Studies in Second Language Acquisition (2024), **46**, 355–377
doi:10.1017/S0272263123000517

RESEARCH ARTICLE

Usage events and constructional knowledge: A study of two variants of the introductory-*it* construction*

Sakol Suethanapornkul[1] and Sarut Supasiraprapa[2]

[1]Independent scholar and [2]Graduate School of Language and Communication, National Institute of Development Administration, Thailand
Corresponding author: Sarut Supasiraprapa; Email: sarut.sup@nida.ac.th

(Received 09 February 2022; Revised 21 August 2023; Accepted 11 September 2023)

Abstract

Usage-based theories hold that mental representation of language is shaped by a lifetime of usage. Both input to which first language (L1) and second language (L2) users are exposed and their own language production affect their construction learning and entrenchment. The present study investigates L2 users' knowledge of two introductory-*it* variants, Adj-that (e.g., *it is clear that* …) and Adj-to (e.g., *it is difficult to* …). We probed the extent to which adjective–variant associations in an academic section of COCA and L2 users' engagement with academic writing affected learners' generation of adjectives distinctively attracted to the two variants. An analysis of cue-outcome contingency was conducted to establish adjective–variant associations, and an elicitation task was carried out, probing L2 users' ability to generate adjectives when prompted with the variants (e.g., *it is [blank] to*). The participants were 84 graduate students in the United States, 44 from L1 English and 40 from L1 Thai backgrounds. The results indicated that the adjective–variant associations predicted L2 users' generation of adjectives. However, academic writing engagement did not affect learners' performance. The findings suggest that statistical information in the input affects L2 users' constructional representation.

Introduction

Usage-based theories hold that our mental representation of language is shaped by a lifetime of usage. Although the precise nature of such representation remains hotly contested (e.g., Ambridge, 2020), a consensus emerges that usage events shape first (L1) and second language (L2) users' constructional knowledge and use. There is substantial evidence that L1 and L2 users are sensitive to frequencies of constructions of all grain sizes and recruit this information in processing and learning (Ambridge et al., 2015; Supasiraprapa, 2019). In addition to frequencies of individual linguistic

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

© The Author(s), 2023. Published by Cambridge University Press.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

items, L1 and L2 users have been found to encode statistical associations of lexemes and constructions. For instance, the degree to which verbs (e.g., *regard*) were attracted to the *as*-predicative construction (e.g., *We regarded it as a serious problem*)—as established by verb-construction association strength scores—was found to predict L1 English speakers' use of the construction (Gries et al., 2005). Likewise, L1 Dutch speakers' production of double-object and prepositional dative constructions was explained by association strength of verbs and constructions (Colleman & Bernolet, 2012). With respect to L2 users, Ellis and Ferreira-Junior (2009) found that verbs produced by naturalistic L2 English learners in verb-argument constructions (VACs) such as a verb locative (VL) construction (e.g., *he walked across the field*) were those that are strongly attracted to the constructions, whereas Gries and Wulff (2009) demonstrated that the strength of associations between verbs and *to*-infinitive/gerundial complements (e.g., *keep going* vs. *want to know*) significantly predicted a complementation pattern that L1 German-L2 English learners supplied in a sentence-completion task (see also Azazil, 2020). In addition, the effect of verb-construction contingencies can be observed in L2 comprehension. L1 Korean-L2 English learners rated resultative sentences (e.g., *David made the room dark*) as more grammatical when they were paired with strongly attracted verbs (e.g., *make* or *get*) than when paired with weakly attracted verbs (e.g., *wipe* or *paint*; Sung & Kim, 2022).

Although extensive research during the past 2 decades has demonstrated that frequency and other forms of statistical regularities in the input shape L1 and L2 speaker's knowledge of constructions, usage-based researchers have long argued that usage events constitute both language input and output (Kemmer & Barlow, 2000) and that our linguistic knowledge is shaped as much by what we produce as by what we comprehend (Ellis, 2019, 2022; Schmid, 2015). For example, as Ellis (2022) clearly pointed out, hearing, reading, and speaking are usage events that promote construction learning and entrenchment. Despite this long-standing view of "usage," we are not aware of any studies that investigate how L2 speakers' own constructional use—independent of input statistics to which L2 users are exposed—further affects their constructional knowledge.

Addressing this research gap, the current study attempted to operationalize language output and probed the effect of both input and output on L2 users' constructional knowledge. We focused on two closely related adjectival complementation patterns in the introductory-*it* construction, as in *it is clear that their plan wasn't successful* (an adjective + *that* clause, which we will refer to as an Adj-that variant) and *it is difficult to know what they stand for* (an adjective + *to*-infinitive clause or an Adj-to variant henceforth). Corpus studies have found that the two variants—along with other patterns—of the introductory-*it* construction are attested much more frequently in academic discourse (Biber et al., 1999; Larsson, 2016). This focus afforded us an opportunity to investigate the role of both input and output on L1 and L2 users' academic knowledge of this construction, as they are highly likely to be exposed to this construction and use it in their academic studies. Specifically for the two variants, we found via our cue-outcome contingency analysis with the measure ΔP (Gries & Ellis, 2015) that each variant did attract certain adjectives, which will be reported in the Method. With this in mind, we asked whether L1 Thai-L2 English users, who were graduate students at U.S. universities, would be able to generate adjectives that were distinctly associated with the target variants (i.e., *cued responses* of the variants). In an elicitation task, L2 users were prompted with schematic frames (e.g., *It is [blank] that …*) and instructed to complete each one with as many words as they could within an allotted time. We reasoned that, through constant exposure to and active use of

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

academic English, this group of L2 users would be attuned to adjective–variant contingencies, thus being able to supply cued responses of the target variants. To test the effect of input statistics on L2 users' constructional knowledge, we predicted learners' responses on the elicitation task from the association strength of adjectives and variants in the academic section of Corpus of Contemporary American English (COCA), which contains approximately 121 million words in peer-reviewed academic articles from a wide variety of disciplines (Davies, 2019). Thus, as in previous usage-based research (Conklin & Thul, 2023), we used data from a large corpus as an index of language input to which L1 and L2 users are exposed. Further, to test whether learners' own constructional use affects their knowledge, L2 users rated how frequently they completed various academic writing tasks in their graduate studies. We entered this indirect measure of constructional use into a statistical analysis. Finally, as L2 construction use expands as a function of L2 proficiency (e.g., Eskildsen, 2012; Römer & Berger, 2019), we also probed whether L2 users' generation of cued responses can be accounted for by their L2 proficiency.

The two variants of the introductory-*it* construction

The introductory-*it* construction, also known as anticipatory-*it* or *it*-extraposition, consists of various patterns, some of which are shown in (1). What these diverse patterns share syntactically is a matrix predicate with the subject pronoun *it* that does not have anaphoric reference and a finite (e.g., *that*-clause) or nonfinite (e.g., *to*-infinitive) subordinate clause (Quirk et al., 1985). The construction serves primarily to introduce and evaluate propositions in an objective and neutral manner (Groom, 2005; Herriman, 2000; Hewings & Hewings, 2002). Objectivity is achieved, it is argued, through the use of the dummy *it*. With personal pronouns as subject of the matrix predicate (e.g., *I am glad the event went well*), an evaluation can be attributed to a specific person. Conversely, an evaluator is obscured with the pronoun *it* (e.g., *it is good that the event went well*). That an evaluative stance toward propositions cannot be assigned to a specific person helps depersonalize an evaluation (Kaltenböck, 2005) and creates the impression that the evaluation is impersonal and objective. The introductory-*it* construction is thus well suited to academic writing, which emphasizes objectivity, among many other things (Hyland & Jiang, 2017), and previous corpus studies have shown that the construction is highly frequent in academic prose (Biber et al., 1999; Groom, 2005).

(1) a. *It is a good idea that schools are designed for shared use.*
b. *It seems that they cannot handle the situation.*
c. *It is not apparent why the rule should apply.*
d. *It is clear that a resolution of this controversy will require communication* (COCA ACA 4003682).
e. *It is difficult to envisage how elicitation of verbal responses, in isolation, might produce a similar effect* (COCA ACA 4001311).

The present study focuses on two patterns of the introductory-*it* construction: the Adj-that (schematically, *it V ADJ that*) and Adj-to variants (*it V ADJ to-infinitive*), as exemplified by (1d) and (1e), respectively. We investigated the two variants primarily for two reasons. First, they were the two most frequent types of the construction in academic prose (Larsson, 2016, 2017), and we expected our target participants to be familiar with such a genre and by extension the two variants because of their graduate-level studies. Second, although less frequent patterns (e.g., *it V that* as exemplified by *it*

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

follows that …) can be restrictive with respect to lexical items they select, the two variants under investigation do accept a diverse set of adjectives, thus affording us the opportunity to elicit multiple responses per variant in our experiment.

Previous corpus studies have shown that the two variants perform specific rhetorical functions in academic English (i.e., Adj-that assessing the likelihood or validity of propositions and Adj-to evaluating the necessity or difficulty of procedures); consequently, they seem to co-occur with adjectives that are semantically compatible (Groom, 2005; Herriman, 2000; Larsson, 2016). For example, inspecting Tables 4 and 6 in Peacock (2011), we can see that *clear* and *evident* appear to be used only with Adj-that and *difficult* and *hard* only with Adj-to (see also Groom, 2005, p. 269). Though some adjectives such as *important* and *possible* do occur with both variants, they do not encode the same meaning. For example, *possible* + *to-infinitive* denotes the difficulty of a process, whereas *possible* + *that* functions as a hedge (Hewings & Hewings, 2002; Peacock, 2011). Therefore, the two constructions are not synonymous (Römer, 2009).

In this study, we identified the degree to which individual adjectives were attracted to one of the introductory-*it* variants. A similar type of analysis was previously done to identify adjective–infinitive pairings in the Adj-to variant (e.g., *it is important to note* …; Larsson & Kaatari, 2019). Usage-based studies have consistently shown that association strength between lexemes and constructions in the input determines L2 constructional learning and use. For example, Ellis and Ferreira-Junior (2009) found that verbs that were distinctively associated with VACs (e.g., *go* for the VL construction) were acquired early. Likewise, words that are strongly attracted to particular constructions are judged to be more acceptable in language comprehension and accessed first in language production (Ellis et al., 2014; Gries & Wulff, 2009; Sung & Kim, 2022). Our cue-outcome contingency analysis was conducted to establish the adjective–variant associations in both Adj-that and Adj-to. The objective was to probe whether attraction between adjectives and variants in the input would play an instrumental part in shaping the knowledge of the introductory-*it* construction of competent L2 users, who have had extensive exposure to the construction through their academic studies.

Assessing the contribution of language production in L2 users' constructional knowledge

Although the importance of input has been consistently demonstrated in L2 studies from usage-based perspectives, input itself is not the sole determinant of L2 constructional knowledge. Usage-based researchers have long argued that L1 and L2 users' linguistic knowledge is built up from what users comprehend in their idiosyncratic linguistic environment as well as from what they themselves produce to accomplish communication goals (Kemmer & Barlow, 2000). More recent proposals in usage-based models have also placed both language input and output at the center of language development and use (Ellis, 2019, 2022; Schmid, 2015).

Despite an equally important role of input and output in most—if not all—usage-based models, research investigating L2 users' knowledge and use of abstract, schematic constructions such as VACs has focused almost exclusively on the effect of input statistics, as evidenced by the body of research reviewed thus far (e.g., Azazil, 2020; Ellis et al., 2014). Although more recent usage-based studies have begun to address L2 output, the focus is still on determining the effect of input. Crossley et al. (2016), for example, showed that the words L2 users produced were the ones they heard in their input.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

In this study, we sought to establish the relative contribution of L2 users' own language production on their constructional knowledge. Specifically, we quantified how actively L2 users engaged in academic writing and related this indirect measure of active constructional use to L2 users' performance in an elicitation task. Dąbrowska (2015) demonstrated that L1 speakers' constructional knowledge differed as a function of linguistic experience. For instance, graduate students and university lecturers performed significantly better than skilled workers on English comprehension tests involving complex structures such as parasitic gaps. Such a difference, Dąbrowska argues, is because university students and professors are exposed to more complex language and they also *produce* complex structures more often.

Previous corpus studies of the introductory-*it* construction have likewise suggested that the use of the two variants may be affected by how often L2 writers produce the forms themselves (Römer, 2009), although this issue has never been empirically validated with a statistical analysis. Evidence to date has come from corpus studies comparing L2 writers with published authors. The latter group, which presumably consists of both L1 and L2 English writers, is generally thought to make more use of the construction.[1] It was found, for example, that as opposed to more experienced writers, L2 users often made use of "extreme" adjectives that are not suitable for academic writing (e.g., *dangerous*, *fascinating*, and *funny*; Larsson, 2019; Römer, 2009).

How might L2 users' active use of constructions further promote their constructional knowledge? Regarding the introductory-*it* construction, it may be that to conform to an academic register, L2 users will more likely than not make use of the construction in their own language production. Thus, the more L2 users perform academic writing, the more opportunities that they have to use both Adj-that and Adj-to variants. This in turn should attune L2 users to statistical associations between adjectives and variants that exist in the construction. To explore this possibility, we collected as part of the experiment participants' self-rated academic writing engagement and used the scores to indirectly assess L2 users' active use of the introductory-*it* construction.

The current study

In this study, we assessed the relative contribution of the two sides of usage to L2 users' constructional knowledge. Specifically, we asked whether adjective–variant association strength in the input and L1 Thai-L2 English users' engagement with academic writing, a proxy for language users' production of introductory-*it* variants (among many other constructions), would affect learners' generation of lexemes of the Adj-that and Adj-to variants. To elicit adjectives that were attracted to each variant, we adapted a task that has been used extensively to assess L1 and L2 speakers' knowledge of verb–VAC associations (e.g., Ellis et al., 2014; Römer et al., 2020). Although the role of statistical contingencies in the input has been well documented for L2 processing and learning (Ellis & Ferreira-Junior, 2009; Gries & Wulff, 2009), the effect of L2 users' own production on their constructional knowledge is largely underexplored. Additionally, previous corpus-based studies have demonstrated that L2 users' constructional knowledge expands as a function of their L2 proficiency. For example, more advanced L2 users produce more types of VACs and a wider range of verbs within each VAC (Römer

[1]In several corpus studies, published authors are generally thought to have more academic writing experience by virtue of their research publications.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

& Berger, 2019). As a result, we also tested whether English L2 proficiency—in addition to academic writing engagement—would increase L2 users' ability to generate cued responses of the Adj-that and Adj-to variants. The two research questions guiding the present study are as follows:

1. To what extent do adjective–variant association strength in the input and engagement with academic English writing predict L2 users' generation of cued responses?
2. To what extent is L2 users' generation of cued responses explained by their L2 English proficiency?

Given the available evidence in usage-based literature, it is predicted that L2 users' ability to supply lexemes, when prompted with introductory-*it* variants, will be influenced by statistical associations of adjectives and variants that are present in the input. We further hypothesize that the degree to which L2 users engage in various academic writing tasks will positively affect their performance on the elicitation task. Specifically, we expect L2 users who perform academic writing tasks more frequently overall to be able to supply more cued responses of the target variants. Finally, we predict that English L2 proficiency will positively affect L2 users' responses. We test our prediction in a mixed-effects logistic regression model; significant effects of one or more predictors offer evidence in support of our prediction(s).

Method

Participants

We recruited 89 participants from two L1 backgrounds: English and Thai.[2] The L1 Thai-L2 English users were recruited to guard against any transfer effect, as it is generally argued that there is no equivalent construction for the two variants in Thai (Chainarongdejagul, 2018). Both groups were graduate students pursuing a master's or doctoral degree in the United States at the time of data collection. Five participants were subsequently excluded for the following reasons: currently not pursuing a degree (n = 1), not following the instructions (n = 2), and not pursuing a graduate degree in the United States (n = 2). Thus, 84 participants remained. The L2 group consisted of 40 L1 Thai-L2 English speakers (21 PhDs and 19 master's degrees; 23 female, 16 male, and 1 nonspecified). The L2 participants' average TOEFL iBT score was 102 out of 120 (SD = 10.3, range: 78–117). The L2 group started learning English at approximately 6.1 years of age (SD = 3.37). Despite having an early start, all but two participants reported not having studied or lived in an English immersion environment before the age of 10. Their stay in an English-speaking country ranged from 8 months to 8 years (M = 3.21 years, SD = 2.02). The other two participants indicated having been exposed to English in a more immersive environment (i.e., living in a family or studying in a school where English was spoken) at an age earlier than 10.[3] In addition to the L1 Thai-L2 English group, 44 native English-speaking participants were recruited (23 PhDs and 21 master's degrees; 25 female and 19 male).

[2]To reduce researcher degrees of freedom, recruitment continued until the number of participants per degree per L1 reached 20.

[3]Removing these two participants from the analysis did not change the results presented in Table 9. We therefore decided to retain responses from the participants in our analysis.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Experiment

Stimuli

To construct target items, each of the two introductory-*it* variants was paired with two linking verbs, *is* and *seem* (e.g., *it is [blank] that …* and *it seems [blank] that …*). These two verbs were chosen to represent verbs appearing in the construction (Francis et al., 1998). Doing so yielded a total of four target items. Filler items were constructed from two types: "*It is a/an [blank] noun*" and "*It seems like a/an [blank] noun*" such that they resembled the target items. Fillers, however, required attributive rather than predicative adjectives. Our decision to elicit adjectives even with fillers was informed by responses collected during our piloting with a different sample of 23 participants. It was found that with fillers eliciting verbs (e.g., *It [blank] across*), nonadjective answers in the target items accounted for approximately 45% of the total responses; we obtained far fewer nonadjective responses in the target items with the current set of fillers, as will be shown below. To construct fillers, we selected nouns from the Academic Word List (Coxhead, 2000) and, for each one, performed searches on an academic section of COCA. We ensured that the first 50 frequency-sorted attributive adjectives of each noun were not among the top 25 adjectives strongly cued by either variant. This was done to guard against any spillover effect from filler to target items. There were nine fillers in total.

Writing experience

Participants completed an 18-item academic writing experience questionnaire (AWE-Q), developed by the first author of the present study and validated with item response theory models in a previous unpublished study (Suethanapornkul & McKay, 2018). Designed to gauge writers' experience with academic writing, the questionnaire asked participants to rate, on a scale from 1 (*very rarely*) to 5 (*very often*), how frequently they performed each of the 18 writing tasks commonly found in academic settings (e.g., submitting an abstract for a conference presentation, submitting a grant proposal, and writing a part of the thesis or dissertation). Participants checked "Does not apply" if certain tasks were not applicable. The experimental stimuli and questionnaire can be found in the Online Supplementary Materials.

For each participant, we dropped items with no responses and calculated average scores, which served as an indirect measure of constructional use in both L1 English and L1 Thai-L2 English participants. The L1 English group engaged in academic writing somewhat frequently ($M = 3.43$, $SD = 0.88$, range: 1.78–5); academic-related emails ($M = 4.57$, $SD = 0.76$) and course papers ($M = 4.40$, $SD = 0.96$) were the two most common tasks the group performed. The average scores of the L2 English group were similar ($M = 3.60$, $SD = 0.96$, range: 1.56–5), with academic-related emails ($M = 4.75$, $SD = 0.67$) and academic presentations ($M = 4.24$, $SD = 0.89$) being the two most common tasks.

Procedure

The full experiment was delivered via Qualtrics. Participants were informed that they would complete an academic English vocabulary test by typing in as many academic English words that were appropriate for a given item in 1 min. After signing an electronic consent form, participants took three practice trials (e.g., *as a/an [blank] of …*). For each one, after participants completed their answers, the experiment

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

presented three commonly used words in academic writing (e.g., the three words for the frame *as a/an [blank] of …* were *example*, *percentage*, and *result*).

For the main part of the experiment, each participant saw two target items, one from each variant, interspersed with three fillers, selected randomly from the nine items. The order of the presentation of these two target items was counterbalanced across participants. For methodological comparability, we followed the procedure in Ellis et al. (2014) and Römer et al. (2020). Participants completed each of the five trials by typing their answers in the boxes provided, and their responses remained on the screen until the end of each trial (= 1 min). Once the elicitation task was completed, the participants took the AWE-Q and completed a language background questionnaire, also on Qualtrics.

Adjective–variant association strength

To establish adjective–variant association strength in the input of academic English to which language users were potentially exposed, we computed a directional cue-outcome contingency measure, ΔP (Gries & Ellis, 2015). The measure expresses the likelihood of an outcome occurring when a cue is present as opposed to when it is absent. There are two versions of ΔP; the two differ with respect to what linguistic element is the cue and what the outcome is (Gries, 2013). In the context of the present study, the cue can be an adjective and the outcome a variant or vice versa. Because in the experiment participants were given the two introductory-*it* variants and prompted to complete an empty slot with adjectives, we calculated a version of ΔP where the presence or absence of a variant was the cue and the choice of adjective was the outcome and represented such contingency as $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$. For each adjective–variant pair, a ΔP score thus indicates how strongly the variant cues the adjective.

To obtain the $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores, we first retrieved from an academic section of COCA through the CQPweb interface (Hardie, 2012) instances of the two introductory-*it* variants. Our search terms combined part-of-speech tags for verbs with a length requirement, from one to three words, to retrieve different verb tenses and modal verbs (e.g., *is*, *seemed*, *might be*, and *could have been*). The terms also allowed for intervening adverbial and prepositional phrases before or after an adjective (e.g., *it is important for researchers to understand the process*). Last, the queries included the *to*-infinitive and *that*-clause; thus, only instances where *that* was explicitly mentioned were retrieved (see Hyland & Jiang, 2018, for a similar approach). We retrieved 44,120 hits in total (16,339 Adj-that and 27,781 Adj-to). The search terms, along with example sentences retrieved, are included in the Online Supplementary Materials.

A concordance file was generated such that the two variants were the node with surrounding contexts (+10 words on each side). We then manually removed instances where *it* had anaphoric reference (e.g., *it was able to perform …*) or the *to*-infinitive or *that*-clause was preceded by other elements but adjectives (e.g., *it was good proof that …*). We independently coded the first 10% of the file and calculated an interrater agreement, which was 96%. We resolved any discrepancy in our coding through discussion before further coding. After manual editing, we retained 15,862 and 23,627 instances of Adj-that and Adj-to, respectively. Our final preprocessing step included extracting adjectives from the nodes and coding the variant with which each adjective was attested. We additionally lemmatized comparative and superlative

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Table 1. Descriptive information of the two target variants

Variants	Types	Tokens	Type-token ratios (TTRs)	Hapax legomena (%)	H_{norm}
Adj-that	262	15,862	1.65	49 (18.7)	0.60
Adj-to	495	23,627	2.10	152 (30.7)	0.59

Note. H_{norm} = normalized entropy.

adjectives (e.g., *good* for *best* and *better*) and made the spelling consistent (e.g., *okay* for *o.k.* and *ok*). Table 1 presents descriptive information for the two variants. In total, there were 585 types: 262 types appeared with Adj-that and 495 with Adj-to. Of the 262 types, 90 were attested only with Adj-that (49 hapaxes). Likewise, of the 495 types, 323 types appeared solely with Adj-to (152 hapaxes). The two variants shared 172 types. Table 2 presents the top 15 most frequent adjectives of each variant, along with co-occurrence frequencies with each of the constructions.

In addition to the two target variants, we identified and extracted from COCA five other instantiations of the introductory-*it* construction, as documented in Francis et al. (1998) and Larsson (2016). In each of these patterns, a finite or nonfinite subordinate clause is preceded by a matrix predicate with an adjective as the predicate lemma (that is, an element that contributes most to the meaning of the predicate; see Larsson, 2016). Schematic examples of these patterns are *it V ADJ wh-* (e.g., *it is not clear whether the test was reliable*) and *it V N as ADJ that* (e.g., *it strikes me as rather odd that nothing was accomplished*). Including frequencies of adjectives in these five patterns into a total count enabled us to compute two $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores for every adjective, one for each variant, such as $\Delta P_{(\text{Adj-that} \rightarrow \text{possible})}$ and $\Delta P_{(\text{Adj-to} \rightarrow \text{possible})}$. In total, we identified 2,300 instances of the five patterns; examples of each pattern can be found in the Supplementary Materials.

A $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ score for each adjective–variant combination was calculated from frequency counts in a 2 × 2 contingency table as differences of proportions (Gries & Ellis, 2015, p. 240), and its values range from −1 to +1. Positive ΔP values indicate

Table 2. The 15 most frequent adjectives of each variant in COCA

Adj-that				Adj-to			
		Frequencies in				Frequencies in	
Ranks	Adjectives	Adj-that	Adj-to	Ranks	Adjectives	Adj-to	Adj-that
1	clear	**2,826**	2	1	important	**4,615**	853
2	possible	**1,908**	2,374	2	difficult	**2,662**	0
3	likely	**1,281**	58	3	possible	**2,374**	1,908
4	true	**1,084**	13	4	necessary	**1,587**	120
5	surprising	**917**	85	5	easy	**1,512**	0
6	important	**853**	4,615	6	hard	**1,318**	0
7	unlikely	**785**	22	7	impossible	**1,132**	41
8	apparent	**610**	1	8	interesting	**648**	228
9	evident	**504**	0	9	reasonable	**523**	49
10	obvious	**477**	1	10	good	**380**	26
11	imperative	**315**	134	11	useful	**332**	3
12	essential	**271**	319	12	essential	**319**	271
13	noteworthy	**238**	3	13	appropriate	**270**	46
14	significant	**230**	34	14	safe	**263**	0
15	interesting	**228**	648	15	helpful	**212**	1

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Table 3. The 2 × 2 co-occurrence frequencies of clear

Variants	*clear*	¬ *clear*	Row totals
Adj-that	2,826	13,036	**15,862**
¬ Adj-that	523	25,404	**25,927**
Column totals	**3,349**	**38,440**	**41,789**

that the presence of the cue increases the likelihood of the outcome, whereas negative values indicate the opposite. To illustrate, we can establish how strongly Adj-that cues *clear* by obtaining frequencies when the adjective appears with the variant ($n = 2{,}826$) as opposed to when it occurs with Adj-to and other instantiations ($n = 523$) and summing frequencies of all other adjectives that are attested with Adj-that and with the rest of the patterns but Adj-that (see Table 3). Here, an observed value of $\Delta P_{(\text{Adj-that} \rightarrow \text{clear})}$ is 0.16 (that is, from $\frac{2{,}826}{(2{,}826 + 13{,}036)} - \frac{523}{(523 + 25{,}404)}$). Similarly, we can assess how strongly Adj-to attracts *clear* by obtaining co-occurrence frequencies between the two elements ($n = 2$), along with other frequency counts, and constructing a 2 × 2 contingency table, similar to Table 3. Inputting the raw frequencies into a formula, we obtain a $\Delta P_{(\text{Adj-to} \rightarrow \text{clear})}$ score of −0.18. We can therefore see that Adj-that attracts *clear* but Adj-to repels it. We calculated observed $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores for all of the 585 types identified from the corpus analysis.

The ΔP scores capture adjective–variant contingencies that are present in the input of academic English to which language users are exposed. To ensure that the ΔP scores we included in our statistical analysis (see below) solely reflected such associations, we completed additional steps, as discussed in Gries (2022) and recommended by one of the reviewers. First, we computed two theoretically possible ΔP scores for each adjective–variant combination. For example, Table 4 presents one scenario, in which all instances of *clear* occur with Adj-that. In this case, a ΔP score would indicate *the strongest possible attraction* between Adj-that and *clear*. As all counts of *clear* are now in the upper-left cell (that is, 2,826 + 523 = 3,349), the frequency counts in the other cells are adjusted so that they sum to actual frequencies of adjectives and variants in the column and row totals, respectively. Note that the actual frequencies are kept unchanged from Tables 3 to 4; this ensures that the effect of frequency is controlled for (see Gries, 2022, p. 24). We obtain a $\Delta P_{(\text{Adj-that} \rightarrow \text{clear})}$ score of 0.21 from Table 4. For the other scenario, none of the instances of *clear* are in Adj-that (that is, the value of 3,349 is moved to the lower-left cell), so a ΔP score would indicate *the strongest possible repulsion* between Adj-that and *clear*. In this scenario, we obtain a ΔP score of −0.13 (that is, from $\frac{0}{(0 + 15{,}862)} - \frac{3{,}349}{(3{,}349 + 22{,}578)}$). Second, we determined where observed ΔP scores were with respect to the two theoretically possible ΔP scores on either end of the repulsion–attraction

Table 4. A contingency table demonstrating the strongest possible attraction between Adj-that and clear

Variants	*clear*	¬ *clear*	Row totals
Adj-that	3,349	12,513	**15,862**
¬ Adj-that	0	25,927	**25,927**
Column totals	**3,349**	**38,440**	**41,789**

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Table 5. The top 15 adjectives of each variant ranked by observed ΔP scores

Attracted to Adj-that				Attracted to Adj-to			
		$\Delta P_{(\text{Adj-that} \rightarrow \text{adjective})}$				$\Delta P_{(\text{Adj-to} \rightarrow \text{adjective})}$	
Ranks	Adjectives	Observed	Transformed	Ranks	Adjectives	Observed	Transformed
1	clear	0.16	0.84	1	important	0.15	0.84
2	likely	0.08	0.96	2	difficult	0.11	0.99
3	true	0.07	0.99	3	easy	0.06	0.98
4	surprising	0.05	0.88	4	necessary	0.06	0.93
5	unlikely	0.05	0.97	5	hard	0.06	0.99
6	apparent	0.04	0.97	6	impossible	0.05	0.96
7	evident	0.03	0.97	7	reasonable	0.02	0.91
8	obvious	0.03	0.88	8	useful	0.01	0.99
9	possible	0.03	0.44	9	good	0.01	0.92
10	noteworthy	0.02	0.98	10	interesting	0.01	0.73
11	imperative	0.02	0.70	11	safe	0.01	1.0
12	significant	0.01	0.86	12	tempting	0.01	1.0
13	probable	0.01	1.0	13	helpful	0.01	0.96
14	conceivable	0.01	0.99	14	appropriate	0.01	0.85
15	doubtful	0.01	0.79	15	enough	0.01	0.98

continuum. To illustrate, for the association of Adj-that and *clear*, we determine where the observed value of 0.16 falls within a theoretically possible range of −0.13 and 0.21. A value of 0.85 is observed—that is, from $\frac{0.16-(-0.13)}{0.21-(-0.13)}$. With this transformation, the ΔP scores are between 0 and 1, and the higher the score, the more strongly attracted an adjective is to a particular variant. Table 5 presents the top 15 adjectives of each variant ranked by observed ΔP scores, along with transformed ΔP scores. In the next section, we discuss the statistical analysis we conducted to test the contribution of this item-level predictor, in addition to other predictors, to responses from the elicitation task.

Statistical analysis of the elicitation data

We completed the following steps prior to statistical analysis. First, we dropped 42 partial, misspelled, or nonadjective responses (e.g., *typic*, *possile*, and *surprisingly*) from the elicitation data, which altogether constituted approximately 4% of the total observations (n = 1,057). Second, we removed past-participle responses (e.g., *argued* and *understood*), which accounted for 10.72% (n = 62 out of 578) and 11.89% (52/437) of the responses in L1 English and L1 Thai-L2 English participants, respectively. Finally, we lemmatized answers that were comparative or superlative adjectives (i.e., *good* for *better* and *best*) and retained 901 responses from 84 participants (L1 English: 516; L1 Thai-L2 English: 385). As can be seen in Table 6, summing over the two variants, *important* was the most common answer (n = 53), followed by *necessary* (29), *obvious* (26), *likely* (25), and *possible* (24).

To address our first research question, we ran a mixed-effects logistic regression model that related adjective–variant contingencies in the input and participants' engagement with academic English writing—among other things—to their responses in the elicitation task. We began by creating a binary outcome variable that indicated whether each of the 901 responses generated by the participants "fit" a particular variant. Take, for instance, four hypothetical answers—*clear*, *likely*,

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Table 6. The 15 most frequent adjectives of each variant in the experiment (L1 and L2 groups combined)

Adj-that			Adj-to		
Ranks	Adjectives	Frequencies	Ranks	Adjectives	Frequencies
1	important	21	1	important	32
2	obvious	15	2	necessary	17
3	likely	14	3	easy	13
4	clear	14	4	difficult	13
5	necessary	12	5	possible	12
6	possible	12	6	good	12
7	interesting	11	7	crucial	12
8	probable	11	8	obvious	11
9	good	9	9	likely	11
10	unlikely	9	10	essential	10
11	evident	9	11	logical	9
12	impossible	8	12	reasonable	9
13	logical	6	13	interesting	8
14	reasonable	6	14	hard	8
15	true	6	15	impossible	7

difficult, and *important*—to the target item *it is [blank] that*. Based on the observed $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores, *clear* and *likely* were attracted to Adj-that, whereas *difficult* and *important* were cued by Adj-to (see Table 5). In fact, *difficult* and *important* were repelled by Adj-that (i.e., $\Delta P_{(\text{Adj-that} \rightarrow \text{difficult})} = -0.10$, and $\Delta P_{(\text{Adj-that} \rightarrow \text{important})} = -0.12$). Therefore, given the Adj-that variant in the target item, we would code *clear* and *likely* each as a cued response (cued = 1) but not *difficult* and *important* (noncued = 0). With this outcome variable, we were thus able to capture one critical aspect of language users' constructional knowledge—that is, the knowledge of the adjective–variant associations in the two introductory-*it* variants.

Of the total 901 responses, 116 were unattested with either variant in our corpus analysis (e.g., *able*, *reliable*, and *excellent*) and thus were dropped, leaving 785 observations in the data set (L1 English: 458; L1 Thai-L2 English: 327).[4] Altogether, there were 183 types, 63 of which were hapax legomena. On average, the L1 English group supplied 5.34 ($SD = 2.86$, range: 1–13) responses for Adj-that and 5.19 ($SD = 2.11$, range: 1–10) answers for Adj-to, whereas the L2 group attempted 4.29 ($SD = 2.42$, range: 1–11) and 4.54 ($SD = 2.23$, range: 1–14) responses for Adj-that and Adj-to, respectively. A linear mixed-effects model with participants as random intercepts showed that L1, variant, and the interaction between the two predictors did not predict the number of responses (all $ps > .05$).

We conducted a mixed-effects logistic regression analysis on the 785 observations, predicting the binary outcome—cued responses—in the elicitation task, from the following set of predictors:

[4]One reviewer suggested that a ΔP score of -1 be given to unattested responses for them to be retained in the analysis. Such a score would also indicate that adjectives were repelled by the variants. However, one issue with keeping these responses in the analysis was how to estimate frequencies of these items. Though several options are available, most—if not all—of them have inherent limitations (see Brysbaert & Diependaele, 2013). With this in mind, we decided against including the 116 unattested items in our regression analysis.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

(1) **Introductory-*it* variants**: a categorical variable for the variant of a target item for which participants supplied their answers, either Adj-that (e.g., *It is [blank] that*) or Adj-to (e.g., *It is [blank] to*). We chose this predictor primarily because the two variants differed in the number of attested types (see Table 1; see also Larsson, 2016) and because usage-based studies have shown that high type frequency is critical to constructional generalization (Azazil, 2020).
(2) **$\Delta P_{\text{(variant} \rightarrow \text{adjective)}}$ scores**: a continuous variable for the adjective–variant associations. Given the design of the elicitation task, we entered as a predictor transformed $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores, which ranged from 0.01 to 1 in the data set. We used transformed ΔP rather than observed ΔP to control for the effect of frequencies (see Gries, 2022). We used ΔP scores of adjectives in the variant to which they were attracted regardless of the variant of a target item. For instance, a response *clear* to the items *it is [blank] that* and *it is [blank] to* was given the same ΔP score of 0.84 for the word's attraction to Adj-that.
(3) **Adjective frequencies**: a continuous variable for the co-occurrence frequencies of adjectives with the variant to which they were attracted, as obtained from the academic section of COCA. Take, for example, two answers—*clear* and *difficult*—to the item *it is [blank] that*. We assigned the frequencies of 2,826 to *clear* for its co-occurrences with Adj-that and 4,615 to *important* for its co-occurrences with Adj-to.[5]
(4) **L1s**: a categorical variable for participants' L1, either English or Thai.
(5) **Degrees**: a categorical variable for a graduate degree participants were pursuing, either a master's or PhD.
(6) **AWE-Q scores**: a continuous variable for participants' average AWE-Q scores. We used this predictor as an indirect measure of participants' active use of the two introductory-*it* variants, among many other constructions, in their academic studies.

In addition to the above predictors, we included two control variables as follows:

(7) **Linking verbs**: a categorical variable for the verb embedded inside a target item, either *is* (e.g., *It is [blank] that*) or *seems* (e.g., *It seems [blank] to*).
(8) **Trials**: a continuous variable for the order in which responses were supplied in each target item by the participants.

We applied the following transformation to the above predictors. Continuous variables were grand-mean centered and standardized, and adjective frequencies were $\log_2$ transformed prior to centering and scaling, but transformed ΔP scores were not because doing so did not affect the shape of the distribution. Categorical predictors were sum-coded.

We took a bottom-up approach in building our mixed-effects models, as recommended by Hox et al. (2017). First, we added into the model item-level (i.e., variants, frequencies, ΔP scores, linking verbs, and trials) and participant-level predictors (i.e., L1s, degrees, and AWE-Q scores) as fixed effects and assessed whether they were correlated.

[5]Assigning different transformed ΔP scores and frequencies to each answer depending on which variant was present in the target item (e.g., ΔP scores of 0.844 and 0.001 for *clear* when the adjective was supplied for *it is [blank] that* and *it is [blank] to*, respectively) produced extremely large and statistically significant estimates for all predictors, an indication that (quasi-)complete separation occurred. Such an issue was likely caused by how the outcome variable in the present study was created (that is, based on ΔP scores).

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

We found that all variance inflation factors (VIFs) were close to 1. We then included intralevel interaction terms (e.g., the interaction between variants and ΔP scores) and checked the VIFs. Along with the fixed-effect structure, we placed random intercepts on participants and items. For item-specific random intercepts, we considered only adjectives with frequency ≥ 2 as individual factor values and coded all 63 hapax legomena as "other." Thus, instead of 183, there were 121 factor values for items.

Second, we assessed whether random slopes were justified. Initially, our model consisted of by-participant random slopes for variants, adjective frequencies, and ΔP scores and by-item random slopes for L1s. As singular fit was reported, we ran a principal components analysis on the random-effect structure (see Gries, 2021) and subsequently simplified the model by dropping random slopes for—in this exact order—L1s, ΔP scores, and adjective frequencies. Likelihood-ratio tests (LRTs) confirmed the removal (all ps > .75). In the end, the random-effect structure consisted of random intercepts for participants and items as well as by-participant random slopes for variants.

Last, we incorporated cross-level interaction terms into the model and tested whether any of the fixed-effect terms could be removed (see Gries, 2021).[6] The final model from which we drew inferences was significantly different from the null model —$\chi^2(9) = 179.81$, $p < .001$; $AIC_{final\ model} = 825.52$; and $AIC_{null\ model} = 987.33$—with all VIFs near 1. The value for $R^2_{marginal}$ was .36, and $R^2_{conditional}$ was .52. The model's C score was 0.89, which was above the 0.80 threshold commonly used in linguistics.

To address the second research question vis-à-vis L2 users' English proficiency and their performance on the elicitation task, we performed a mixed-effects logistic regression analysis with only data from the L2 English participants. Because four participants did not report their TOEFL scores, they were removed from further analysis. In total, there were 301 data points. Initially, we included the same set of predictors used to address the first research question, except L1s and related interactions. Subsequently, we simplified the random-effect structure of the model, retaining only the by-participant random intercepts and random slopes for variants, and tested which predictors could be dropped with LRTs (all ps > .25). At this stage, as the AWE-Q scores were nonsignificant, the predictor was dropped. The final model from which we drew inference was significantly different from the null model—$\chi^2(5) = 86.87$, $p < .001$; $AIC_{final\ model} = 299.16$; $AIC_{null\ model} = 376.03$. All predictors had VIFs close to 1. The $R^2_{marginal}$ was .42, and $R^2_{conditional}$ was .60. The model's C score was 0.91.

Results

To address the research questions, we extracted the predicted probability of cued responses from the models. We use the term proportion instead of probability to facilitate readers in their comprehension of results.

RQ 1: The effect of input and output on elicitation task performance

Before discussing the main results, we present a crosstab of cued responses by variants and L1s. As can be seen in Table 7, the participants were largely able to generate cued responses for each introductory-*it* variant. The only exception, however, is the L2 users'

[6]At this stage, as the linear effect of AWE-Q was nonsignificant, we explored the curvature of AWE-Q scores with a second-degree polynomial, per one reviewer's comment. However, the LRT indicated that a model with the curved effect of AWE-Q scores was not significantly better, $\chi^2(1) = 0.38$, $p = .54$.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Table 7. Crosstab of cued responses by variants and L1s

	Adj-that		Adj-to	
L1	Cued	Noncued	Cued	Noncued
English	136	99	156	67
Thai	56	94	133	44

performance on Adj-that; only 37.3% of the responses were considered attracted to the variant. On average, Adj-to elicited a higher proportion of cued responses (> 70%), whereas only 46.8% of the responses were cued by Adj-that.

Table 8 presents cued and noncued responses of each variant across the two groups of participants. Zooming in on the cued responses, the 10 most common answers of the L1 English group were *important* (n = 16); *necessary* (13); *clear*, *likely*, and *obvious* (10 each); *evident* and *probable* (8 each); and *good*, *reasonable*, and *unlikely* (6 each). Six of these words were cued by Adj-that. All answers had transformed ΔP scores above 0.80, and only four (*important* and *necessary* for Adj-to and *clear* and *likely* for Adj-that) occurred more than 1,000 times with their respective variants in COCA. The L2 English participants produced *important* most often (16), followed by *crucial* and *possible* (8 each); *difficult* and *easy* (7 each); *good* and *interesting* (6 each); and *acceptable*, *impossible*, and *obvious* (5 each). Of the 10 answers, only *obvious* and *possible* were attracted to Adj-that; the rest were cued by Adj-to. Half of the answers occurred more than 1,000 times with their respective variants in COCA, and seven (i.e., *acceptable*, *difficult*, *easy*, *good*, *important*, *impossible*, and *obvious*) had $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores above 0.80.

Regarding noncued responses, the L1 English group generated *important* most often (12), followed by *likely*, *necessary*, and *obvious* (8 each); *logical* (6); *essential*, *good*, *interesting*, and *possible* (5 each); and *unlikely* (4). For the L2 English group, the 10 most common noncued responses were *important* (9); *impossible* and *possible* (7 each); *interesting* (6); *essential* (5); *good* and *necessary* (4 each); and *correct*, *crucial*, and *reasonable* (3 each).

As shown in Table 9, the mixed-effects logistic regression model revealed a significant and positive effect of $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores. The significant effect was driven by the participants' generation of adjectives that were distinctive responses of the variants. Examples include *difficult*, *easy*, *important*, and *necessary* for Adj-to and *clear*

Table 8. The 10 most frequent responses of each variant separated by their cued/noncued status

	Adj-that					Adj-to			
Rank	Cued		Noncued		Rank	Cued		Noncued	
1	obvious	15	important	21	1	important	32	possible	12
2	clear	14	necessary	12	2	necessary	17	likely	11
3	likely	14	interesting	11	3	difficult	13	obvious	11
4	possible	12	good	9	4	easy	13	essential	10
5	probable	11	impossible	8	5	crucial	12	logical	9
6	evident	9	reasonable	6	6	good	12	unlikely	5
7	unlikely	9	bad	4	7	reasonable	9	vital	4
8	logical	6	correct	4	8	hard	8	clear	3
9	true	6	crucial	4	9	interesting	8	critical	3
10	imperative	5	false	4	10	impossible	7	imperative	3

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Table 9. Summary of components in the mixed-effects logistic regression model predicting cued responses

Parameters	Estimates	*SE*	95% CI	*z*	*p*
Intercept	0.64	0.13	[0.39, 0.89]	4.98	<.001***
Trial	−0.23	0.11	[−0.43, −0.02]	−2.16	.031*
Variant [Adj-that]	−0.71	0.16	[−1.02, −0.39]	−4.39	<.001***
ΔP score	0.55	0.14	[0.27, 0.83]	3.85	<.001***
Adjective frequency	−0.05	0.13	[−0.30, 0.20]	−0.40	.689
L1 [English]	0.13	0.11	[−0.09, 0.36]	1.18	.238
AWE-Q	0.05	0.11	[−0.16, 0.25]	0.44	.659
Variant [Adj-that] × L1 [English]	0.44	0.15	[0.14, 0.74]	2.89	.003**
Variant [Adj-that] × Frequency	0.37	0.11	[0.15, 0.59]	3.32	<.001***
Variant [Adj-that] × ΔP score	−1.24	0.14	[−1.52, −0.96]	−8.63	<.001***

Note. The 95% CIs were approximated using the Wald method; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

and *obvious* for Adj-that. Adjective frequencies, however, did not influence the participants' generation of cued responses. Contrary to our prediction, academic writing engagement, which we used to indirectly assess participants' active use of the introductory-*it* construction, was nonsignificant. Participants who were more actively engaged with academic writing did not produce more cued responses.

In addition to the two target predictors, the main effect of variants was significant, with Adj-that eliciting considerably fewer cued responses than Adj-to. Consistent with the descriptive statistics in Table 7, the model estimated that 50.1% of the answers in Adj-that were cued responses of the variant, as opposed to 73.0% in Adj-to. Unlike variants, participants' L1s were not a significant predictor, though according to the model the L1 English group generated more cued responses (64.5%) than the L1 Thai-L2 English group (58.0%). One control variable, trials, was significant: cued responses were generated earlier, and the probability that an answer would become attracted to a variant decreased for each additional response (see Ellis et al., 2014, p. 75).

There were three significant interactions in the model. For the first one, although the main effect of L1s was nonsignificant, the interaction between L1s and variants was. In Figure 1, we can see a sharp drop in the estimated mean proportion of cued responses in the L2 English group, from Adj-to to Adj-that (75.0% vs. 38.9%). In contrast, such a drop was much less pronounced in the L1 English group (69.8% vs. 59.7%). Figure 1 graphically highlights this interaction; the L1 Thai-L2 participants supplied more cued responses given Adj-to than their L1 English counterparts, but their performance on Adj-that was much less robust. Fewer than 40% of the responses supplied by the L2 group were distinctive adjectives of Adj-that.

Regarding the other two interactions, we found that participants' generation of distinctive adjectives varied as a function of variants and input statistics (i.e., frequencies and Δ*P* scores). The left panel of Figure 2 presents the interaction between the variants and adjective frequencies, and the right panel illustrates the interaction between the variants and transformed Δ*P* scores. Note that in Figure 2 the two predictors are in their original scale ($\log_2$ adjective frequencies and transformed Δ*P* scores). To explain the findings, it is worth inspecting a list of cued and noncued responses of each variant in Table 8. When prompted with the Adj-that variant, participants supplied extremely frequent lexemes of the construction, with the response *logical* being the exception. This is why, in the left panel of Figure 2, the

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

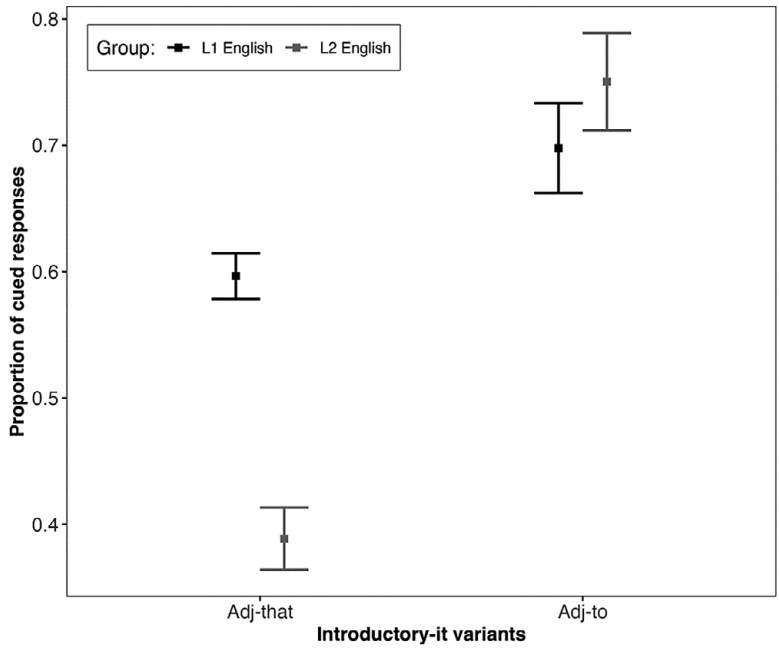


Figure 1. Estimated mean proportion of cued responses by L1 and variants. The values plotted in this graph were estimated from the mixed-effects logistic regression model. The error bars represent bootstrapped 95% confidence intervals from 100 simulations.

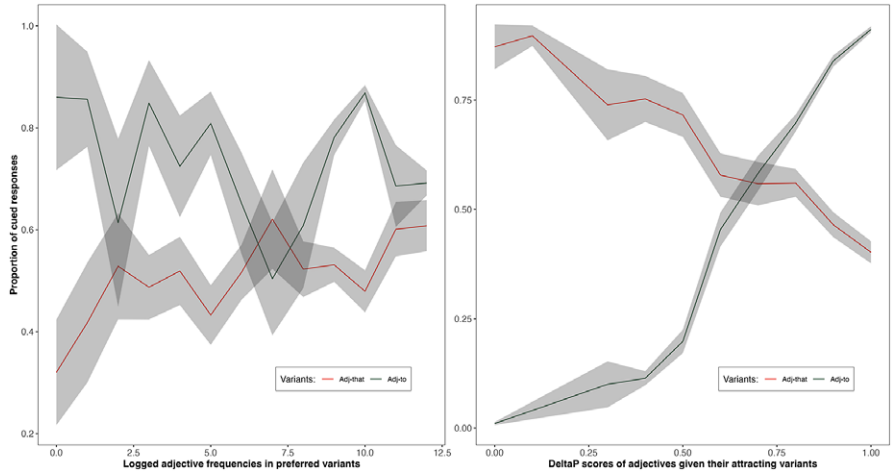


Figure 2. Estimated mean proportion of cued responses as a function of variants and input statistics. The values plotted in this graph were estimated from the mixed-effects logistic regression model. Error bands are bootstrapped 95% confidence intervals from 100 simulations.

mean proportion of cued responses of Adj-that increased as adjective frequencies were higher. The model estimated that, at logged frequency of zero (raw frequency = 1), the mean proportion of cued responses of Adj-that was at 32.0% (95% CI = [21.7%, 42.4%]). But at logged frequency of 11 (raw frequency = 2,048), the mean

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

proportion of cued responses went up to 60.0% (95% CI = [54.8%, 65.4%]). For the Adj-to variant, we found that participants often generated responses such as *possible*, *essential*, and *logical*, which despite being cued by Adj-that appeared slightly more frequently with Adj-to in COCA (for frequencies of *possible* and *essential*, see Table 2). In the context of the present study, these answers were treated as noncued responses of the variant; thus, the lower mean proportion of cued responses in Adj-to as frequencies increased.

The right panel of Figure 2 presents the mean proportion of cued responses as a function of introductory-*it* variants and transformed ΔP scores. This interaction was driven mainly by the observation that as ΔP scores increased, the proportion of distinctive lexemes of Adj-that decreased from 87.2% (95% CI = [82.1%, 92.3%]) to 40.2% (95% CI = [37.7%, 42.7%]). This is because, as shown in Table 8, participants completed the item *it is [blank] that* with such adjectives as *important*, *necessary*, *interesting*, and *good*, which were strongly cued by Adj-to (transformed $\Delta P_{(\text{Adj-to} \rightarrow \text{adjective})}$ scores > 0.7) and repelled by Adj-that. Though these responses had high ΔP scores for their attraction to Adj-to, they were considered noncued responses of the Adj-that variant.

RQ 2: The effect of L2 users' English proficiency on elicitation task performance

The model revealed that L2 proficiency had a positive, yet insignificant effect on cued responses, which contradicted our prediction, $\beta = 0.34$, $SE = 0.18$, 95% CI = [−0.02, 0.70], $p = .061$. In other words, L2 users' ability to generate responses that fit a given construction did not depend on their L2 proficiency. Like the first question, the model identified a significant effect of variants, $\beta = -1.17$, $SE = 0.29$, 95% CI = [−1.74, −0.61], $p < .001$, and ΔP scores, $\beta = 0.48$, $SE = 0.22$, 95% CI = [0.05, 0.91], $p = .030$. As is evident in Table 7, the L1 Thai-L2 English group generated fewer cued responses for Adj-that, thus the negative effect of variants. The effect of ΔP scores was driven in large part by responses the L2 group generated for Adj-to, as discussed above. Note that though frequencies had a positive effect on cued responses, their effect was not significant, $\beta = 0.08$, $SE = 0.17$, 95% CI = [−0.26, 0.42], $p = .653$. Last, a significant interaction between variants and ΔP scores was observed, $\beta = -1.33$, $SE = 0.22$, 95% CI = [−1.76, −0.90], $p < .001$. Like the previous analysis, the mean proportion of cued responses of Adj-that decreased because the L1 Thai-L2 English participants completed the target item *it is [blank] that* with adjectives that were highly distinctive to Adj-to.

Discussion

The results from the elicitation task demonstrated that both L1 and L2 users of English were able to supply adjectives that were attracted to the introductory-*it* variants. As predicted, we observed the effect of association strength of adjectives and variants on the elicitation task performance, particularly for the L2 groups. However, we were not able to confirm the other two predictions regarding the effect of academic writing engagement and L2 proficiency. As reported in the first analysis, participants who performed academic writing tasks more frequently, and by extension used the two introductory-*it* variants more often, did not produce more cued responses. With respect to L2 users, we predicted that their English proficiency would affect their task performance, but this prediction was not supported by the data.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Given the available evidence for the effect of lexeme–construction associations, particularly those between verbs and VACs, in both L1 and L2 speakers (Azazil, 2020; Ellis & Ferreira-Junior, 2009; Sung & Kim, 2022), the effect of adjective–variant contingencies on language users' generation of cued responses was expected. Both L1 and L2 users supplied adjectives that are distinctively attracted to the two variants. This was particularly true with the L1 users, whose most common cued responses were among the top 10 lexemes of the two variants (e.g., *important*, *necessary* for Adj-to; *clear, likely, obvious* for Adj-that). Larsson (2016) showed that these adjectives were prototypical exemplars of their respective variants and, because of this, often accessed first in language production (Ellis et al., 2014, Experiment 1). The use of transformed ΔP scores, which captured adjective–variant contingencies without the effect of frequency (see Gries, 2022), in our mixed-effects logistic regression models helps rule out the possibility that participants' generation of cued responses was driven by the co-occurrence frequencies between adjectives and their preferred variants. This claim is further aided in that adjective frequency was not a significant predictor in both models. Nonetheless, we note that for L2 users the top two answers, *important* and *possible*, were the two most frequent adjectives in our COCA data (summing across the two variants: *important* = 5,468 and *possible* = 4,282). In fact, *important* and *possible* are in the top 10 most frequent predicative adjectives in academic discourse (Biber et al., 1999). Moreover, the L2 group's most common cued responses generally had lower $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores than those produced by the L1 participants.

As shown in Figure 1, we can see that the participants overall generated far fewer cued responses for Adj-that than for Adj-to. The L2 group showed a greater drop than did their L1 peers, being able to supply adjectives that were attracted to Adj-that only 40% of the time. That the Adj-to variant elicited a higher proportion of cued responses may be due in part to higher productivity of the variant. There were almost twice as many adjective types in Adj-to as in Adj-that (495 vs. 262; see Table 1). Usage-based studies have shown that high type frequency affects constructional generalization and language production (Ambridge et al., 2015; Gries & Ellis, 2015; Sung & Kim, 2022), and we see a similar effect at play in the present study. We also found that the L2 group generated more cued responses for Adj-to than did the L1 group (that is, 75.0% vs. 69.8%) and that almost all of their top 10 answers were attracted to Adj-to. This may be the case because—with higher type frequencies (as well as token frequencies) in Adj-to—L2 users were able to acquire and form a schematic representation for the construction much better than they did with the Adj-that variant (see Azazil, 2020, for similar results). As one reviewer suggested, given findings from corpus-based studies that a VAC may contain verbs from heterogeneous semantic classes (e.g., Perek, 2014), another reason for the higher proportion of elicited cued responses for the Adj-to variant may be the semantic openness of the adjectives in this variant. This is therefore a possibility that could be explored further in future research.

As discussed in the Method section, we classified participants' answers into cued or noncued responses based on $\Delta P_{(\text{variant} \rightarrow \text{adjective})}$ scores, with the goal of capturing L1 and L2 users' knowledge of lexeme–construction contingencies. Thus, noncued responses are by and large not malformed instances of the construction but rather adjectives that were repelled by a particular variant. For the L2 group, we found that many of their noncued responses were synonymous or semantically related to cued responses, and these noncued responses were equally frequent. For example, for the item *it is [blank] that …*, L2 users supplied *possible*, a cued response, eight times and *impossible*, which is distinctively attracted to Adj-to, seven times. Moreover, for the item *it is [blank] to …*, they supplied *crucial*, a cued response, eight times and *essential*, a lexeme of Adj-that, five times. It is

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

therefore possible that the L2 users' generation of adjectives was driven in part by how closely adjectives are related. This finding is also in line with previous research findings that L2 users, even those at an advanced level, may overgeneralize the use of synonymous or semantically related adjectives (e.g., Sonbul, 2015).

The current study constituted the first attempt to operationalize language output and investigate usage-based researchers' proposal that language users' own constructional use–independent of input statistics to which they are exposed—strengthens their constructional knowledge (e.g., Ellis, 2022; Schmid, 2015). It was hypothesized that L1 and L2 users' greater use of the Adj-to and Adj-that variants, as assessed with an academic writing questionnaire, should further attune these users to adjective–variant associations. However, we did not find evidence in the present work that language output influenced elicitation task performance. Our results indicated that L1 and L2 users' academic writing engagement did not predict cued responses. There are two possible explanations for this null effect. First and most importantly, the scales of the questionnaire—from 0 to 5—encompass a narrow range, and a substantial number of participants in each group rated their engagement as high (i.e., 12 out of 44 L1 English participants and 12 out of 40 L2 English participants with rating of 4 or higher). This was to be expected, given the research-intensive nature of graduate school. However, because of these negatively skewed rating scores over a narrow range of possible values, the effect was estimated imprecisely, as evidenced in the standard error (0.11) that was twice as large as the mean (0.05). Second, what the AWE-Q captures may not be sensitive to what was assessed in the elicitation task. We employed the questionnaire, which assesses the level of engagement with academic writing in English, as a proxy for participants' active use of the introductory-*it* construction. Because of its focus on various writing tasks rather than specific instances of language use, the questionnaire can only capture linguistic experience holistically. In contrast, the elicitation task tapped language users' ability to generate adjectives that were associated with the Adj-that and Adj-to variants. It is highly possible that this form of statistical association may have been too fine-grained to be precisely detected by such a questionnaire as the AWE-Q. We must stress that the null effect of academic writing engagement does not negate the role of language production but rather raises an important question about how to assess output in a way that is both consistent with predictions from usage-based models and useful for L2 research.

In the present study, L2 users' generation of cued responses did not depend on their L2 proficiency scores. One explanation for the null effect of L2 proficiency is that the TOEFL scores no longer reflected L2 users' current language abilities. Recall that we recruited L1 Thai-L2 English participants who were graduate students in U.S. universities, with half of them being doctoral students. The L2 group reported having spent 3.3 years (SD = 2.2 years) in an English-speaking country by the time of data collection (4.3 years [SD = 2.3] for PhD students). It is reasonable to assume, given constant exposure to academic English for such an extended period, that the L2 participants had made significant gains in their English beyond what was captured by the L2 proficiency test.

Limitations and conclusion

As with any study, some limitations and outstanding questions remain. First, although previous corpus studies on VACs reported that more proficient L2 users are more likely to learn and use less frequent verb-VAC combinations, indicating their knowledge of

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

verbs beyond prototypical verbs (Kyle et al., 2021), the current study did not probe whether more proficient L2 users are more likely to use adjectives that are less strongly attracted to the two variants. Second, whether L2 users' knowledge of adjective–variant contingencies is mediated by other types of associations that are also present in each variant remains to be seen. From usage-based perspectives, it is likely that L2 users' ability to form associations between individual adjectives and their preferred variants is facilitated by additional cues, such as the pairings between adjectives and infinitives for the Adj-to variant (Larsson & Kaatari, 2019). Future work can address this issue, for instance, by including verbs—or other lexical items—into a target frame (e.g., *It is [blank] to note …*). This design can increase the naturalness of the test stimuli while still being able to address lexeme–construction contingencies (see Dąbrowska, 2009). By investigating how adjective–variant associations and other types of contingencies interact as L2 users traverse their developmental trajectories, L2 researchers will be able to have a fuller understanding of learners' constructional knowledge.

Despite these limitations, the current study has demonstrated that lexeme–construction contingencies shape L2 users' knowledge of introductory-*it* constructions in much the same way that they influence knowledge of VACs. The effect of associations varies by constructions and is likely to be mediated by constructional frequencies (i.e., type frequencies). Although L2 users' constructional use—as operationalized in the present study—does not affect elicitation task performance, the null effect should serve to heighten the need for more robust assessment tools that can capture learners' active use of a target construction. Only when both sides of usage, language input and output, are taken into consideration will we be able to paint a more complete picture of L2 users' constructional knowledge.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263123000517.

Data availability statement. The experiment in this article earned an Open Data badge for transparent practices. The data and the R script can be found at https://github.com/suesakol/introductory-it.

Acknowledgments. We thank Lourdes Ortega and Brandon Tullock for their feedback at various stages of this project. Part of this work was presented at the American Association for Corpus Linguistics (2018) in Atlanta, GA.

Competing interest. The authors declare none.

References

Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, *40*, 509–559. https://doi.org/10.1177/0142723719869731

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*, 239–273. https://doi.org/10.1017/S030500091400049X

Azazil, L. (2020). Frequency effects in the L2 acquisition of the catenative verb construction: Evidence from experimental and corpus data. *Cognitive Linguistics*, *31*, 417–451. https://doi.org/10.1515/cog-2018-0139

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.

Brysbaert, M., & Diependaele, K. (2013). Dealing with zero word frequencies: A review of the existing rules of thumb and a suggestion for an evidence-based choice. *Behavior Research Methods*, *45*, 422–430.

Chainarongdejagul, S. (2018). *A study on strategies used in translating it-cleft sentences in Pirunrat's translation of Agatha Christie's* The Murder of Roger Ackroyd [*Unpublished master's thesis*]. Chulalongkorn University.

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Colleman, T., & Bernolet, S. (2012). Alternation biases in corpora vs. picture description experiments: DO-biased and PD-biased verbs in the Dutch dative alternation. In D. Divjak & S. Th. Gries (Eds.), *Frequency effects in language representation* (pp. 87–125). Mouton de Gruyter.

Conklin, K., & Thul, R. (2023). Word and multiword processing. In A. Godfroid & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 203–215). Routledge.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213–238. https://doi.org/10.2307/3587951

Crossley, S., Kyle, K., & Salsbury, T. (2016). A usage-based investigation of L2 lexical acquisition: The role of input and output. *Modern Language Journal*, *100*, 702–751. https://doi.org/10.1111/modl.12344

Dąbrowska, E. (2009). Words as constructions. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 201–233). John Benjamins.

Dąbrowska, E. (2015). Individual differences in grammatical knowledge. In E. Dąbrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics* (pp. 650–668). Mouton de Gruyter.

Davies, M. (2019). *The Corpus of Contemporary American English: One billion words, 1990–present.* https://www.english-corpora.org/coca/

Ellis, N. C. (2019). Essentials of a theory of language cognition. *Modern Language Journal*, *103*, 39–60. https://doi.org/10.1111/modl.12532

Ellis, N. C. (2022). Fuzzy representations. *Bilingualism: Language and Cognition*, *25*, 210–211. https://doi.org/10.1017/S1366728921000638

Ellis, N. C., & Ferreira-Junior, F. (2009). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, *7*, 187–220. https://doi.org/10.1075/arcl.7.08ell

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics*, *25*, 55–98. https://doi.org/10.1515/cog-2013-0031

Eskildsen, S. W. (2012). L2 negation constructions at work. *Language Learning*, *62*, 335–372. https://doi.org/10.1111/j.1467-9922.2012.00698.x

Francis, G., Hunston, S., & Manning, E. (1998). *Grammar patterns 2: Nouns and adjectives*. Harper Collins.

Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next …. *International Journal of Corpus Linguistics*, *18*, 137–165. https://doi.org/10.1075/ijcl.18.1.09gri

Gries, S. Th. (2021). (Generalized linear) mixed-effects modeling: A learner corpus example. *Language Learning*, *71*, 757–798. https://doi.org/10.1111/lang.12448

Gries, S. Th. (2022). What do (some of) our association measures measure (most)? Associations? *Journal of Second Language Studies*, *5*, 1–33. https://doi.org/10.1075/jsls.21028.gri

Gries, S. Th., & Ellis, N.C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, *65*, 228–255. https://doi.org/10.1111/lang.12119

Gries, S. Th., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics*, *16*, 635–676. https://doi.org/10.1515/cogl.2005.16.4.635

Gries, S. Th., & Wulff, S. (2009). Psycholinguistic and corpus-linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics*, *7*, 164–187. https://doi.org/10.1075/arcl.7.07gri

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, *4*, 257–277. https://doi.org/10.1016/j.jeap.2005.03.002

Hardie, A. (2012). CQPweb–combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*, 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Herriman, J. (2000). Extraposition in English: A study of the interaction between the matrix predicate and the type of extraposed clause. *English Studies*, *81*, 582–599. https://doi.org/10.1076/enst.81.6.582.9180

Hewings, M., & Hewings, A. (2002). "It is interesting to note that …": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, *21*, 367–383. https://doi.org/10.1016/S0889-4906(01)00016-3

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Hyland, K., & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, *45*, 40–51. https://doi.org/10.1016/j.esp.2016.09.001

Hyland, K., & Jiang, F. (2018). 'We believe that…': Changes in an academic stance marker. *Australian Journal of Linguistics*, *38*, 139–161. https://doi.org/10.1080/07268602.2018.1400498

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press

Kaltenböck, G. (2005). *It*-extraposition in English: A functional view. *International Journal of Corpus Linguistics*, *10*, 119–159. https://doi.org/10.1075/ijcl.10.2.02kal

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (pp. vii–xxviii). CSLI Publications.

Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, *43*, 781–812. https://doi.org/10.1017/S0272263120000546

Larsson, T. (2016). The introductory *it* pattern: Variability explored in learner and expert writing. *Journal of English for Academic Purposes*, *22*, 64–79. https://doi.org/10.1016/j.jeap.2016.01.007

Larsson, T. (2017). A functional classification of the introductory *it* pattern: Investigating academic writing by non-native-speaker and native-speaker students. *English for Specific Purposes*, *48*, 57–70. https://doi.org/10.1016/j.esp.2017.06.001

Larsson, T. (2019). Grammatical stance marking across registers: Revisiting the formal-informal dichotomy. *Register Studies*, *1*, 243–268. https://doi.org/10.1075/rs.18009.lar

Larsson, T., & Kaatari, H. (2019). Extraposition in learner and expert writing: Exploring (in)formality and the impact of register. *International Journal of Learner Corpus Research*, *5*, 33–62. https://doi.org/10.1075/ijlcr.17014.lar

Peacock, M. (2011). A comparative study of introductory it in research articles across eight disciplines. *International Journal of Corpus Linguistics*, *16*, 72–100. https://doi.org/10.1075/ijcl.16.1.04pea

Perek, F. (2014). Rethinking constructional polysemy. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 61–85). John Benjamins.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, *7*, 140–162. https://doi.org/10.1075/arcl.7.06rom

Römer, U., & Berger, C. M. (2019). Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition*, *41*, 1089–1110. https://doi.org/10.1017/S0272263119000202

Römer, U., Skalicky, S. C., & Ellis, N. C. (2020). Verb-argument constructions in advanced L2 English learner production: Insights from corpora and verbal fluency tasks. *Corpus Linguistics and Linguistic Theory*, *16*, 303–331. https://doi.org/10.1515/cllt-2016-0055

Schmid, H.-J. (2015). A blueprint of the entrenchment-and-conventionalization model. *Yearbook of the German Cognitive Linguistics Association*, *3*, 3–25. https://doi.org/10.1515/gcla-2015-0002

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, *18*, 419–437. https://doi.org/10.1017/S1366728914000674

Suethanapornkul, S., & McKay, T. H. (2018). *The academic writing experience questionnaire (AWE-Q): A development and (some) validation study.* Unpublished manuscript.

Sung, M.-C., & Kim, H. (2022). Effects of verb-construction association on second language constructional generalizations in production and comprehension. *Second Language Research*, *38*, 233–257. https://doi.org/10.1177/0267658320932625

Supasiraprapa, S. (2019). Frequency effects on first and second language compositional phrase comprehension and production. *Applied Psycholinguistics*, *40*, 987–1017. https://doi.org/10.1017/S0142716419000109

Cite this article: Suethanapornkul, S., & Supasiraprapa, S. (2024). Usage events and constructional knowledge: A study of two variants of the introductory-*it* construction. *Studies in Second Language Acquisition,* 46: 355–377. https://doi.org/10.1017/S0272263123000517

https://doi.org/10.1017/S0272263123000517 Published online by Cambridge University Press