found that the reading penalty due to ungrammatical use of *can* in the context of epistemic possibility could be clearly distinguished from semantically ambiguous use of *may* in the context of ability expression. In the same experiment with L1 speakers, both modals were shown to be equally acceptable to express permission. The choice of one or the other in the latter case depended on pragmatic preferences. It is important to emphasize though, that pragmatic preference is a type of distinction that is qualitatively different from the differences that stem from syntactic or semantic origins.

### **Universal path of modal acquisition**

Uniqueness of the modality domain is, to some extent, rooted in the historical development of language, which is well documented in empirical investigations of world languages and studies on grammaticalization (Bybee et al., 1994; Dittmar & Terborg, 1991; Giacalone Ramat, 1992; Traugott & Dasher, 2004). Grammaticalization as a process of linguistic change over time explains how linguistic units may develop out of lexical items and become more subject to the rules of grammar. Diachronically, the development of modal meaning follows the route from nonepistemic (agent-oriented and speaker-oriented) modality denoting externally determined situations towards meanings denoting internally (perceptually, cognitively) defined situations (Traugott & Dasher, 2004). Working with data from numerous world languages, Bybee et al. (1994) demonstrated the path of development from modal expression of physical ability and mental ability to root possibility and permission on the one hand and to epistemic possibility on the other. This change is explained as a metaphorical extension, a shift to a different domain: from externally imposed meaning to internal domains where epistemic meanings are encoded.

Literature on L1 English acquisition—that is, research recording child language production (e.g., Wells & Nicholls, 1985)—has extensively documented an acquisition sequence suggesting that nonepistemic modal meaning is acquired earlier than epistemic meaning. Furthermore, experiments testing child comprehension (e.g., Ozturk & Papafragou, 2015; Papafragou & Ozturk, 2006) have shown that older children (7- and 9-year-olds) perform better than 4- and 5-year-olds on tasks involving epistemic possibility and epistemic necessity.

From the available research, it appears that L2 English acquisition of modal auxiliaries exhibits a very similar “nonepistemic before epistemic” acquisitional sequence (Dittmar & Terborg, 1991; Giacalone Ramat, 1992). For example, Giacalone Ramat (1992) studied the acquisition of L2 Italian by learners from various L1 backgrounds (Chinese, Tigrinya, Persian, German, and English). At early stages of acquisition only nonepistemic use of modal verbs was observed in grammaticalized, inflected verb forms, whereas epistemic meaning was expressed by epistemic adverbs such as *forse* and *magari* (Italian for “perhaps” or “maybe”). Use of basic formulaic expressions such as *non (lo)so* (“I don’t know”) was also observed as a substitute for a modal, indicating “zero probability” (p. 312). In a more recent study, Granget et al. (2018) confirmed the comprehension advantage for adult learners of L2 French epistemic modality in comparison with child and adolescent L2 learners.

In short, studies on grammaticalization and individual development suggest that modality in both L1 and L2 language acquisition, as well as in diachronic processes,

develops from pragmatic and lexical means to grammaticalized verb forms. This is explained by the fact that at early stages of acquisition language users do not have the grammatical (morphological) means to mark temporal and modal relations. In these processes, as research shows, nonepistemic meanings precede the acquisition of epistemic meaning.

### *Research into L2 English acquisition of modal auxiliaries*

The challenges related to L2 acquisition of English modals are shared, to a certain degree, among all L2 learners, although the scope of difficulty may depend on similarities and differences between L1 and L2. In particular, the modal polysemy presents a mapping and learning problem that involves matching a single lexeme to multiple meanings and functions. In turn, a single meaning can be covered by multiple lexemes. In L2 research, studies carried out to investigate the use of English modals by L2 learners (e.g., Ayoun & Gilbert, 2017; Gibbs, 1990; Hinkel, 2009) have demonstrated that both EFL and ESL learners experience persistent difficulties when attempting to use English modals. In an examination of Punjabi primary and secondary pupils' acquisition of English modality, Gibbs (1990) focused on four modals: *can*, *could*, *may*, and *might*. The study confirmed the universally ascertained order of modal acquisition, with nonepistemic meanings (ability, permission, and root possibility) acquired earlier than the hypothetical and epistemic possibility meanings.

Despite much research conducted to explore the use of L2 English modals, almost all previous studies have focused on production rather than comprehension. A different perspective, with an interest in comprehension, was offered in a small-scale SPR study with L1 Croatian advanced learners of L2 English (Mifka-Profozic, 2017). The study compared the L2 online processing of modals *may* and *can* with L1 English speakers' processing, and detected differences between the two groups in the processing of epistemic possibility. This study was perhaps the first to investigate comprehension of L2 modals at the level of sentence processing. The present study builds on those findings and fills the existing gap in research by introducing an acceptability judgment task (AJT) to compare offline ratings with the performance on an online SPR task.

### *Online and offline tasks to determine the status of learner knowledge*

Processing investigations are important, since research evidence suggests that only tasks involving real-time, online comprehension, where learners' attention is entirely focused on meaning, make the indirect assessment of implicit knowledge possible (Suzuki, 2017; Suzuki & DeKeyser, 2015, 2017; Vafae et al., 2017). This is because in online processing for meaning, the conscious access to learners' explicit knowledge is precluded. Support for this view is found in theories of comprehension and in empirical evidence showing that sentence processing is incremental, which means that syntactic analysis is computed immediately on each word before the next word is encountered (Jegerski, 2012, 2014; Keating & Jegerski, 2015). In semantic analysis, the process is slightly different because, here, context plays an important role (Altmann & Steedman, 1988) while each ensuing word of a sentence is processed and checked against the previous context to facilitate interpretation and possible lexical ambiguity resolution.

In online processing for meaning language users interpret sentences word-by-word while reading, rather than at the end of sentence (Just et al., 1982). Research on monolingual sentence processing shows that native speakers vary their reading times (RTs) on a word-by-word basis and make reading adjustments according to word

properties such as length, frequency, and word complexity (Just et al. 1982; Keating & Jegerski, 2015). L1 speakers incur a processing cost when encountering syntactic anomalies and mismatches between previous information and incoming input (e.g., Roberts & Liszka, 2013; Stewart et al., 2009). As words are processed incrementally, an increase in RT is detectable either on the target anomalous word(s) or as a spillover on words immediately following, indicating problems in form-meaning assignment when grammatical or logical/semantic incongruencies make sentence meaning ambiguous.

In psycholinguistic studies<sup>4</sup> with L2 learners, online processing is compared with the processing of L1 speakers and with the performance on judgment tasks to determine the status of L2 learner knowledge (e.g., Hopp, 2006, 2016; Jegerski, 2012, 2016; Pliatskias & Marinis, 2013; Roberts & Felser, 2011; Roberts & Liszka, 2013). In L2 research, the term “grammaticality judgments” has generally been preferred to “acceptability judgments” (Spinner & Gass, 2019). More recently, “acceptability” has also been used (e.g., Maie & Godfroid, 2022) as a theoretically more appropriate term because “grammaticality” is an abstract concept referring to competence (Sprouse, 2013), whereas acceptability judgments are perceptions of how acceptable a sentence or a language feature is; thus, the data elicited this way are behavioral and refer to performance.

Notwithstanding the controversies around their usage and terminology, judgment tasks are seen as a useful tool for understanding the developmental stage of the learner interlanguage. Studies in L2 research have traditionally used binary, categorical selection between grammatically acceptable and unacceptable items or sentences. The binary judgment task administered with no time pressure has been validated as a measure of explicit knowledge in Ellis (2005) and used in numerous studies examining the knowledge of L2 morphosyntax (e.g., Coughlin & Trembley, 2013; Ellis et al., 2009; Godfroid et al., 2015; Jiang et al., 2011; Zhang, 2015). In SPR studies comparing L2 and L1 online processing, a graded AJT has been the preferred offline choice in measuring participants’ knowledge (Jegerski, 2012, 2015, 2016; Kaltsa et al. 2016; Roberts & Felser, 2011; Roberts & Liszka, 2013).

With a different target, the present study contributes to this body of research by juxtaposing online processing and offline (untimed) judgements in investigating modal knowledge. A graded AJT using a Likert-type scale is well suited for present purposes, where the use of modals is evaluated in relation to the surrounding context. To date, the research conducted within the implicit/explicit knowledge paradigm has been focused almost exclusively on morphosyntactic language features. The present study, however, introduces a new linguistic target into this paradigm and fills an important gap by investigating modality, or more precisely, two modal auxiliaries, *may* and *can*, involving their semantic, syntactic, and pragmatic functions.

### **The present study**

The study sought to answer the following research questions (RQs):

RQ1:

- (a) To what extent does a match/mismatch between the context and modals expressing ability and epistemic possibility affect acceptability judgement

<sup>4</sup>As pointed out by an anonymous reviewer, SPR can also be used in SLA studies, for example, Lee et al. (2022) in which SPR was employed to demonstrate improvement in comprehension and faster processing following treatment with processing instruction.

ratings in target sentences for upper-intermediate L2 English (L1 Croatian) adult speakers compared with L1 English speakers?

- (b) Does the same alternation, *may* vs. *can*, affect the ratings of sentences conveying the meaning of permission?

RQ2:

- (a) To what extent does a match/mismatch in sentences conveying the meaning of ability and epistemic possibility affect reading times in online processing of target sentences for these participants?
- (b) Does the same alternation, *may* vs. *can*, affect online processing in sentences conveying the meaning of permission?

An untimed AJT task was administered to both L2 and L1 speakers, and the results were compared with the performance of both groups on a SPR task. In the Materials and methods subsection of Method and the Results section below we cover the tasks in the order of the RQs—that is, AJT followed by SPR. However, in the procedure subsection of Method, we cover the tasks in the chronological order that participants encountered them—that is, SPR then AJT.

## Method

All research materials, analytic methods, and data files associated with this study are available in the Open Science Framework (OSF) via the following link: <https://osf.io/6qynj/>. The analyses were performed using R (R Core Team, 2022) with various packages and functions.

## Participants

### L1 English

The L1 English participants were undergraduate students from the Education, Biology, and Economics departments at a UK University (mean age = 20.95,  $SD = 2.17$ , range: 18–35,  $N = 44$ ). The SPR data from 40 L1 English participants and AJT data from 42 L1 English participants were used. Originally, a total of 44 participants completed the SPR, but data from four of these participants (including two who also did not show up for the AJT) were removed after more than two comprehension questions were answered incorrectly (<90% accuracy).

### L2 English

L2 English participants were first- and second-year undergraduate students majoring in English at a Croatian university (mean age 20.6,  $SD = 1.08$ , range: 19–24,  $N = 42$ ). The SPR data from 35 L1 Croatian (hereafter “L2 English”) participants and AJT data from 41 L2 English participants were used. Originally, 42 L2 English participants completed an Oxford Placement Test and achieved a mean accuracy score of 80.48% ( $SD = 5.62$ , range: 70–92), which is comparable to a high B2/low C1 CEFR proficiency level or to an approximate TOEFL overall score range of 84 to 106. The AJT data were obtained from 41 participants, as one participant did not attend the session. SPR data were retained for 35 L2 English participants. A further seven (including one who did not attend the AJT)



completed the SPR task but were removed for answering more than two comprehension questions incorrectly.

### **Materials and methods**

#### *Acceptability judgment task (AJT)*

The AJT stimuli contained the same sentences as those used in SPR (see below), minus the final sentence, which we removed because it was nonessential for eliciting acceptability ratings and to reduce the reading burden. Target sentences were presented in bold to emphasize the focal point of participants' acceptability ratings. The AJT used a Likert-type scale from 1 to 6, where 1 was *least acceptable*, and 6 was *most acceptable*. At the top of the sheet, the instructions asked participants to read sentences carefully, to indicate in each case, the acceptability of the sentence, and for less/unacceptable sentences (i.e., those rated 1–3), to also underline the perceived error.

#### *Self-paced reading (SPR) task*

The SPR stimuli were 36 target items comprising sentences with the modals *can* and *may* manipulated so that each appeared in a matching or a mismatching context relative to the surrounding text, referred to as “congruent/incongruent” for ability/sensation and epistemic possibility. For permission/offers we use the term “formally marked/unmarked” (but for convenience, we use congruent/marked and incongruent/unmarked interchangeably below). There were 18 target items (six each for ability/sensation, epistemic possibility, and permission/offers) in one of the two conditions (congruent/formally marked and incongruent/formally unmarked). Following Stewart et al. (2009), each item comprised three sentences, the first providing a situational context, the second the target sentence containing the modal, and the third serving to wrap up the situation described. Here we provide an example of each modal category as they appeared in the SPR task (target sentences are in bold).

(9) A modal indicating ability:

*Sara is a very experienced driver. Surprisingly, **she can/may drive a van, but she is not able to ride a bike.** She has been driving for more than twenty years on all sorts of roads.*

(10) A modal indicating epistemic possibility:<sup>5</sup>

*Carol is waiting for her friends to pick her up, but they haven't arrived yet. **“They may/can be waiting in the car,” her mum says.** Carol is very impatient.*

(11) A modal indicating permission:

<sup>5</sup>An anonymous reviewer pointed to the potential problem with the ‘modal + HAVE + past participle’ construction of certain epistemic possibility items, which, from a morphosyntactic perspective, is arguably more complex than a ‘modal + base form’ verb construction and could therefore require more cognitive resources to process. However, such constructions do not appear to have had any effect on L1 online processing because the slowdown reaction to the mismatching modal in the context of epistemic possibility starts on Segment 3, earlier than in sentences expressing ability that all use the ‘modal + verb base form’ construction.

*Andrea was sorting the books on her bookshelf when her friend Lisa came in. “You can/may take some if you wish, but please don’t keep them for long,” said Andrea. Lisa was happy to take several books.*

There were 56 items in total, comprising 36 experimental target items and 20 fillers, which also contained three sentences but with a second sentence unrelated to modals (see the stimuli list on the study’s OSF page). Participants also answered 20 comprehension questions (10 each for target items and fillers) to confirm that they were reading the sentences, focusing on meaning, and not skipping through. Target items, adapted from British National Corpus (BNC Consortium, 2007), Coates (2014), and Palmer (1986, 1990), were designed with an equivalent number of syllables as far as possible (Jegerski, 2014). As shown in Table 1, the first word in the sentence (Segment 0) was either a personal pronoun or a one-syllable name, the second word (Segment 1) was the modal (*can/may*), the third (Segment 2) was a one syllable word in 34 out of 36 sentences (two words had two syllables), the fourth (Segment 3) a one syllable word in 29 of the target sentences, and the remaining seven words had two syllables. Experimental target items and fillers were pseudorandomized (see Procedure).

### *Instrument reliability*

*Overview.* For methodological transparency and to provide useful information on the psychometric properties of these stimuli (Plonsky & Derrick, 2016; Marsden et al., 2018), we make the full set of AJT and SPR instrument reliability estimates available on the study’s OSF page, where information on the procedure for estimating instrument reliability can also be found.

*AJT task reliability summary.* Instrument reliability estimation for the AJT data was challenged by nonconvergence in seven out of 24 analyses, with items having no or little variance due to participants rating them as clearly highly acceptable or unacceptable (i.e., all/most participants selecting 6 or 1) or a negative loading or loading of 1 on a single factor involved in reliability computation. Because such rating patterns are theoretically valid, we retained all items for the main analyses (note also that item was modeled as a random effect). However, we generally needed to remove statistically “rogue” items when computing AJT instrument reliability estimates (for full details see “All reliability estimates” on the study’s OSF page). The estimates summarized immediately below reflect these issues, offering a best possible (but imperfect) picture of AJT instrument reliability in the present study (for further information see “AJT and SPR instrument reliability summary interpretation” on the study’s OSF page).

For the AJT data, 24 estimates were available. For L1 English participants, AJT reliability was reasonably high overall (12 estimates, *Mdn* = .79, *IQR* = 0.24), but more

**Table 1.** Distribution of segments in target verb phrases

Modality type	Segment						
	0	1	2	3	4	5	6
<b>Ability<sup>a</sup></b>	<i>He</i>	<i>can/may</i>	<i>read</i>	<i>books</i>	<i>in</i>	<i>six</i>	<i>languages</i>
<b>Epistemic</b>	<i>They</i>	<i>may/can</i>	<i>have</i>	<i>left</i>	<i>for</i>	<i>South</i>	<i>America</i>
<b>Permission<sup>b</sup></b>	<i>You</i>	<i>may/can</i>	<i>talk</i>	<i>to</i>	<i>his</i>	<i>assistant,</i>	<i>says...</i>

<sup>a</sup>Agent-oriented modality

<sup>b</sup>Speaker-oriented modality

varied for congruent and incongruent items (six estimates each, respectively,  $Mdn = .58$ ,  $IQR = 0.59$ ,  $0.09$ ) and different types of modality (four estimates per type, ability:  $Mdn = .39$ ,  $IQR = 0.68$ ; epistemic:  $Mdn = .79$ ,  $IQR = 0.10$ ; permission:  $Mdn = .86$ ,  $IQR = 0.15$ ), while consistently high for stimuli versions 1 and 2 (six estimates each, respectively,  $Mdn = .76$ ,  $IQR = 0.25$  and  $Mdn = .82$ ,  $IQR = 0.22$ ). For the L2 English participants, AJT reliability was generally high overall (12 estimates,  $Mdn = .83$ ,  $IQR = 0.09$ ), for congruent and incongruent items (six estimates each, respectively,  $Mdn = .80$ ,  $.86$ ,  $IQR = 0.08$ ,  $0.07$ ), different types of modality (four estimates per type, ability:  $Mdn = .83$ ,  $IQR = 0.15$ ; epistemic:  $Mdn = .87$ ,  $IQR = 0.05$ ; permission:  $Mdn = .79$ ,  $IQR = 0.03$ ), and stimuli versions 1 and 2 (six estimates each, respectively,  $Mdn = .82$ ,  $IQR = 0.10$  and  $Mdn = .84$ ,  $IQR = 0.08$ ).

*SPR instrument reliability summary.* Given that the SPR task also included data for each sentence segment, 168 reliability estimates were available. For the 84 estimates pertaining to L1 English participants (see also Mifka-Profozic et al., 2020), internal consistency was high overall ( $Mdn = .94$ , range:  $.43$ – $.99$ ,  $IQR = 0.06$ ), as well as for congruent and incongruent items (42 estimates each, respectively,  $Mdn = .94$ ,  $.93$ ,  $IQR = 0.04$ ,  $0.08$ ), different types of modality (28 estimates per type, ability:  $Mdn = .95$ ,  $IQR = 0.06$ ; epistemic:  $Mdn = .94$ ,  $IQR = 0.04$ ; permission:  $Mdn = .92$ ,  $SD = 0.16$ ), stimuli versions 1 and 2 (42 estimates each, respectively,  $Mdn = .97$ ,  $IQR = 0.03$  and  $Mdn = .91$ ,  $IQR = 0.11$ ), and across all six segments ( $Mdn = .92$  to  $.97$ ,  $IQR = 0.04$  to  $0.10$ ). For the L2 English participants, internal consistency was similarly high overall for the 84 estimates ( $Mdn = .96$ , range:  $.86$ – $.99$ ,  $IQR = 0.05$ ), for congruent and incongruent items (42 estimates each, respectively,  $Mdn = .96$ ,  $.96$ ,  $IQR = 0.03$ ,  $0.05$ ), different modality types (28 estimates per type, ability:  $Mdn = .96$ ,  $IQR = 0.03$ ; epistemic:  $Mdn = .96$ ,  $IQR = 0.03$ ; permission:  $Mdn = .97$ ,  $SD = 0.07$ ), for stimuli versions 1 and 2 (42 estimates each, respectively,  $Mdn = .97$ ,  $IQR = 0.03$  and  $Mdn = .95$ ,  $IQR = 0.05$ ), and across all six segments ( $Mdn = .92$  to  $.98$ ,  $IQR = 0.01$  to  $0.05$ ).

### Procedure

In this section, we cover tasks in the order the participants encountered them (i.e., SPR then AJT). Following the procedure suggested in Keating and Jegerski (2015), we first administered the SPR task and then AJT to avoid any possibility that participants consciously notice the presence of ungrammatical items in the SPR, as the AJT may make them metalinguistically aware. As common practice in psycholinguistic and SLA research shows, studies that measure both online processing and offline performance on tests of explicit knowledge administer the tests one immediately after another: implicit first, then explicit (e.g., Coughlin & Tremblay, 2013; Jegerski, 2012, 2015, 2016; Maie & Godfroid, 2022; Roberts & Liszka, 2013). We administered the two tasks with a 1- or 2-day difference at both research sites (the UK and Croatia) depending on whether a participant had completed the SPR on Day 1 or Day 2. The reason for spreading the administration of the SPR over two days was that each participant was tested individually, having to spend at least 15–20 min with the research assistant: to read the information about the study, ask questions and receive answers, have the procedure explained, read and sign the ethical consent, and have practice before completing the task itself. We are confident that the split administration of the two tasks could not have any effect on results because, despite using the same sentences, the two tasks were entirely different.

*SPR task*

The SPR task was administered using the freely available Psychopy software (Peirce, 2007, 2009). Participants were instructed to read each sentence at a normal speed and press the space bar to proceed to the next word, with each word on the screen disappearing before the next word appeared. A centred noncumulative “stationary window” method was used, with the white experimental items/fillers text appearing on the black screen word by word until the end of the entire task. Only one word was visible at a time. After each set of sentences, an instruction appeared on the screen to remind participants what they were required to do.

Before starting the main task, participants read three items for practice to help them become familiar with the task. The practice items had a structure similar to that of the experimental items but were unrelated to the modals. For the main task, participants read through the 56 items that appeared in the general order of two experimental target items followed by one filler (a target-item-to-filler ratio of 2:1 for 32 target items and 1:1 for two target items), with 20 randomly appearing comprehension questions, 10 following the 36 target items (comprehension was tested 28% of the time) and 10 following the 20 fillers (filler comprehension was tested 50% of the time). Comprehension questions were unrelated to the use of modals to avoid interfering with target item processing (Roberts & Liszka, 2013).

To account for possible order effects (see also Data analysis), the experimental target items and fillers were counterbalanced into two versions (1 and 2), each encountered by only half of the participants. Thus, half of the participants encountered target items in the order 1–36 and fillers in the order 1–20 (Version 1), and the other half encountered target items in the order 36–1 and fillers in the order 20–1 (Version 2). If an item in Version 1 had a congruent modal, the same item in Version 2 had the incongruent modal and vice versa; otherwise, the two versions were identical. At the end of the SPR task, a sentence appeared explaining that this was the end and thanking participants.

*AJT task*

The AJT task was administered as a pen-and-paper task. First, verbal and written instructions were given. Participants were asked to ignore any spelling or punctuation mistakes and not to go back and change a response once decided. One example sentence (not involving a modal) was then provided, with a corresponding rating and underlining, as it had lower acceptability.

The AJT task used the same target items and fillers as the SPR task and was also counterbalanced into two versions, but it did not contain comprehension questions. Participants completed the same version encountered in SPR, with the difference that the AJT task presented all 56 items (36 target items and 20 fillers) in a single list, with either two items followed by one filler (14 times) or one target item followed by one filler (6 times), whereas the last two items were target items. Target items and fillers were randomized within this ordering (e.g., target items did not follow the order 1–36).

**Data analysis**

To address RQ1 we ran ordinal mixed-effects regression analyses separately for each modal type (ability, epistemic, permission) using the *clmm* function in the *ordinal* package in R (Christensen, 2019). These analyses comprised six AJT models. We ran separate analyses for the L1 English participants (AJT models 1–3 for ability, epistemic, and permission, respectively) and L2 English participants (AJT models 4–6 for the

respective modality types) and visually compared models, as this offered more nuanced and interpretable findings than entering “group” (L1/L2) as a predictor in an interaction with “congruency” (but interested readers can construct such analyses using the OSF R script). All analyses had one ordinal outcome variable, “rating” (responses spanned a 1–6 scale covering *least* to *most acceptable*). The AJT models 1–6 had one fixed effect, “congruency” with two levels (congruent, incongruent), sum-coded to compare the rating mean for a given level with the overall rating mean for both levels, with by-participant and by-sentence random slopes.

To address RQ2, we first collected all SPR target item reading times (RTs) calculated in milliseconds from Segment 0 (the word preceding the modal) to Segments 1, 2, 3, 4, 5, and 6 and responses to the 20 comprehension questions (10 following target items, 10 following fillers). As noted above (see Participants), four L1 English and seven L2 English participants with less than 90% comprehension accuracy (>2/20 responses incorrect) were removed before further analyses. Given that the L2 English participants were upper-intermediate proficiency, we applied the same comprehension accuracy standard as for the L1 English participants. These steps resulted in 40 L1 English participants and 35 L2 English participants retained for the main SPR analyses.

Before proceeding with main analyses, we conducted an item analysis to evaluate the scope and functioning of the items across both tasks.<sup>6</sup> A qualitative reexamination of items 26 and 27 (adapted from Coates’s corpus) revealed that although they would satisfy Bybee’s definition of speaker-oriented modality, it was preferable to remove them from all SPR and AJT analyses because they are fixed phrases. Quantitatively, an examination of the L1 participants’ AJT average ratings and item discriminability analyses revealed no systemic problems with the statistical function of items.

In line with recent methodological syntheses of L2 SPR methodology and outlier treatment in applied linguistics studies, we considered RTs in terms of their potential legitimacy and distribution rather than using standard deviation boundaries (Nicklin & Plonsky, 2020). First, we set the lower boundary for a legitimate RT at 150 ms, the point at which L1 magnetoencephalography (MEG) research suggests that lexicality (i.e., word form identification) likely begins, although findings vary (Hsu et al, 2011; Nicklin & Plonsky, 2020). Although the 150-ms boundary is derived from L1 MEG research, we also used it for the L2 English participants given their reasonably high proficiency and in the absence of equivalent L2 MEG research (Nicklin & Plonsky, 2020). The upper boundary was set at the commonly used 2,000-ms level for both L1 and L2 English participants, a potentially strict cutoff for mid and lower L2ers but reasonable, we argue, for upper-intermediate L2ers and considering that Plonsky and Nicklin (2020) found the median RTs in 18/19 L2 studies (some including sentence/phrase rather than word-by-word presentation) to lie below this boundary. These steps resulted in the removal of 26/9,520 L1 English RTs (0.27%) and 78/8,330 L2 English RTs (0.94%) after the exclusion of items 26 and 27. The distribution of the data were then checked, and for each SPR model, RTs were log transformed to reduce positive skew.

For the main analyses, we ran linear mixed-effects regression analyses separately for each modal type (ability, epistemic, permission) using the *lmer* function in the *lme4* R package (Bates et al., 2015). These analyses comprised six SPR models. Again, we ran separate analyses on the L1 English data (SPR models 1–3 for ability, epistemic, and

<sup>6</sup>We thank the anonymous reviewers for suggesting this preliminary step before proceeding with further analyses.

permission modality, respectively) and L2 English data (SPR models 4–6 for the respective modality types), in each case looking at which sentence segments readers significantly slowed down on (and by how much) after encountering an incongruent/formally unmarked modal. This offered a more nuanced and interpretable set of findings compared with entering “group” as predictor in a three-way interaction with “congruency” and “segment” (again, interested readers are referred to the OSF R script). These analyses had one continuous outcome variable, “log RT,” and two categorical fixed effect predictors entered as an interaction, “congruency” (sum-coded: congruent as 1, incongruent as -1)  $\times$  “segment” (seven levels, coded [segment] 0, 1, 2, 3, 4, 5, and 6).

The *buildmer* R package (Voeten, 2022) was used to automatically identify optimal models based on which terms made significant contributions to log-ratio likelihood. Thus, for the SPR analyses, random effects were maximally specified as by-participant and by-sentence random intercepts and slopes for the congruency-segment interaction (Barr et al., 2013), with *buildmer* arriving at optimal, simpler structures—namely, the intercept and slope of congruency conditioned on “participant” for all models and on “sentence” for all models except SPR Model 6 (L2 participants, permission), for which only the intercept was used. Given the counterbalancing in our experimental design and random effects pertaining to items, we opted not to enter stimuli version as an additional fixed or random effect (cf. Mifka-Profozic et al., 2020).

For the AJT models, we report effect estimates and corresponding 95% confidence intervals (CIs), standard errors, degrees of freedom,  $z$  values (equal to the estimate divided by the standard error),  $p$  values, and, estimated independently from the models (Jegerski, 2018), standardized effect sizes (Cohen’s  $d$ ) and 95% confidence intervals for within-participants comparisons of mean ratings for congruent versus incongruent items by modal and participant type, calculated using the *effsize* R package taking into account within-participant variation (Torchiano, 2020). Both AJT and SPR effects (see below) are interpreted using Cohen’s (1988) benchmarks of  $d = .20$  (small),  $d = .50$  (medium), and  $d = .80$  (large) and by noting 95% confidence intervals that did not pass through zero, indicating a reliable effect (cf. Jegerski, 2018). As Plonsky and Oswald’s (2014) L2-field-specific scale for within-group contrasts of  $d = .60$  (small),  $d = 1.0$  (medium), and  $d = 1.4$  (large) was derived from a meta-analysis of a large number of pedagogical interventions, it had limited applicability given the current study’s instrumentation, design, and specific focus.

For the SPR models, we report effect estimates and corresponding 95% CIs, standard errors, degrees of freedom, Wald  $t$  values (the estimate divided by the standard error),  $p$  values, and standardized effect sizes (Cohen’s  $d$ ) and 95% confidence intervals, as described for the AJT, with the addition of by-segment analyses and consideration of Avery and Marsden’s (2019) indicators of reliable SPR ambiguity resolution to interpret ability modal effect sizes (for L1  $d = .23$  [.13, .32]; for L2  $d = .19$  [.12, .25]) and anomaly detection to interpret epistemic modal effect sizes (for L1  $d = .41$  [.29, .54]; for L2  $d = .19$  [.09, .29]). As permission modals expressed a pragmatic function, Avery and Marsden’s (2019) findings were not appropriate for interpreting effect sizes for this type of modal. Any by-segment results for SPR are interpreted from Segment 2 (the lexical verb immediately following the modal) onward.

To establish the statistical significance of regression estimates, the conventional alpha values of .05, .01, and .001 were lowered to .0083, .0017, and .00017 to reduce the chance of a Type I error (i.e., dividing these alpha values by six, given that in any analysis the highest order interaction had six parameters, any of which could support our predictions). The  $p$  values for estimates are thus interpreted against these adjusted

alpha values. For simplicity, we report unadjusted 95% CIs of estimates rather than 99.167% CIs (i.e., the equivalent precision of 95% CIs adjusted for six repeated tests).

For all models we used the *r2* function in the *performance* R package (Lüdtke et al., 2021) to compute and report marginal and conditional  $R^2$  to show the proportion of variance explained, respectively, by the fixed effects alone and combined fixed and random effects. The  $R^2$  values were interpreted as small (.18), medium (.32), or large (.51) based on the amount of variance in AJT ratings and SPR log RTs explained (Plonsky & Ghanbar, 2018).

## Results

In the Results section we cover tasks in the order of the RQs (AJT then SPR).

### AJT analyses (RQ1)

In this section we report the results of the ordinal mixed-effects regression analyses of the AJT data (AJT models 1–6) focusing on model estimates that directly address RQ1 and Cohen's *d* effect sizes (see the study's OSF page for descriptive statistics and the R script to obtain the full list of model estimates).

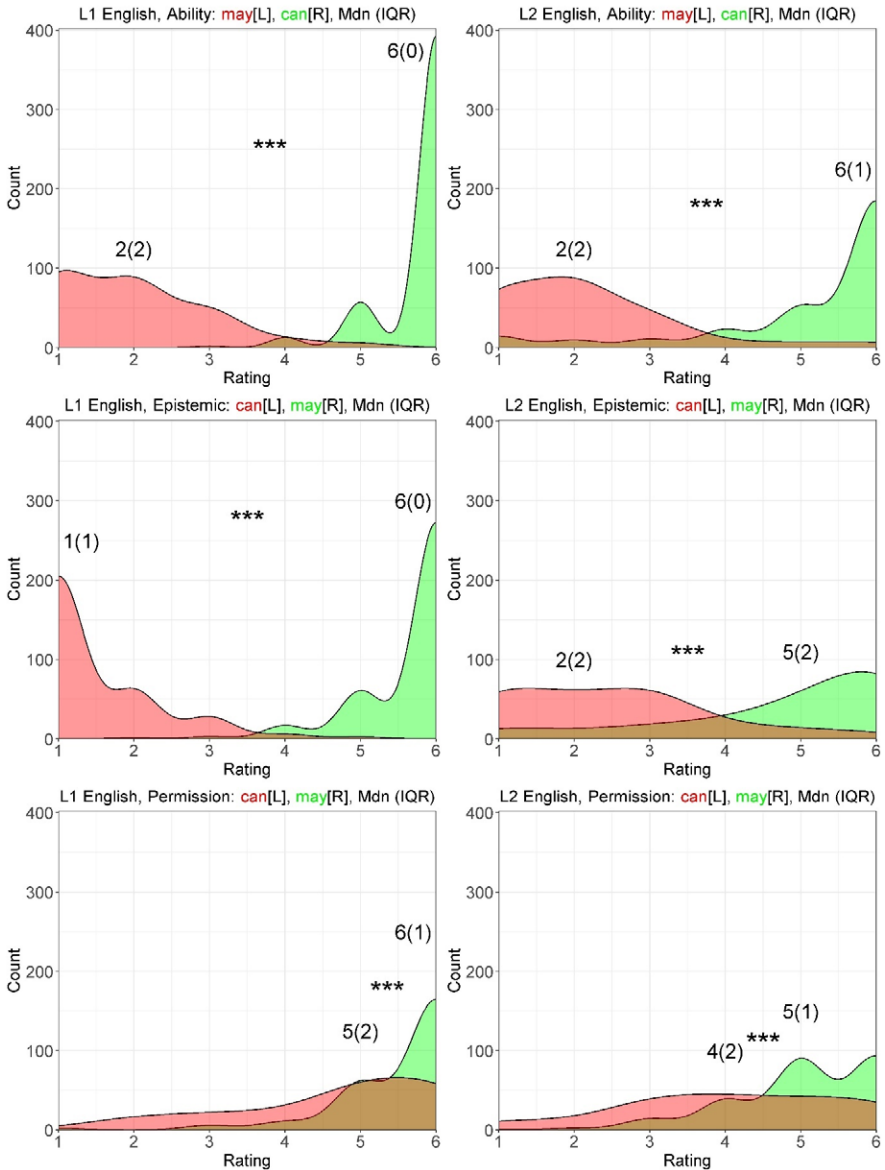
In AJT models 1–3 (L1 English participants), the fixed effects explained large, large, and negligible (<.18) proportions of rating variance for ability, epistemic, and permission modals respectively (marginal  $R^2 = .80, .86, .17$ ), and the combined fixed and random effects explained large amounts of variance in all cases (conditional  $R^2 = .88, .87, .51$ ). In AJT models 4–6 (L2 English participants), the fixed effects explained large, medium, and negligible proportions of rating variance for ability, epistemic, and permission modals respectively (marginal  $R^2 = .53, .37, .14$ ), and large, large, and medium proportions of variance were explained by the combined fixed and random effects (conditional  $R^2 = .62, .52, .35$ ).

Table 2 and Figure 1 show that both L1 and L2 groups rated congruent/formally marked items as significantly more acceptable than incongruent/formally unmarked items across all three modal types.

The estimates in Table 2 and Cohen's *d* effect sizes in Table 3 show that the effect of congruency (ability and epistemic modals) was stronger for L1 English participants than for the L2 English participants, but for formality (permission modals) the effect was slightly stronger for the L2 participants. Items were rated significantly differently for all modal types by both participant groups, with Cohen's *d* effect size confidence intervals not passing through zero. Effects were (very) large for ability (L1  $d = 9.64$  [6.47,

**Table 2.** Summary of fixed effect predictors of rating (outcome) for L1 English AJT data (AJT models 1–3) and L2 English AJT data (AJT models 4–6)

Data	AJT model	Modality type	Parameter	Estimate [95%CIs]	SE	z	p
L1 English participants (N = 42)	1	Ability	Congruency1	4.655 [3.724, 5.586]	0.475	9.800	<.001***
	2	Epistemic	Congruency1	4.666 [3.848, 5.484]	0.417	11.180	<.001***
	3	Permission	Formality1	1.050 [0.531, 1.568]	0.265	3.968	<.001***
L2 English participants (N = 41)	4	Ability	Congruency1	2.131 [1.774, 2.488]	0.182	11.690	<.001***
	5	Epistemic	Congruency1	1.587 [1.145, 2.029]	0.225	7.042	<.001***
	6	Permission	Formality1	0.830 [0.389, 1.271]	0.225	3.688	<.001***



**Figure 1.** Density plots (smoothed histograms), within the chart space showing median (*Mdn*), interquartile range (*IQR*), and distribution of ratings on the left [L] or right [R] for L1 English participants (AJT models 1–3) and L2 English participants (AJT models 4–6) for congruent/formally marked items (green) and incongruent/formally unmarked items (red) by different modality types, significant differences (\*\* $p < .001$ ).

12.81]; L2  $d = 4.21$  [2.71, 5.72]) and epistemic possibility (L1  $d = 11.15$  [7.92, 14.37]; L2  $d = 3.17$  [1.97, 4.38]), and large for permission modals (L1  $d = 1.09$  [0.57, 1.61]; L2  $d = 1.17$  [0.61, 1.73]).



**Table 3.** Cohen's *d* effect sizes for within-participants comparisons of mean ratings for congruent/formally marked versus incongruent/unmarked items by modal type for L1 English and L2 English participants with two interpretation frameworks

Data	Type	Cohen's <i>d</i> [95% CIs]	Effect size interpretation framework	
			Cohen (1988) <sup>a</sup>	CIs do not pass through 0
L1 English participants ( <i>N</i> = 42)	Ability	9.64 [6.47, 12.81]	large	1
	Epistemic	11.15 [7.92, 14.37]	large	1
	Permission	1.09 [0.57, 1.61]	large	1
L2 English participants ( <i>N</i> = 41)	Ability	4.21 [2.71, 5.72]	large	1
	Epistemic	3.17 [1.97, 4.38]	large	1
	Permission	1.17 [0.61, 1.73]	large	1

<sup>a</sup>Cohen (1988) effect size magnitude: small (.20), medium (.50), large (.80).

The medians and interquartile ranges contained in Figure 1 show that on average, both L1 and L2 participants rated congruent items on the upper half of the acceptability scale (i.e., 4–6) and incongruent items on the lower half of the scale (i.e., 1–3) for ability and epistemic possibility. Both groups rated both formally marked and unmarked items on the upper half of the scale for permission modal, with *may* significantly higher in each case.

To summarize, the analyses of the AJT data showed large effects for both the L1 and L2 English groups, with each rating congruent/formally marked items as significantly more acceptable than incongruent/formally unmarked items for all modals. For both L1 and L2 English participants, effects were most pronounced with ability and epistemic possibility. The two groups were most distant from one another in the way they rated epistemic possibility, which for L2ers showed a comparatively smaller (but still large) effect than for ability modals. For ability and epistemic possibility, the L1 English participants consistently provided more extreme ratings at either end of the rating scale compared than did the L2 participants. The L1 English participants rated both marked and unmarked items slightly higher on the whole (medians in the 5–6 range) than the L2 English participants did (medians in the 4–5 range).

### SPR analyses (RQ2)

In this section we report the results of the mixed-effects regression analyses of the SPR data (SPR models 1–6) and Cohen's *d* effect sizes. For each model, we again focus on the fixed-effect estimates that directly address the research questions—namely the Congruency × Segment interaction (see the study's OSF page for descriptive statistics and the R script to obtain the full list of model estimates).

The fixed effects explained a relatively small proportion of log-RT variance for ability, epistemic, and permission modals, respectively, (SPR models 1–3 marginal  $R^2 = .06, .04, .01$ ; SPR models 4–6 marginal  $R^2 = .03, .01, .02$ ), whereas the combined fixed and random effects explained a larger amount (SPR models 1–3 conditional  $R^2 = .64, .55, .53$ , respectively; SPR models 4–6 conditional  $R^2 = .57, .64, .66$ , respectively).

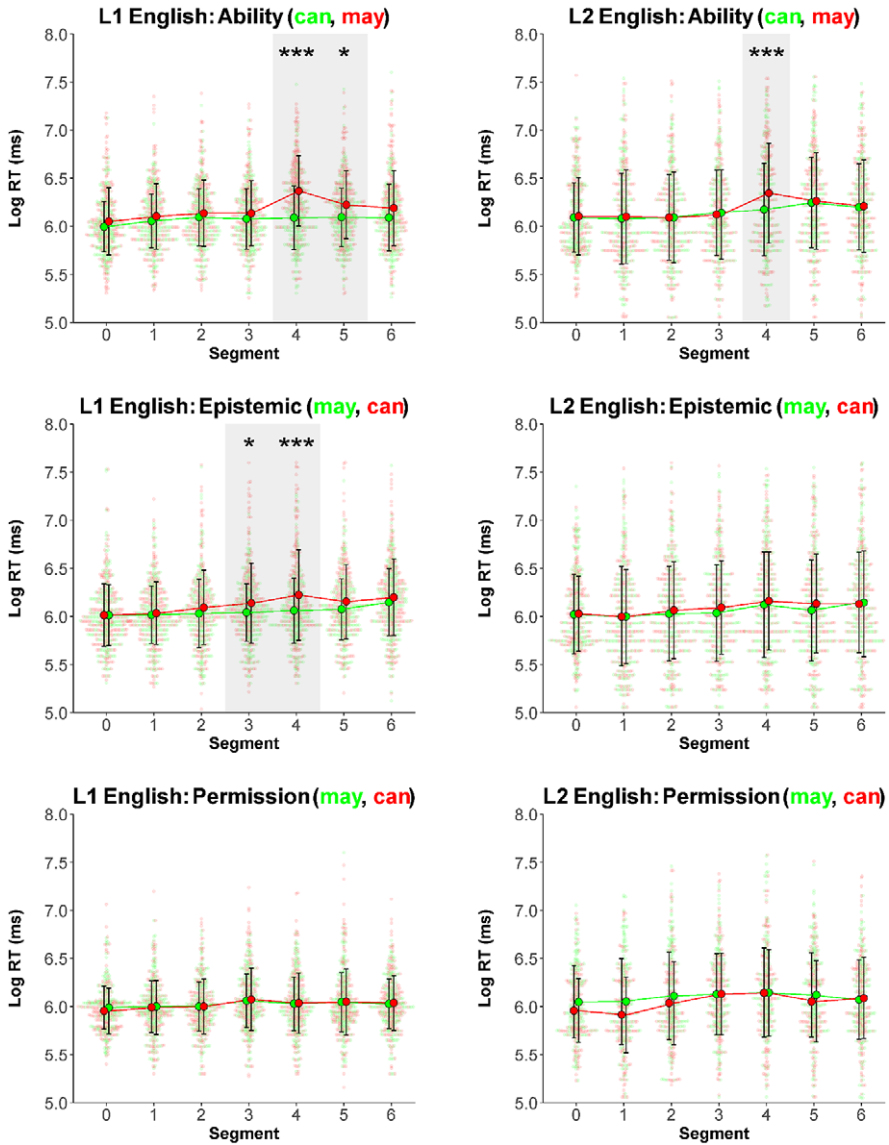
The estimates in Table 4 (visualized in Figure 2) and Cohen's *d* effect sizes in Table 5 show that both participant groups displayed sensitivity to semantic ambiguities with the modal denoting ability. Specifically, the L1 English participants had significantly slower log RTs for ambiguous sentences at Segment 4 (estimate = -0.109 [-0.136, -0.83],  $t = -8.095$ ,  $p < .001$ ,  $d = -0.92$  [-1.24, -0.59]) and Segment 5 (estimate = -0.038 [-0.064,

**Table 4.** Summary of fixed effect predictors of log RTs (outcome) for L1 English SPR data (SPR models 1–3) and L2 English SPR data (SPR models 1–3), significant predictors in bold, results interpreted from Segment 2 (the lexical verb immediately following the modal) onward

Type	Parameter	L1 English participants (N = 40): Model 1 = Ability; Model 2 = Epistemic; Model 3 = Permission					L2 English participants (N = 35): Model 4 = Ability; Model 5 = Epistemic; Model 6 = Permission				
		Estimate [95% CIs]	SE	df	t	p	Estimate [95% CIs]	SE	df	t	p
Ability	(Intercept)	<b>6.023 [5.938, 6.108]</b>	<b>0.043</b>	<b>52.432</b>	<b>138.866</b>	<b>&lt;.001***</b>	<b>6.096 [5.971, 6.221]</b>	<b>0.064</b>	<b>47.792</b>	<b>95.492</b>	<b>&lt;.001***</b>
	Cong1:Seg1	0.004 [−0.022, 0.030]	0.013	3241.020	0.299	.765	−0.003 [−0.045, 0.039]	0.022	2812.422	−0.133	.894
	Cong1:Seg2	0.006 [−0.020, 0.033]	0.013	3241.020	0.474	.635	0.006 [−0.036, 0.048]	0.022	2812.504	0.271	.786
	Cong1:Seg3	−0.003 [−0.029, 0.024]	0.013	3241.035	−0.188	.851	0.018 [−0.024, 0.060]	0.022	2812.483	0.831	.406
	Cong1:Seg4	<b>−0.109 [−0.136, −0.083]</b>	<b>0.014</b>	<b>3241.133</b>	<b>−8.095</b>	<b>&lt;.001***</b>	<b>−0.084 [−0.127, −0.042]</b>	<b>0.022</b>	<b>2812.877</b>	<b>−3.899</b>	<b>&lt;.001***</b>
	Cong1:Seg5	<b>−0.038 [−0.064, −0.011]</b>	<b>0.013</b>	<b>3241.035</b>	<b>−2.783</b>	<b>.005*</b>	−0.001 [−0.043, 0.042]	0.022	2812.593	−0.023	.982
	Cong1:Seg6	−0.023 [−0.049, 0.004]	0.013	3241.035	−1.671	.095	−0.003 [−0.045, 0.040]	0.022	2812.168	−0.117	.907
Epistemic	(Intercept)	<b>6.013 [5.926, 6.101]</b>	<b>0.044</b>	<b>52.872</b>	<b>135.126</b>	<b>&lt;.001***</b>	<b>6.025 [5.880, 6.169]</b>	<b>0.074</b>	<b>47.163</b>	<b>81.652</b>	<b>&lt;.001***</b>
	Cong1:Seg1	−0.009 [−0.040, 0.022]	0.016	3230.699	−0.574	.566	0.002 [−0.039, 0.043]	0.021	2794.965	0.091	.927
	Cong1:Seg2	−0.032 [−0.063, −0.001]	0.016	3230.736	−2.000	.046	−0.017 [−0.058, 0.024]	0.021	2794.971	−0.804	.422
	Cong1:Seg3	<b>−0.049 [−0.080, −0.018]</b>	<b>0.016</b>	<b>3230.734</b>	<b>−3.065</b>	<b>.002*</b>	−0.027 [−0.069, 0.014]	0.021	2795.346	−1.297	.195
	Cong1:Seg4	<b>−0.089 [−0.121, −0.058]</b>	<b>0.016</b>	<b>3231.292</b>	<b>−5.588</b>	<b>&lt;.001***</b>	−0.016 [−0.057, 0.026]	0.021	2795.084	−0.743	.458
	Cong1:Seg5	−0.041 [−0.073, −0.010]	0.016	3230.794	−2.604	.009	−0.034 [−0.075, 0.008]	0.021	2795.181	−1.591	.112
	Cong1:Seg6	−0.026 [−0.057, 0.005]	0.016	3230.761	−1.658	.097	−0.003 [−0.038, 0.045]	0.021	2795.318	0.143	.886
Permission	(Intercept)	<b>5.971 [5.896, 6.046]</b>	<b>0.038</b>	<b>44.249</b>	<b>156.266</b>	<b>&lt;.001***</b>	<b>6.003 [5.886, 6.120]</b>	<b>0.060</b>	<b>45.146</b>	<b>100.546</b>	<b>&lt;.001***</b>
	Form1:Seg1	−0.012 [−0.040, 0.015]	0.014	2690.038	−0.880	.379	0.024 [−0.013, 0.061]	0.019	2350.946	1.251	.211
	Form1:Seg2	−0.016 [−0.044, 0.011]	0.014	2690.038	−1.166	.244	−0.006 [−0.043, 0.032]	0.019	2350.960	−0.293	.769
	Form1:Seg3	−0.024 [−0.051, 0.003]	0.014	2690.038	−1.713	.087	−0.048 [−0.085, −0.011]	0.019	2350.954	−2.519	.012
	Form1:Seg4	−0.020 [−0.047, 0.008]	0.014	2690.069	−1.399	.162	−0.042 [−0.079, −0.005]	0.019	2350.946	−2.232	.026
	Form1:Seg5	−0.017 [−0.045, 0.010]	0.014	2690.038	−1.241	.215	−0.009 [−0.046, 0.028]	0.019	2350.954	−0.481	.630
	Form1:Seg6	−0.022 [−0.050, 0.005]	0.014	2690.069	−1.605	.109	−0.049 [−0.086, −0.012]	0.019	2350.992	−2.581	.010

Note. Alpha values adjusted to correct for six repeated tests. SE = standard error; df = degree of freedom; Cong/Form = congruency/formality; Seg = Segment.

\*p < .0083; \*\*p < .0017; \*\*\*p < .00017.



**Figure 2.** Modal log RTs (small dots = individual log RTs, larger points = mean log RTs connected by lines, vertical black bars =  $\pm 1 \times$  standard deviation, green = congruent/marked, red = incongruent/unmarked, y-axis truncated at 5.0, shaded areas and asterisks show segments where significant slowdowns occurred (\* $p < .0083$ ; \*\* $p < .0017$ ; \*\*\* $p < .00017$ ).

-0.011],  $t = -2.783$ ,  $p = .005$ ,  $d = -0.49$  [-0.82, -0.16]), whereas the L2 English participants slowed down at Segment 4 only (estimate = -0.084 [-0.127, -0.042],  $t = -3.899$ ,  $p < .001$ ,  $d = -0.53$  [-0.78, -0.27]). For epistemic modality, only the L1 English participants displayed sensitivity to grammatical violation, showing significantly slower log RTs at Segment 3 (estimate = -0.049 [-0.080, -0.018],  $t = -3.065$ ,  $p = .002$ ,  $d = -0.43$  [-0.68,

**Table 5.** Cohen's *d* effect sizes for within-participants comparisons of mean SPR RTs for raw congruent versus incongruent items by modal type and segment for L1 English and L2 English participants with three interpretation frameworks, results interpreted from Segment 2 (the lexical verb immediately following the modal) onward

Type	Segment	L1 English participants ( <i>N</i> = 40)				L2 English participants ( <i>N</i> = 35)			
		Cohen's <i>d</i> [95% CIs]	Effect size interpretation framework			Cohen's <i>d</i> [95% CIs]	Effect size interpretation framework		
			Cohen (1988) <sup>a</sup>	Avery and Marsden (2019) sensitivity <sup>b</sup>	CIs do not pass through 0		Cohen (1988) <sup>a</sup>	Avery and Marsden (2019) sensitivity <sup>b</sup>	CIs do not pass through 0
Ability	0	-0.25 [-0.56, 0.06]	-	-	-	-0.07 [-0.24, 0.10]	-	-	-
	1	-0.20 [-0.43, 0.02]	-	-	-	-0.05 [-0.26, 0.16]	-	-	-
	2	-0.19 [-0.42, 0.03]	-	-	-	-0.04 [-0.19, 0.10]	-	-	-
	3	-0.22 [-0.49, 0.06]	small	-	-	0.05 [-0.13, 0.22]	-	-	-
	4	-0.92 [-1.24, -0.59]	large	1	1	-0.53 [-0.78, -0.27]	medium	1	1
	5	-0.49 [-0.82, -0.16]	small	1	1	-0.03 [-0.23, 0.17]	-	-	-
6	-0.34 [-0.61, -0.07]	small	1	1	-0.06 [-0.24, 0.11]	-	-	-	
Epistemic	0	0.02 [-0.13, 0.17]	-	-	-	0.02 [-0.13, 0.18]	-	-	-
	1	-0.10 [-0.25, 0.05]	-	-	-	0.03 [-0.12, 0.18]	-	-	-
	2	-0.25 [-0.49, -0.01]	small	-	1	-0.08 [-0.30, 0.14]	-	-	-
	3	-0.43 [-0.68, -0.17]	small	1	1	-0.08 [-0.27, 0.11]	-	-	-
	4	-0.64 [-0.94, -0.35]	medium	1	1	0.00 [-0.28, 0.28]	-	-	-
	5	-0.34 [-0.55, -0.13]	small	-	1	-0.15 [-0.36, 0.06]	-	-	-
6	-0.26 [-0.50, -0.01]	small	-	1	-0.02 [-0.17, 0.14]	-	-	-	
Permission	0	0.11 [-0.13, 0.36]	-	-	-	0.29 [0.08, 0.50]	-	-	-
	1	0.02 [-0.19, 0.22]	-	-	-	0.36 [0.14, 0.58]	-	-	-
	2	-0.04 [-0.24, 0.17]	-	-	-	0.20 [-0.12, 0.52]	small	-	-
	3	-0.13 [-0.37, 0.11]	-	-	-	-0.02 [-0.22, 0.18]	-	-	-
	4	-0.07 [-0.32, 0.18]	-	-	-	0.05 [-0.29, 0.38]	-	-	-
	5	-0.06 [-0.34, 0.22]	-	-	-	0.22 [-0.06, 0.49]	small	-	-
6	-0.09 [-0.28, 0.09]	-	-	-	-0.04 [-0.30, 0.22]	-	-	-	

Note. - = effect size not interpreted using this framework; blank space = effect size was negligible by this framework.

<sup>a</sup>Cohen (1988) effect size magnitude: small (.20), medium (.50), large (.80).

<sup>b</sup>Avery and Marsden's (2019) meta-analysis showing reliable SPR sensitivity to ambiguity resolution (used to interpret ability modals) for L1ers ( $d = .23$  [.13, .32]) and L2ers ( $d = .19$  [.12, .25]) and grammatical anomaly (used to interpret epistemic modals) for L1ers ( $d = .41$  [.29, .54]) and L2ers ( $d = .19$  [.09, .29]). This framework was not applicable for permission modals.

-0.17]) and Segment 4 (estimate = -0.089 [-0.121, -0.058],  $t = -5.588$ ,  $p < .001$ ,  $d = -0.64$  [-0.94, -0.35]). For permission, there was no significant difference between RTs associated with the two modals for neither the L1 nor L2 participants.

Cohen's  $d$  effect sizes show the slowdown on incongruent sentences was greatest for L1 ability Segment 4 (a large effect size with reliable sensitivity to semantic ambiguity), followed by L1 epistemic Segment 4 (medium with reliable sensitivity to grammatical violation), and L2 ability Segment 4 (medium with reliable sensitivity to semantic ambiguity). Table 5 shows that various other L1 ability, L1 epistemic, and L2 permission segments had small effects (Cohen, 1988), reliable sensitivity to grammatical anomaly and ambiguity indicated (Avery & Marsden, 2019), and confidence intervals not passing through zero, but as shown in Table 4 and Figure 2, the mixed-effects model estimates for these segments were not significant with the adjusted alpha value.

To summarize, analyses of the SPR data showed that for ability modal L1 and L2 English participants were similarly sensitive to context–modal mismatches involving the incongruent *may* (SPR Models 1 and 4) with an observable spillover effect commencing at Segment 4, which was large for L1ers and medium for L2ers. For epistemic modality, only the L1 English participants were sensitive to mismatches involving the incongruent *can*, with a small, observable spillover effect commencing in significance at Segment 3 (SPR model 2). For permission, no significant difference was shown in either group, between the RTs associated with the use of either *can* or *may*.

## Discussion

The present study sought to investigate the nature of L2 English knowledge for modals *may* and *can* and how this compares with that of L1 speakers. RQ1 asked about how a match/mismatch between the context and modals expressing agent-oriented ability and epistemic possibility as well as the formal context in speaker-oriented permission affect offline acceptability judgement ratings for the two participant groups. RQ2 asked about how a match/mismatch and the formal context affect reading times in online processing of sentences containing these modals, indirectly tapping into implicit knowledge.

The results revealed two major findings. The first is that L1 and L2 English speakers rate the acceptability of sentences containing modals *can* and *may* in an offline AJT similarly—that is, sentences containing a modal congruent with the context were rated significantly higher than sentences with the modal mismatching the context, with very large effect sizes observed. Likewise, the pragmatic use of *may* in formal situations was rated by both groups significantly higher than the use of *can*, with large effect sizes observed; however, *can* was still rated on the upper half of the scale, which suggests that the participants do not reject it as they do in incongruent use of the modal mismatching the context for ability and epistemic possibility. The difference between the reliable acceptance of context/modal match and rejection of the mismatch referring to the modals expressing ability (*can*) and epistemic possibility (*may*) suggests that L2 learners exhibit a level of explicit knowledge related to the semantics of the two modals. A difference between the L1 and L2 English speakers can be seen in the levels of their certainty. The L1 speakers consistently used extreme ratings (with a smaller spread) in both acceptance and rejection, showing a high level of certainty. The L2 speakers, on the other hand, seem to be more hesitant and apply more cautious ratings (with somewhat more spread), although still generally in accord with the L1 group. This may be interpreted as characteristic of explicit knowledge, which is known to be variable and unsystematic in L2 learners (Ellis et al., 2009).

The second major finding of the study is that the L2 learners differed from native speakers in their processing of epistemic modality—that is, L2 learners did not show sensitivity to grammatical violation when *can* was used instead of *may* in a context suggesting epistemic possibility. However, they were sensitive to mismatches, like L1 speakers, when processing sentences using the modal *may* instead of *can* to express ability, with a large effect for L1ers and a medium effect for L2ers. This suggests that the upper-intermediate L2 learners have developed their representations of the semantics contained in the modal *can*, but this cannot be confirmed for the epistemic meaning encoded in the modal *may*.

The differential behavior of L2 speakers in the two types of task (AJT and SPR) is significant because it indicates that two different types of knowledge are being tapped into while rating sentence acceptability, on the one hand, and real-time processing, on the other. Our study sought to contribute an examination of modality to the discussion on implicit/explicit L2 knowledge. As such it is one of the rare studies that used graded acceptability judgments to tap into explicit linguistic knowledge. However, one may argue that modal knowledge cannot be acquired declaratively (i.e., by learning the pedagogical rules) because it is a different type of knowledge than knowledge of morphosyntactic features, which sometimes can be learned explicitly. Specifically, one may ask to what extent epistemic meaning can be taught or learned explicitly or declaratively.

Even though it is fair to say that epistemic meanings can be acquired only inductively or by experience, an important consideration is that it is not always possible to make a direct link between explicit knowledge and explicit learning or between implicit knowledge and implicit learning (DeKeyser, 2003). In other words, explicit knowledge does not have to be the product of only explicit teaching/learning. In Bialystok's (1994) model of learning built on the cognitive processes of analysis and control, implicit knowledge may become explicit through the process of analysis. Furthermore, as Bialystok explains, both the L1 and L2 develop through the same cognitive processes of analysis and control. Therefore, it should not be surprising to see a level of explicit knowledge relating to the areas of language that are usually acquired inductively by L2 learners or even to the natively acquired L1. A recent longitudinal study by Kim and Godfroid (2023) provided evidence of the development path from explicit to implicit knowledge in L2 learners. As for the L2 participants in the current study, it merits mentioning that they are students majoring in English who study language, thus modals and their usage are included in their curriculum. Language study also involves much practice with language, reading, writing, speaking, listening, etc. It is likely that all such experience will contribute to the development of explicit knowledge in absence of any pedagogical rules and, with time, possibly to the development of implicit knowledge.

What matters here is the fact that in providing their acceptability judgments without time constraint, both L2 and L1 speakers were able to think about, analyze, and report their perceptions, be they built on declarative memory or drawn from native intuitions. Sprouse (2013) refers to AJTs as “consciously reported perceptions of acceptability that arise when native speakers attempt to comprehend a (spoken or written) utterance” (p. 97). Because the analyzed knowledge built on native intuitions is by definition more stable than acquired knowledge in a second language, it is not surprising that the L2 speakers demonstrated more variability in their acceptability judgments. In this case, the levels of certainty and variability may just reflect the levels of proficiency.

Evidence demonstrating a difference between L2 speakers' offline acceptability judgments and their online processing has previously been found in L2 studies

conducted within the explicit/implicit framework (Jegerski, 2015; Jiang et al., 2011; Roberts & Liszka, 2013; Tokowicz & Warren, 2010). The majority of these studies focused on morphosyntax such as inflectional agreement and syntactic ambiguities such as garden-path sentences. The present study extends this research agenda to modality and, along with Roberts and Liszka (2013), contributes to the investigation of L2 acquisition of tense, aspect, and modality from the perspective of sentence processing. Our results suggest that implicit knowledge as a component of language competence can be developed relatively early (Tokowicz & Warren, 2010) for some modals and some meanings—namely, those expressing ability<sup>7</sup>—but not for other meanings—namely, epistemic possibility. As the results show, native-like linguistic behavior can be seen in the L2 learners' processing of sentences containing an incongruent modal expressing ability/sensation, where both L1 and L2 groups experienced a slowdown caused by a mismatch between the modal and the context. However, L2 readers recovered sooner whereas L1 readers took more time for disambiguation. We explain this as evidence of a somewhat easier/quicker recovery from semantic ambiguity for L2 readers, which may suggest that L1 speakers, operating on their native linguistic competence, experience a greater disruption when encountering semantic ambiguity and it takes them longer to recover. Similarly, L1 speakers take longer to recover from syntactic violation, to which L2 speakers did not show sensitivity.

No significant change in reading time for either L1 or L2 speakers was observed in sentences using *may* or *can* for speaker-oriented meaning expressing permission. As previously pointed out, this function of the two modals is of a different nature than the meanings contained in their syntactic and semantic roles: speaker-oriented *may* and *can* denoting permission serve the pragmatic function. Our findings offer further proof that for giving or asking permission, both modals are today used interchangeably. Although *may* has typically been considered pragmatically suitable in more formal settings, Leech's (2003) analysis of written and spoken corpora of British and American English found that the use of *may* for permission had declined over the 30-year period from 1961 to 1991, whereas the use of *can* in formal situations had increased. It is possible that the trend is present now, too.

In addition to the substantive findings, the study provides valuable data on instrument reliability in AJT and SPR stimuli (Marsden et al., 2018) and, adding to those in Mifka-Profozic et al. (2020), shows how estimates varied across participant and instrument features. It also sheds light on the psychometric properties and error associated with these types of instrumentation. For these purposes, the instrument reliability analyses benefitted from the application of superior coefficients to the much used (but often misapplied) Cronbach's alpha (McNeish, 2018; cf. Raykov & Marcoulides, 2019), steps we encourage other researchers to take.

For the main analyses, mixed-effects regression offered a more robust and nuanced approach to modeling the SPR data than analysis of variance (Plonsky & Oswald, 2017), and ordinal mixed-effects regression offered a powerful method for handling the ordinal AJT outcome variable, avoiding the need to (mis)treat this variable as continuous. The standardized effect sizes reported allowed us to consider the magnitude of effects alongside statistical significance, providing evidence that is less confounded by sample size in standardized measurement units (standard deviations) that enable

<sup>7</sup> Anecdotal evidence from EFL teachers suggests that, at least for Croatian L1 speakers, the modal *can* is learned very early, as one of the first verbs used.

systematic comparison across studies with different designs, foci, participants, and instrumentation (Avery & Marsden, 2019). In combination, this nuanced methodology helped us build on Mifka-Profozic et al.'s (2020) findings, extending the inquiry to establish how linguistics knowledge and online processing of English modals manifest across first and second languages for these types of learners.

### *Limitations*

In conducting SPR in the current study, we made all efforts to strictly follow the procedures recommended by Keating and Jegerski (2015) and Jegerski (2012, 2014). However, we note several potential limitations. First, experimental items were designed to be as comparable as possible in terms of the number of syllables (Jegerski, 2014), but ensuring natural-sounding sentences meant there was not always perfect consistency. The first word in the sentence (Segment 0) was always either a personal pronoun or a name, the second word (Segment 1) was always the modal (may/can), the third word (Segment 2) was a one-syllable word in 34 out of 36 sentences (two words had two syllables), the fourth word (Segment 3) was a one syllable word in 29 of the target 36 sentences, and the remaining seven words had two syllables. Nevertheless, as we controlled for random effects linked to items, we believe that these minor discrepancies in the number of syllables did not affect the results.

A second possible limitation is that each target item was not always followed by a filler or a comprehension question. We used a target-item-to-filler ratio of 2:1 for 32 target items and 1:1 for four target items, with 20 randomly appearing comprehension questions, 10 following the 36 target items and 10 following the 20 fillers. The reason for such a decision was that in our study each item consisted of three sentences rather than of isolated sentences as in most other SPR studies; thus, we reduced the number of fillers to avoid the participants' fatigue. Overall, we used a ratio of 36 target sentences (one in each item of three sentences) versus 108 nontarget sentences in experimental items plus 60 filler sentences, unrelated to the target items (168 in total). In making clear these potential limitations, we hope to aid future researchers seeking to achieve the optimal balance between item authenticity, coverage, and participant attention/fatigue.

### *Suggestions for further research*

Well-attested syntactic and semantic complexities of the English modals, along with their complex pragmatic interpretations, contribute to L2 modal acquisition in various ways. However, we still do not know what the single contribution of each specific aspect may be. Therefore, further studies are needed on other modals to more fully understand how acquisition of these modals links to the type of knowledge developed. The participants in the present study were upper-intermediate speakers of L2 English, but research suggests that only highly proficient or near-native L2 speakers perform at the level comparable to that of L1 speakers in online processing (Jackson, 2008; Jegerski, 2016; Hopp 2016; Hopp & Lemmerth, 2018). It is possible that the participants in the current study are still developing their comprehension of epistemic meaning while having achieved a fairly advanced level in other aspects of linguistic performance. Immersion may play a role in accelerating the acquisition. Thus, further investigation with highly proficient, near-native L2 users of English would be welcome.



A possible factor here is the late acquisition of epistemic meaning individually in both L1 and in L2 (Giacalone Ramat, 1992; Granget et al., 2018; Ozturk & Papafragou, 2015; Traugott & Dasher, 2004) and diachronically from pragmatic and lexical means to grammaticalized forms (Bybee et al., 1994). If Giacalone-Ramat's hypothesis regarding L2 modal acquisition is correct, longitudinal studies following L2 learners from their beginner to more advanced stages would be able to show in what ways L2 acquisition takes place. In this strand of research, online sentence processing could also reveal the stages that L2 learners go through to achieve L2 competence.

**Acknowledgements.** We would like to thank the anonymous reviewers and Journal Editors for their helpful feedback on earlier versions of the article, and Norbert Vanek and Giulia Bovolenta for valuable statistical advice.

**Data availability statement.** The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at <https://osf.io/6qynj/> and via the IRIS database at <https://www.iris-database.org>.

**Competing interest.** We declare no competing interests.

## References

- Altmann, G. T. M., & Steedman, M. J. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238. [https://doi.org/10.1016/0010-0277\(88\)90020-0](https://doi.org/10.1016/0010-0277(88)90020-0)
- Avery, N., & Marsden, E. (2019). A meta-analysis of sensitivity to grammatical information during self-paced reading: Towards a framework of reference for reading time effect sizes. *Studies in Second Language Acquisition*, 41, 1055–1087. <https://doi.org/10.1017/S0272263119000196>
- Ayoun, D., & Gilbert, C. (2017). The acquisition of modal auxiliaries in English by advanced Francophone learners. In M. Howard & P. Leclercq (Eds.), *Tense-aspect-modality in a second language: Contemporary perspectives* (181–209). John Benjamins. <https://doi.org/10.1075/sibil.50.07ayo>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <http://doi.org/10.18637/jss.v067.i01>.
- Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, 16, 157–168. <http://www.jstor.org/stable/44487721>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.
- BNC Consortium. (2007). *British National Corpus, XML edition*. Oxford Text Archive. <http://hdl.handle.net/20.500.12024/2554>.
- Bolinger, D. (1989). Extrinsic possibility and intrinsic potentiality: 7 on may and can + 1. *Journal of Pragmatics*, 13, 1–23. [https://doi.org/10.1016/0378-2166\(89\)90107-0](https://doi.org/10.1016/0378-2166(89)90107-0)
- Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, 33, 247–271. <https://doi.org/10.1017/S0272263110000756>
- Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36, 1–22. <https://doi.org/10.1093/applin/amt018>
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Christensen, R. H. B. (2019). *ordinal: Regression models for ordinal data* [Computer software] (R package version 2019.12-10). R Foundation for Statistical Computing. <https://CRAN.R-project.org/package=ordinal>.
- Coates, J. (2014). *The semantics of modal auxiliaries*. Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum

- Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for non-native speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, 34, 615–646. <https://doi.org/10.1017/S0142716411000890>
- DeKeyser, R. (2003). Implicit and explicit learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of second language acquisition* (pp. 313–348). Blackwell. <https://doi.org/10.1002/9780470756492.ch11>
- Dittmar, N., & Terborg, H. (1991). Modality and second language learning: A challenge for linguistic theory. In T. Huebner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 347–384). John Benjamins. <https://doi.org/10.1075/lald.2.19dit>
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141–172. <https://doi.org/10.1017/S0272263105050096>
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Multilingual Matters. <https://doi.org/10.21832/9781847691767>
- Giacalone Ramat, A. (1992). Grammaticalization processes in the area of temporal and modal relations. *Studies in Second Language Acquisition*, 14, 297–322. <https://doi.org/10.1017/S027226310001113X>
- Gibbs, D. A. (1990). Second language acquisition of the English modal auxiliaries can, could, may, and might. *Applied Linguistics*, 11, 297–314. <https://doi.org/10.1093/applin/11.3.297>
- Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, 37, 269–297. <https://doi.org/10.1017/S0272263114000850>
- Granget, C., Dat, M.-A., Cuët, C., El Haj, P., Albochi, F., & Allawama, A. (2018). Effets de l'âge d'immersion sur la compréhension initiale d'expressions modales épistémiques en français L2. *SHS Web of Conferences*, 46, Article 10005. <https://doi.org/10.1051/shsconf/20184610005>
- Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic writing. *Journal of Pragmatics*, 41, 667–683. <https://doi.org/10.1016/j.pragma.2008.09.029>
- Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, 22, 369–397. <https://doi.org/10.1191/0267658306sr272oa>
- Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, 37, 1253–1280. <https://doi.org/10.1017/S0142716415000569>
- Hopp, H., & Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, 40, 171–199. <https://doi.org/10.1017/S0272263116000437>
- Hsu, C.-H., Lee, C.-Y., & Marantz, A. (2011). Effects of visual complexity and sublexical information in the occipitotemporal cortex in the reading of Chinese phonograms: A single-trial analysis with MEG. *Brain and Language*, 117, 1–11. <https://doi.org/10.1016/j.bandl.2010.10.002>
- Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, 58, 875–909. <https://doi.org/10.1111/j.1467-9922.2008.00481.x>
- Jegerski, J. (2012). The processing of subject–object ambiguities in native and near-native Mexican Spanish. *Bilingualism: Language and Cognition*, 15, 721–35. <https://doi.org/10.1017/S1366728911000654>
- Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.
- Jegerski, J. (2015). The processing of case in near-native Spanish. *Second Language Research*, 31, 281–307. <https://doi.org/10.1177/0267658314563880>
- Jegerski, J. (2016). Number attraction effects in near-native Spanish sentence comprehension. *Studies in Second Language Acquisition*, 38, 5–33. <https://www.jstor.org/stable/26330996>
- Jegerski, J. (2018). Sentence processing in Spanish as a heritage language: A self-paced reading study of relative clause attachment. *Language Learning*, 68, 598–634. doi:10.1111/lang.12289
- Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes: Morphological congruency. *Language Learning*, 61, 940–967. <https://doi.org/10.1111/j.1467-9922.2010.00627.x>
- Just, M. A., Carpenter, P., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228–238. <https://doi.org/10.1037//0096-3445.111.2.228>
- Kaltsa, M., Tsimpli, I. M., Marinis, T., & Stavrou, M. (2016). Processing coordinate subject-verb agreement in L1 and L2 Greek. *Frontiers in Psychology*, 7, Article 648. <https://doi.org/10.3389/fpsyg.2016.00648>

- Keating, G., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, 37, 1–32. <https://doi.org/10.1017/S0272263114000187>
- Kim, K. M., & Godfroid, A. (2023). The interface of explicit and implicit second-language knowledge: A longitudinal study. *Bilingualism: Language and Cognition*, 26, 709–723. <https://doi.org/10.1017/S1366728922000773>.
- Lee, J. F., Malovrh, P., Doherty, S., & Nichols, A. (2022). A self-paced reading (SPR) study of the effects of processing instruction on the L2 processing of active and passive sentences. *Language Teaching Research*, 26, 1133–1157. <https://doi.org/10.1177/1362168820914025>
- Leech, G. (2003). Modality on the move: The English modal auxiliaries 1961–1992. In R. Facchinetti, M. G. Krug & F. R. Palmer (Eds.), *Modality in contemporary English* (pp. 223–240). Mouton de Gruyter. <https://doi.org/10.1515/9783110895339.223>
- Lüdecke D., Ben-Shachar M., Patil I., Waggoner P., & Makowski D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6, Article 3139. <https://doi.org/10.21105/joss.03139>
- Lyons, J. (1977). *Semantics, volume 2*. Cambridge University Press.
- Maie, R., & Godfroid, A. (2022). Controlled and automatic processing in the acceptability judgment task: An eye-tracking study. *Language Learning*, 72, 158–197. <https://doi.org/10.1111/lang.12474>
- Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, 39, 861–904. <https://doi.org/10.1017/S0142716418000036>
- Mifka-Profozic, N. (2017). Processing epistemic modality in a second language: A self-paced reading study. *IRAL International Review of Applied Linguistics in Language Teaching*, 55, 245–264. <https://doi.org/10.1515/iral-2017-0107>
- Mifka-Profozic, N., O'Reilly, D., & Guo, J. (2020). Sensitivity to syntactic violation and semantic ambiguity in English modal verbs: A self-paced reading study. *Applied Psycholinguistics*, 41, 1017–1043. <https://doi.org/10.1017/S0142716420000338>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23, 412–433. <https://doi.org/10.1037/met0000144>
- Nicklin, C., & Plonsky, L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics*, 40, 26–55. <https://doi.org/10.1017/S0267190520000057>
- Norris J., & Ortega, L. (2013). Assessing learner knowledge. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 573–589). Routledge. doi: 10.4324/9780203808184
- Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development*, 11, 191–214. <https://doi.org/10.1080/15475441.2014.905169>
- Palmer, F. (1986). *Mood and modality*. Cambridge University Press.
- Palmer, F. (1990). *Modality and the English modals*, 2nd ed. Longman.
- Palmer, F. (2003). Modality in English: Theoretical, descriptive and typological issues. In R. Facchinetti, M. Krug, & F. Palmer (Eds.), *Modality in contemporary English* (pp. 1–17). Mouton de Gruyter. <https://doi.org/10.1515/9783110895339.1>
- Papafragou, A., & Ozturk, O. (2006). *Children's acquisition of epistemic modality*. Cascadilla Proceedings project.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. John Benjamins. <https://doi.org/10.1075/sibil.40>
- Peirce, J. W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods*, 162, 8–13.
- Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform*, 2. <https://doi.org/10.3389/neuro.11.010.2008>
- Platiatskias, C., & Marinis, T. (2013). Processing of regular and irregular past tense morphology in highly proficient second language learners of English: A self-paced reading study. *Applied Psycholinguistics*, 34, 943–970. <https://doi.org/10.1017/S0142716412000082>
- Plonsky, L., & Derrick, D. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, 100, 538–553. <https://doi.org/10.1111/modl.12335>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting  $R^2$  values. *The Modern Language Journal*, 102, 713–731. <https://doi.org/10.1111/modl.12509>

- Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, 36, 583–621. <https://doi.org/10.1177/0267658319828413>
- Plonsky, L., & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. <https://doi.org/10.1111/lang.12079>
- Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, 39, 579–592. <https://doi.org/10.1017/S0272263116000231>
- R Core Team. (2022). *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing. <https://www.R-project.org/>)
- Raykov, T., & G. A. Marcoulides. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, 79, 200–210. <https://doi.org/10.1177/0013164417725127>
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, 63, 595–626. <https://doi.org/10.1111/lang.12010>
- Roberts, L., & Felser, C. (2011). Plausibility and recovery from garden paths in second language sentence processing. *Applied Psycholinguistics*, 32, 299–331. <https://doi.org/10.1017/S0142716410000421>
- Roberts, L., & Liszka, S. (2013). Processing tense/aspect agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, 29, 413–439. <https://doi.org/10.1177/0267658313503171>
- Spinner, P., & Gass, S. M. (2019). *Using judgments in second language acquisition research*. Routledge.
- Sprouse, J. (2013). *Acceptability judgments*. Oxford University Press. <https://doi.org/10.1093/obo/9780199772810-0097>
- Stewart, A. J., Haigh, M., & Kidd, E. (2009). An investigation into the online processing of counterfactual and indicative conditionals. *Quarterly Journal of Experimental Psychology* 62, 2113–2125. <https://doi.org/10.1080/17470210902973106>
- 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. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge: Elicited imitation and word monitoring. *Language Learning*, 65, 860–895. <https://doi.org/10.1111/lang.12138>
- Suzuki, Y., & DeKeyser, R. (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>
- Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, 22, 1092–1106. <https://doi.org/10.1080/09541440903325178>
- Torchiano, M. (2020). *effsize: Efficient effect size computation* [Computer software] (R package version 0.8.1). R Foundation for Statistical Computing. <https://CRAN.R-project.org/package=effsize>.
- Traugott, E. C., & Dasher, R. B. (2004). *Regularity in semantic change*. Cambridge University Press. <https://doi.org/10.1075/aila.23.03tyl>
- Vafaei, P., Suzuki, Y., & Kachisnke, 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>
- Voeten, C. C. (2022). *buildmer: Stepwise elimination and term reordering for mixed-effects regression* [Computer software] (R package version 2.3). R Foundation for Statistical Computing. <https://CRAN.R-project.org/package=buildmer>
- Wells, C. G., & Nicholls, J. (1985). *Language and learning: An interactional perspective*. Falmer Press.
- Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, 37, 457–486. <https://doi.org/10.1017/S0272263114000370>

---

**Cite this article:** Mifka-Profozic, N., O'Reilly, D., & Lovrovic, L. (2023). Unpacking L2 explicit linguistic knowledge and online processing of the English modals *may* and *can*: A comparison of acceptability judgments and self-paced reading\*. *Studies in Second Language Acquisition*, 1–28. <https://doi.org/10.1017/S0272263123000475>