<LRH>Distributional Factors in the Use of L2 Morphemes

<LRH>Distributional Factors in the Use of L2 Morphemes

<ArtType>EMPIRICAL STUDY

<AT>Effects of Availability, Contingency, and Formulaicity on the Accuracy of English Grammatical Morphemes in Second Language Writing

<AU>Akira Murakami[a] and Nick C. Ellis[b]

<AF>[a]University of Birmingham and [b]University of Michigan

<AN>

We thank Masaki Eguchi for drawing our attention to Dunn (2018).

Correspondence concerning this article should be addressed to Akira Murakami, University of Birmingham, Edgbaston Campus, Frankland Building, Department of English Language and Linguistics, Birmingham, B15 2TT, United Kingdom. Email: a.murakami@bham.ac.uk

The handling editor for this article was Scott Crossley.

<ABS>

We investigated whether the accuracy of grammatical morphemes in second language (L2) learners' writing is associated with usage-based distributional factors. Specifically, we examined whether the accuracy of L2 English inflectional morphemes is associated with the availability (i.e., token frequency) and contingency (i.e., token frequency relative to other forms with the same lemma) of the inflected word form as well as the formulaicity of the context in which it occurs (i.e., predictability of the form given the surrounding words). Data drawn from a large-scale learner corpus indicated that contingency is a robust predictor of morpheme accuracy, thereby supporting the usage-based view that language learners are sensitive to distributional properties in their input. Furthermore, the relationship of contingency with accuracy does not necessarily lessen when learners' proficiency rises. Contrary to previous research investigating online processing, we did not identify in our study availability and formulaicity as predictors of accuracy of morpheme production in writing.

<KWG>**Keywords** usage-based theories; contingency; grammatical morpheme; learner corpus

<A>Introduction

<TXT>

The acquisition of grammatical morphology by second language (L2) learners is slow, gradual, and sometimes incomplete. Second language acquisition (SLA) research has shown that:

1.Some morphemes are acquired earlier than others as, for example, the classic studies of morpheme acquisition order demonstrated (e.g., Dulay & Burt, 1973, 1974).

2.Within each morpheme, some forms are acquired earlier than others depending on the lemmas the morpheme is attached to, as, for instance, investigations of the aspect hypothesis demonstrated (e.g., Andersen & Shirai, 1994).

3.Acquisition takes place incrementally over several years of usage, if at all. Jia and Fuse (2007) demonstrated that, for children in an English as a second language (L2) environment, the acquisition of a morpheme such as the third-person singular present-tense *-s* can take five years or more to go from zero to 80% provision in obligatory contexts, in spoken production. At the extreme, many adult L2 learners will never acquire total control of L2 morphology, even after tens of years of English immersion (e.g., in spoken production, Johnson & Newport, 1989; Schmidt, 1984).

Grammatical morphemes are ubiquitous across first language (L1) usage: Each day provides tens to thousands of receptive experiences of functional morphemes, and tens to thousands of contexts requiring their productive use. Yet L2 provision is variable.

A wide range of factors have been discussed in relation to this variability in L2 morpheme acquisition: see, for example, Goldschneider and DeKeyser (2001) on perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency; Murakami and Alexopoulou (2016) for a large-scale learner corpus investigation of L1 to L2 transfer effects; N. C. Ellis and Sagarra (2011) on blocking and learned attention; Cintrón-Valentín and Ellis (2016) on salience, attention, modality, and form-focused instruction; Wulff and Ellis (2018) on implicit and explicit learning. Indeed, in 2015 a whole special issue was dedicated to explaining morpheme-order related phenomena from a wide range of different theoretical perspectives; see, for example, Hulstijn (2015), Hulstijn, Ellis, & Eskildsen (2015), and O'Grady (2015).

In our study, we focused on patterns of language experience as they might affect within-morpheme development of inflected forms. We took the perspective of usage-based theories that hold that distributional properties in the input affect the ease of processing, order of acquisition, and accuracy of use of linguistic constructions, that the system is learned incrementally, and that regularities/generalization/productivity emerge from the combined experiences of usage (e.g., Beckner et al., 2009; Bod, Hay, & Jannedy, 2003; Bybee & Hopper, 2001; N. C. Ellis, 2002; Goldberg, 2006; MacWhinney & O'Grady, 2015; Robinson & Ellis, 2008).

The following three interrelated distributional properties have been implicated in usage-based learning: availability, contingency, and formulaicity.

<A>Background Literature

<B>Availability

<TXT>

Availability is concerned with how often learners experience a particular form in their input. If the word *wanted* is more frequent than the word *graduated*, one might expect that learners typically would use past tense *-ed* in *wanted* more accurately than in *graduated*. Availability has been widely demonstrated to impact processing, acquisition, and use in both L1 (e.g., Ambridge, Kidd, Rowland, & Theakston, 2015) and L2 (e.g., N. C. Ellis, 2002; N. C. Ellis, Römer, & O'Donnell, 2016). In L1 acquisition, high token frequency surface forms are produced earlier and more accurately in those forms compared to production in other forms and compared to production of other words that are inflected in lower token frequency forms (Aguado-Orea, 2004; Braine et al., 1990; Finley, 2018; Marchman, 1997). The fact that the frequency of word lemmas plays a lesser role in the accurate retrieval of inflected word-forms compared to the token frequency of an inflected word-form itself has been key evidence in the development of emergentist theories of language acquisition that posit chunk-based learning from usage, construction grammar, and linguistic structure as processing history.

<B>Contingency

<TXT>

Contingency is essentially a probabilistic association between cue and outcome. In the context of morpheme development, a cue can be the lemma of an inflected form with its corresponding outcome the inflected form. One measure of such an association is reliability, which is a conditional probability of an inflected form given its lemma. For example, the frequency of the inflected form *arrived* in the Corpus of Contemporary American English (COCA; Davies, 2008-) is 56,456 and the total frequency of the lemma *arrive* is 99,836. In this case, the reliability of *arrived* is .57 (56,456 / 99,836). This means that when the lemma *arrive* is used, it is often (just over half the time) used to indicate events in the past. This in turn suggests that the lemma *arrive* is likely to be fairly strongly associated with past tense, and may, thus, be processed more easily in past tense forms. Psychological research into animal and human learning alike has demonstrated profound and omnipresent impacts of contingency in the learning of cue-outcome associations (Shanks, 1995). Reliability has been shown to influence language change (Bybee, 1985) as well as accuracy and error patterns in L1 (N. C. Ellis, O'Donnell, & Römer, 2014; Hay, 2001; Matthews & Theakston, 2006; Tatsumi et al., 2018) and L2 processing and acquisition (N. C. Ellis et al., 2014; Sugaya & Shirai, 2009).

Contingency learning is central to the competition model, a psycholinguistic theory of language acquisition and sentence processing (MacWhinney, 1987; MacWhinney & Bates, 1989), as well as to other psycholinguistic models of construction learning such as the rational learning of form-function contingencies (N. C. Ellis, 2006). The competition model focuses on the various morphological, syntactic, and semantic linguistic cues contained in a sentence—for example, case marking, word order, and semantic characteristics such as animacy—that people use to interpret the meaning of that sentence. Each cue is probabilistically associated with a particular interpretation, and the cue-weights combine to allow learners to choose the interpretation with the highest likelihood. Learners, particularly those in naturalistic learning contexts, assign cue-weights inductively over their history of experience and usage. Cue-weights differ between languages because different languages use different cues to signal meanings. Thus, L2 learners must learn which cues are important in which languages. To do this, they begin with cues that are more available in the input and/or those that they are most attuned to from their L1 (MacWhinney, 2008), after which they come to rely upon cues that are more reliable in their interpretations. Cues that are rare and unreliable are learned late and are relatively weaker, even in adults (MacWhinney, 1997).

<B>Formulaicity

<TXT>

Formulaicity refers to the extent to which a given word sequence is a fixed, prefabricated, or memorized expression. Studies have demonstrated that more formulaic language is processed faster and acquired earlier than less formulaic language both in L1 (e.g., Bannard & Matthews, 2008; Tremblay, Derwing, Libben, & Westbury, 2011) and in L2 (for review, see Siyanova-Chanturia & Pellicer-Sánchez, 2018). Formulaicity is typically operationalized via either frequency or association strength. Using frequency alone, however, can be problematic (N. C. Ellis, 2012). The expression *a lot of Americans*, for instance, is frequent as a 4-gram (i.e., a sequence of four words). This, however, is possibly only because the first trigram (i.e., *a lot of*) is frequent and not necessarily because there is an inherent association between *a lot of* and *Americans*. Because *a lot of* co-occurs with a huge number of plural nouns, one needs to examine the probability of observing *Americans* given that the preceding context is *a lot of* in order to properly capture the formulaicity of the phrase. One way to achieve this goal is by using unidirectional association measures such as ΔP (N. C. Ellis, 2006; Gries, 2013). Originally proposed by Allan (1980), ΔP is a directional contingency metric quantifying the strength between a cue and an outcome.[1] Gries (2013) extended the use of the metric to collocation research and showed that, unlike traditional collocational measures, ΔP captures directionality in collocation (e.g., the probability that the preceding word is *of* given the current word *course* is much higher than the probability that the following word is *course* given the current word *of*). The measure can also be applied to multiword expressions to quantify the directional association between the surrounding words and the target inflected word (see Dunn, 2018, for a variety of ΔP-based indices of formulaicity in multiword expressions; see Siyanova-Chanturia & Spina, 2020 for a recent example in L2 writing research).

Formulaic contexts as well as slot-and-frame patterns (Braine, 1976) or frames (Mintz, 2003) can strongly constrain the nature of the slot-filler (e.g., the context *once upon a* __ highly predicts the word *time*; the context *a lot of* __ highly predicts nouns; the frame *to* __ *it* highly predicts verbs, etc.). Such distributional information can be potent in the acquisition of both the grammatical and the semantic properties of the slot-filler (Elman, 1990; Redington, Chater, & Finch, 1998). Mintz, Wang, and Li (2014) compared training situations in which target words in an artificial language occurred in frames (i.e., surrounded by two words that frequently co-occur) against situations in which target words occurred in simpler bigram contexts (i.e., where an immediately adjacent word provides the context for categorization). They found that learners categorized words together when the words occurred in similar frame contexts but not when the words occurred in similar bigram contexts. In a study of L1 English-speaking 2.5-year-olds, Childers and Tomasello (2001) found that a nonce verb was better acquired for subsequent creative use in a transitive utterance when the nonce word was surrounded by pronouns than when it was surrounded by proper nouns or names, suggesting that children's transitive schema may start out with pronouns in preverbal/postverbal positions (i.e., pronoun V pronoun) rather than being fully general. In these ways, formulaic frames might positively promote the processing and productivity of their subcomponent words.

<B>Previous Research on Availability, Contingency, and Formulaicity Using Elicited Imitation Tests

<TXT>

Guo and Ellis (2021) performed two experiments investigating the effects of availability, contingency, and formulaicity on the suppliance of accurate morphemes in an aural to written elicited imitation task (EIT). Specifically, Chinese learners of English were asked to listen to sentences containing grammatical morphemes and to type as much as they could remember of the sentence that they had heard. Targeting past tense *-ed*, progressive *-ing*, third person *-s*, and plural *-s*, Guo and Ellis investigated whether the three distributional factors calculated based on COCA were positively associated with correct suppliance of the morphemes. In both experiments:

1.there were highly significant effects of reliability (i.e., contingency) on the accuracy of provision (particularly of plural *-s*, third-person present-tense *-s*, and progressive *-ing)*;

2.word-form frequency had smaller but significant effects particularly on plural *-s* and third-person present-tense *-s*;

3. there were no effects of lemma frequency;

4. effects of reliability (but not frequency) were greater at lower levels of proficiency; and

5.Experiment 2 also demonstrated phrase-superiority effects whereby higher frequency 4-word strings were associated with increased accuracy of production of the morphemes embedded therein.

Whereas this study was certainly informative, its experimental nature nevertheless limited its scope for the numbers of learners and words targeted. Also, the highly controlled contexts of the morpheme occurrences and the use of elicited imitation as a language processing task afforded experimental validity through controlled decontextualized language perception and repetition at the expense of ecological validity in situations of rich, meaningful, and more-naturalistic communication.

Our study, therefore, aimed to complement and extend the investigations of Guo and Ellis (2021) by drawing data from a large-scale learner corpus of L2 writing. The analysis of a large-scale corpus allowed us to target a greater number and range of words and learners, leading to a study with a larger scope and a more fine-grained picture of the effects of relevant factors than those of the Guo and Ellis study. The difference between EIT and writing, however, warranted attention. In the following section, we discuss how specifically they differ and how the differences might impact the association between distributional factors and morpheme accuracy.

<B>Automatic Versus Controlled Language Processing

<TXT>

Frequency effects are more often evidenced in automatic online language processing rather than in more considered production tasks where there is more time for conscious creation and editing. EIT involves subjects' listening to a single sentence out of context and then their repeating that sentence verbatim in all of its parts as accurately as possible. Unlike written composition (i.e., with content generated by the learner), it requires fluent online perception involving implicit or automatized processing. In designing and validating a battery of language tests that separately tap implicit and explicit learning, R. Ellis (2005) reported factor analyses of measures for L2 learners where elicited imitation, oral narration, and timed grammaticality judgment tests loaded on Implicit Knowledge whereas metalinguistic knowledge tests and untimed grammaticality judgment tests loaded on Explicit Knowledge. Subsequent construct validation studies have provided confirmatory empirical support for the oral EIT as a measure of implicit linguistic knowledge (Bowles, 2011; Erlam, 2006; R. Ellis, Loewen, & Erlam, 2006; Spada, Shiu, & Tomita, 2015), or at least of automatized explicit knowledge (Suzuki & DeKeyser, 2015). Implicit, as well as automatized processing takes place automatically, ballistically, and unconsciously. It runs on the system of linguistic representations that have been tuned by prior usage.

Guo and Ellis (2021) understood the effects of availability, contingency, and formulaicity upon accuracy of morpheme production in the EIT task as follows. The perception and comprehension encoding stages involve three parts: (a) taking word-forms into an auditory/lexical buffer, (b) linking lexical items syntactically, and (c) constructing a meaningful interpretation of the sentence. Psycholinguistic research has demonstrated a variety of frequency effects in the perception and processing of words, morphemes, multiword chunks, and syntactic constructions, and likewise here too the recognition and preservation of the correct form of target words in EIT is influenced by the forces of availability, contingency, and formulaicity. Models of L1 auditory sentence processing such as chunk-and-pass processing (Christiansen & Chater, 2016) suggest that the language system rapidly integrates all available incoming information, interactively satisfying multiple probabilistic constraints as quickly as possible, to update the current interpretation of what has been said so far. Relevant cues include sentence-internal information about lexical and structural biases as well as extra-sentential cues from the referential and pragmatic context (although the decontextualized nature of EIT denies many of these usual additional influences). As the incoming auditory information is chunked, it is rapidly integrated with prior information to recognize words and morphemes that are in turn chunked into larger multiword units. Incremental identification of incoming units is influenced by the sequential probabilities of what has been processed to date: The next word in a well-entrenched word sequence is more easily identified as is an incoming morpheme that is highly predicted in its context. In parsing and interpreting the target morphemes, there are influences of syntactic integrity, for example, auxiliary [*be*] impacting particularly progressive *-ing* and influences of contextual support that could impact the encoding of the past *-ed*. The encoding of third person present *-s* and plural *-s* on subjects is also under the influence of syntactic integrity, although in English, agreement processing is generally less obligatory than processing for tense and aspect (MacWhinney, 1997, 2008). The final stage of EIT, production (whether oral or written), could also be sensitive to frequency effects and sequential probabilities at word, morpheme, and, particularly, phrasal levels: A well-entrenched formulaic phrase can support provision of its component morphemes whether they are analyzed or not (Arnon & Clark, 2011). The more the demand for immediacy and automaticity of production, the greater the likely impact of these factors.

Written composition has much more scope for variability of processing than does spoken EIT even with a written production component. At one extreme, as in texting, a relevant process analysis might look quite like that for speaking (e.g., Levelt, 1989)—something fast, skilled, and automatic that builds upon highly specialized mechanisms dedicated to performing specific subroutines such as retrieving appropriate words, generating morpho-syntactic structure, computing the phonological target shape of syllables, words, phrases and whole utterances, and creating and executing articulatory programs. At this extreme, one might expect probabilistic effects to be at their strongest. Formulaic language is more common in speech than in writing (Erman & Warren, 2000), and the observation that memorized clauses and clause-sequences form a high proportion of the fluent stretches of speech heard in everyday conversation led Pawley and Syder (1983) to propose that it is this use of memorized language that underpins fluency. However, most writing is far less mundane than texting, and at the other extreme, writing can look more like this current article, the result of numerous edits, reworkings, and conscious revisions.

On the whole, writing is considerably slower than speech, and it involves much more conscious analysis and attention “to complex aspects of writing during decision making; problem-solving behaviors involving heuristic searches; and well-differentiated control strategies” (Cumming, 1989, p. 81). These conscious strategies override the distributional norms that are offered up from the automatized/implicit system. Editing disrupts what is first offered up to consciousness by the implicit systems, then rearranges and builds upon what is on the page, trying to find new and interesting ways of better expressing ideas. There is immense scope for creativity in choosing where, and at what levels of grain, to follow the norms and where to exploit them (Hanks, 2013).

Psychometric investigations confirm that writing is more associated with explicit processing. Elder and Ellis (2009) examined learners of English as a L2 with respect to their performance on two measures thought to tap into implicit knowledge (an EIT and a timed grammaticality judgment test) and two measures considered to tap more into explicit knowledge (an untimed grammaticality judgment test and a multilinguistic knowledge test), and the relationship of these scores with those obtained on the different sections of the International English Language Testing System test and showed that implicit knowledge correlated more strongly with the oral skills (speaking and listening) and explicit knowledge correlated more strongly with the written skills (writing and reading).

Automaticity relies upon memorized structures and habits; consciousness affords novelty (N. C. Ellis, 2015). These considerations make clear some important differences between EIT and writing composition that potentially moderate their reliance upon formulaic knowledge and sequential distributional norms. Our study, therefore, helped us to shed light on the potential differences between EIT and writing composition for the association between distributional factors and morpheme accuracy.

<A>The Present Study

<TXT>

The present study addressed the following research questions:

1.Do L2 learners supply English grammatical morphemes in their obligatory contexts more frequently in more frequent and reliable words as well as in the contexts with higher ΔPs?

2. To what extent do the distributional factors of Research Question 1 interact with learners’ L2 proficiency and/or longitudinal development?

Research Question 1 was directly concerned with a conceptual replication of Guo and Ellis (2021). We examined whether the difference in tasks (i.e., EITs involving real-time aural perception and written production in a controlled environment vs., here, morpheme provision in more consciously considered and meaningful free-writing tasks) and scope (here including tens of thousands of participants ranging widely in proficiency, age, language background, geography, and other demographics) yielded any difference in the relationship between distributional factors and the accuracy of morpheme use. Research Question 2 was more exploratory in its investigation of interactions whereby learners may, for instance, be more sensitive to distributional factors when their proficiency is low, whereas morpheme use might be more robust against lower availability, contingency, and formulaicity for high-proficiency learners.

<A>Method

<B>Data

<TXT>

We drew the primary data for our study from the EF-Cambridge Open Language Database (EFCAMDAT; Geertzen, Alexopoulou, & Korhonen, 2014), a partially error-tagged large-scale longitudinal learner corpus. The corpus includes writings submitted to Englishtown, an online school formerly run by EF Education First.[2] The course in Englishtown consisted of 16 levels, each of which included six or eight units depending on the subcorpus. In each unit, students were asked to respond to a prompt in a free writing task; these responses constitute EFCAMDAT. The topics of writing varied widely from self-introduction to letter writing to argumentative writing (see Alexopoulou, Michel, Murakami, & Meurers, 2017; Michel, Murakami, Alexopoulou, & Meurers, 2019). The corpus currently includes approximately 147 million words in 1.2 million writings by 175,000 learners from a wide range of nationality groups. Because the writings were submitted to an online school, it is possible to track the development of individual learners. Furthermore, approximately two-thirds of the writings (782,648 writings; 66%) come with teacher corrections that we used as error annotation to calculate morpheme accuracy.[3] In this study, we used only the error-annotated portion of the corpus.

<C>*Target Morphemes*

<TXT>

In EFCAMDAT, we identified the errors and accurate uses of the same set of four inflectional morphemes as Guo and Ellis (2021) had used: past tense *-ed*, progressive *-ing*, third person *-s*, and plural *-s*. We examined multiple morphemes because we could be more confident that distributional factors such as frequency are the general principles underlying accurate morpheme use if the factors were positively correlated with accuracy across the morphemes. We examined only omission (e.g., the lack of plural *-s*) and misformation errors (e.g., use of third person *-s* when past tense *-ed* was required). We did not analyze overgeneralization errors (e.g., suppliance of plural *-s* when a singular noun was required) because we hypothesized that the direction of the effects of availability, contingency, and formulaicity would differ between omission and misformation errors on the one hand and overgeneralization errors on the other. We expected the three factors to decrease omission and misformation errors because their high values presumably lead to accurate uses of the target morphemes. We would have expected overgeneralization errors to increase, however, because their high values presumably lead to unnecessary uses of the target morphemes.

Past tense *-ed* included only regular past tense forms and not irregular ones (e.g., *went*). We identified candidates for the irregular forms in a bottom-up manner by consulting the COCA. We have reported the detailed procedure in Appendix S1 in the online Supporting Information. We did not target other uses of *-ed* such as those as past participles in perfect aspect. Progressive *-ing* included the *be* + *-ing* form but did not include *be going to* or gerunds. Third person *-s* excluded irregular forms (i.e., *be*, *have*, and *do*). Plural *-s* did not include irregular forms (e.g., *children*), which we also identified in a bottom-up manner through the COCA. It excluded pronouns (e.g., *others*) as well.

<C>*Data Extraction*

<TXT>

We used the following procedure to identify the errors and accurate uses of the target morphemes in EFCAMDAT. If a writing included punctuation, capitalization, or space-related errors, we replaced the original learner writing with the corresponding teacher-corrected portion insofar as these errors were concerned because the first two error types have been demonstrated to lower the accuracy of automated analysis (Huang, Murakami, Alexopoulou, & Korhonen, 2018) and the last error type has often led to erroneous part-of-speech tagging due, for instance, to inaccurate tokenization (e.g., two words concatenated without a space). Spelling errors often result in parsing errors as well, but we retained these spelling errors uncorrected because the correction of grammatical morphemes is occasionally coded as spelling errors in EFCAMDAT. We then annotated the original (i.e., learner-written) and corrected (i.e., teacher-corrected) versions of each writing with part-of-speech tags via TreeTagger (Schmid, 1994) with the provided English model in order to solve ambiguities (e.g., plural *-s* vs. third person *-s*).[4] We standardized spelling variations (e.g., *realised* vs. *realized*, *colours* vs. *colors*) prior to identifying errors and accurate uses.[5]

We used a script in the R software (R Core Team, 2020) to identify the errors and accurate uses of the target morphemes automatically. We counted the lack of a morpheme as an error only when the same lemma was used in both the original form and the corrected form. If, for instance, the clause *I make a party* was corrected into *I am having a party*, we did not count this as an error of progressive *-ing* because the lemmas differed (i.e., *make* vs *have*). We did not count spelling errors (e.g., *partys* instead of *parties*) as morpheme errors, either. We have provided the detailed procedure for identifying errors, accurate uses, and irregular forms, as well as the evaluation of the identification accuracy of errors and accurate uses, in Appendix S1 in the online Supporting Information.

<C>*Subcorpus*

<TXT>

One issue with the original EFCAMDAT data is that a single topic ID was often assigned to writings that were written with different topics, making it difficult to control for topic effects. As part of an endeavor to clean up EFCAMDAT, Shatz (2020) estimated the correct topics of writings in EFCAMDAT, and we used the topic IDs in Shatz (2020) rather than the original IDs in accounting for topic effects. Accordingly, we targeted only the writings whose topics were estimated in Shatz (2020), which included writings from 11 nationality groups spanning the A1 to C1 levels (i.e., 15 Englishtown levels in total) in the Common European Framework of Reference (CEFR; Council of Europe, 2001).

In addition, for the sake of reliability, we targeted only words with a minimum of five occurrences in each of the CEFR levels A1 through C1 in the analysis of the three verbal morphemes. According to EF Education First, the first three levels of Englishtown correspond to A1 in CEFR, the second three levels to A2, and so forth (Geertzen et al., 2014). We followed this alignment in the assignment of CEFR levels to each writing. For plural *-s*, however, we employed different criteria to reduce its otherwise unmanageably large data size. Specifically, we targeted only the learners who contributed 20 or more writings with at least one error or accurate use of plural *-s*, and among them, we included the word forms with a minimum of 20 occurrences (across all the writings of learners who contributed 20 or more writings) at each of the CEFR levels A1 through C1 in the analysis. After the application of the above criteria, we further excluded several word forms manually because they were not used as target morphemes (e.g., *according* as progressive *-ing*). Table 1 shows the resulting number of learners, writings, words, and obligatory contexts for each morpheme. The distribution of the nationality groups was skewed, with Brazilian learners occupying 43% to 48% of observations for each of the morphemes, followed by Chinese, Mexican, and German learners each contributing 6% to 13% for each morpheme. We have provided the full lists of words targeted for each morpheme in Appendix S1 in the online Supporting Information.

<COMP: Place Table 1 near here>

<C>*Distributional Factors*

<TXT>

We computed availability, contingency, and formulaicity based on the COCA version released in March 2020 (Davies, 2008-) that has been used in previous research as a first approximation of L2 learners' input (e.g., Monteiro, Crossley, & Kyle, 2020; Wolter & Yamashita, 2018). We operationalized availability as the surface-form frequency of the inflected word tagged as a verb (in the case of past tense *-ed*, progressive *-ing*, and third person *-s*) or as a noun in the case of plural *-s*.[6] Word forms such as *used*, *getting*, *says*, and *years* were among the high frequency items that we targeted in this study, and those like *requested*, *raining*, *prefers*, and *supermarkets* were among the low frequency items. We operationalized contingency as reliability that we calculated by dividing surface-form frequency (i.e., availability) by the frequency of the corresponding lemma within the same part of speech (i.e., verb or noun). High reliability forms included *decided*, *trying*, *depends*, and *inhabitants*, whereas low reliability forms included *liked*, *having*, *sees*, and *worlds*.

We measured formulaicity as ΔP at the level of individual obligatory contexts of each morpheme. Specifically, we identified trigrams to 5-grams (i.e., sequences of three to five words) including the target inflected form in the corrected writing, calculated ΔP in each sequence, and considered the highest standardized value as the index of formulaicity of the context. The calculation of ΔP is [*p*(outcome | cue)] – [*p*(outcome | ¬ cue)], that is, the probability of the outcome given a cue minus the probability of the outcome without the cue. In our case, a cue was the words surrounding the target inflected word in the corrected writing, and an outcome was the inflected word. Appendix S2 in the online Supporting Information provides an example of the specific calculation of ΔP. ΔP values can vary from +1 to –1, with larger values indicating stronger association and negative values indicating inverse association (i.e., the presence of a cue predicts the absence of the outcome).

We identified the target context in a moving-window manner as illustrated with an example in Table 2. In this example, the context around the target inflected word (*kinds*) was *can deal with all kinds of people and i*. From the context, we extracted all the trigrams, 4-grams, and 5-grams that included the target form and calculated ΔP in each sequence, with the inflected form as the outcome and all the other surrounding words as the cue. We then identified the largest ΔP value in each n-gram (underlined in the ΔP column in Table 2). However, those values were not comparable because their absolute values and variance differed across trigrams to 5-grams. We, therefore, log-transformed those largest ΔP values so that their distributions more closely approached a normal disribution[7] and then standardized them to a mean of 0 and a standard deviation of 1 within each of trigrams, 4-grams, and 5-grams. The largest standardized log-transformed ΔP value (underlined in the rightmost column in the table) was the value that we considered to represent the formulaicity of the particular context.

<COMP: Place Table 2 near here>

In the identification of formulaic sequences, we considered the segment that did not include any alphabet, digit, or underscore as the boundary that a sequence could not cross. If, for example, the target inflected word occurred right before a punctuation mark (e.g., *rarely go on business trips.*), we identified only one trigram (*on business trips*), one 4-gram (*go on business trips*), and one 5-gram (*rarely go on business trips*). Also, we uncontracted contractions (e.g., *’m —> am*) and replaced all sequences of digits (numbers) by the nonword *NUM* because formulaicity is presumably generally stable across the values of numbers (e.g., *10 years ago* vs. *20 years ago*). Sequences with high ΔP included *since i graduated from college*, *and time is running out*, *practice makes perfect*, and *ladies and gentlemen*, and those with low ΔP included *wanted a lot of*, *going is not*, *says do not*, and *in the years of*.

<B>Data Analysis

<TXT>

In each morpheme, we fit a Bayesian mixed-effects binary logistic regression model to predict accurate uses against errors, which were the reference level. Bayesian models were suited for our analysis because the number of target words was not particularly large and Bayesian analyses with weakly informative priors allowed us to more reliably estimate the parameters of interest (see Gudmestad, House, & Geeslin, 2013, for similar arguments in L2 research). More generally, in comparison to frequentist analyses, Bayesian methods yield (a) more reliable parameter estimates in complex statistical models such as the ones used in our study, (b) the full posterior distribution that is more informative than point estimates, and (c) more intuitive uncertainty metrics (Kruschke, 2014; Lambert, 2018; McElreath, 2020; see also Norouzian, de Miranda, & Plonsky, 2018). Fixed-effects variables included learners’ proficiency that we operationalized as their mean Englishtown Level, learners’ writing number representing longitudinal development in an ordinal manner (e.g., 1 for the first writing, 2 for the second writing), L1 type indicating whether an equivalent feature to the target morpheme was obligatory in the language predominantly spoken in the country or region of the learner’s nationality (see Table 3), frequency, reliability, ΔP, and the two-way interactions among the parameters mentioned above. We did not include higher-order interactions or nonlinear terms to avoid excessive complexity of the model (but see Murakami, 2016). We used the treatment contrast to code L1 type with the ABSENT group as the reference group. We log-transformed frequency because the log-transformed values would more likely to result in a linear relationship between frequency and logit-transformed accuracy (i.e., its effects were assumed to be multiplicative). We then standardized all the fixed-effects variables except L1 type to a mean of 0 and a standard deviation of 1.

<COMP: Place Table 3 near here>

For the random-effects structure, the grouping variables included learners, their nationalities, inflected forms, topics of writing, and the interaction between inflected forms and topics of writing. We included the form-topic interaction to account for the influence of prompts and example writings on specific inflected forms. We expected to find the higher accuracy of an inflected form, for instance, if the same form occurred in the task prompt. In addition to random intercepts, we included the following random slopes: (a) by-learner random slopes for writing number, frequency, reliability, and ΔP, (b) by-nationality random slopes for the mean Englishtown level, writing number, frequency, reliability, and ΔP, (c) by-inflected-form random slopes for the mean Englishtown level, writing number, and ΔP, and (d) by-topic random slopes for L1 type and the same set of predictors as (b). This random-effects structure was close to the maximal model (Barr, Levy, Scheepers, & Tily, 2013), except that we did not include the interaction terms as random slopes. We did not perform variable selection due to the many problems associated with this procedure (e.g., Harrell, 2015).

Largely in line with the recommendations offered by the Stan Development Team (2020), we used the following weakly informative prior distributions: normal distribution with the standard deviation of 2 for the intercept, *t* distribution with three degrees of freedom for fixed-effects slope and contrast parameters, half-normal distribution with the standard deviation of 1 for the random-effects standard deviation, and the Lewandowski-Kurowicka-Joe (LKJ) distribution with the prior parameter value of 1 (i.e., flat prior bounded between –1 and +1) for the random-effects correlations. The priors, thus, regularized the intercept, slopes, and random-effects standard deviations towards 0, penalizing extreme values and potentially allowing more reliable parameter estimation despite the complexity of the model. We derived the posterior distributions based on Hamiltonian Monte Carlo with four Markov chains with 5,000 iterations each, including 1,000 warmup iterations. Appendix S3 in the online Supporting Information presents the prior predictive checks (Schad, Betancourt, & Vasishth, 2021). We fit the models with brms (Bürkner, 2017), a front-end R package of Stan (Carpenter et al., 2017). For the sake of transparency, we have made available the R code used for the preprocessing of corpora, data retrieval, and data analysis, together with the data used for modeling (Murakami & Ellis, 2022) at IRIS (www.iris-database.org) and OSF (https://osf.io/ba8mf/).

<A>Results

<B>Descriptive Figures

<TXT>

Figure 1 shows the accuracy of each morpheme across the mean Englishtown levels, writing number, L1 type, frequency, reliability, and ΔP values. The mean Englishtown level appeared to be largely positively correlated with accuracy, but the relationship was less clear in writing number and L1 type. Frequency looked to be positively correlated with accuracy for progressive *-ing*, third person *-s*, and plural *-s* but not for past tense *-ed*. Reliability and accuracy were positively associated for all the morphemes except for progressive *-ing*. ΔP was positively correlated with accuracy for progressive *-ing* and third person *-s*, but the relationship was unclear for the other two morphemes. These observations based on the figure, however, were somewhat impressionistic and also ignored various sources of variability (e.g., individual variation). Statistical modeling allowed us to more formally examine the relationships between our predictors and morpheme accuracy.

<COMP: Place Figure 1 near here>

<B>Results of Statistical Modeling

<TXT>

All the models converged reasonably well (all $\hat{R}$ values < 1.05). Appendix S4 in the online Supporting Information presents trace plots, posterior predictive checks, and a sensitivity analysis. Appendix S5 in the online Supporting Information presents the numerical summary of each parameter in each model as well as the histograms showing the posterior distribution of each fixed-effects parameter. Figure 2 demonstrates the posterior mean and the 95% credible interval of each fixed-effects parameter for each morpheme. As we expected, when the other predictors were at their mean values or the reference level, the mean Englishtown level (i.e., proficiency) was positively associated with accuracy for each morpheme. Writing number was also positively correlated with accuracy, although the 95% credible intervals excluded 0 only for past tense *-ed* and for progressive *-ing*. Similarly, for L1 type, 95% credible intervals excluded 0 only for plural *-s*, where the PRESENT group outperformed the ABSENT group. The lack of clear evidence for L1 influence might seem surprising given its pervasive impact documented in the literature (e.g., Jarvis & Pavlenko, 2007). The finding, however, replicated Murakami (2014), who did not identify a clear effect of L1 on third person *-s* or progressive *-ing* but found some L1 influence on plural *-s* in an earlier version of EFCAMDAT. The results were also partly consistent with Murakami and Alexopoulou (2016), who found that L1 hardly influenced the accuracy of third person *-s* although it exerted a stronger effect on plural *-s* in the Cambridge Learner Corpus (Nicholls, 2003).[8]

The 95% credible intervals of the frequency parameter only excluded 0 for progressive *-ing*, suggesting that the effects of token frequency could not be reliably identified in our data. Reliability, however, was consistently positively correlated with accuracy for all the morphemes. When the other predictors took their mean values or the reference level, the increase of one standard deviation of reliability was associated with an accuracy increase of 0.5% for progressive *-ing* to 5.5% for third person *-s*. Although these accuracy increases may seem small, they were still substantial given that overall accuracy approached ceiling (see Figure 1). As was the case for frequency, ΔP was not strongly associated with accuracy, except for past tense *-ed* where it was negatively correlated with accuracy.

Turning to interaction terms, we observed that most of the 95% credible intervals included 0, which meant that the relationship between morpheme accuracy and our predictor variables did not necessarily differ depending on the other predictors included in the model. When we analyzed our focal variables (i.e., frequency, reliability, and ΔP), the only parameters for which 95% credible intervals did not include 0 were the interaction of the mean Englishtown level and frequency for third person *-s*, the interaction of writing number and frequency for third person *-s*, and the interaction of writing number and ΔP for progressive *-ing*. Notice that all the interactions involved the mean Englishtown level or writing number on the one hand and one of the distributional factors on the other. The question was whether these interactions were strong enough to drive the association between the distributional factors and accuracy to a meaningful level depending on the mean Englishtown levels or writing numbers without extrapolation (i.e., within the range of the mean Englishtown level and writing number observed in this study). To check this, we examined the posterior distribution of the distributional factors across the mean Englishtown levels or writing numbers more closely. With respect to the interactions of writing number and frequency for third person *-s* and of writing number and ΔP for progressive *-ing*, the 95% credible intervals of the posterior distribution of the distributional factors (i.e., frequency and ΔP) only excluded 0 after 112 writings (for a single learner) for the interaction of writing number and frequency for third person *-s* and 113 writings for the interaction of writing number and ΔP for progressive *-ing*. We did not concern ourselves with these two interaction terms because only a small number of learners in our data contributed 112/113 writings (i.e., four learners in the interaction of writing number and frequency for third person *-s* and three learners in the interaction of writing number and ΔP for progressive *-ing*). The interaction of the mean Englishtown level and frequency for third person *-s*, however, deserved our attention. The 95% credible intervals of the posterior distribution of frequency came to exclude 0 at Level 8 Unit 7 (CEFR B1) or higher levels in Englishtown (see Figure 3). In other words, the frequency of the inflected form of third person *-s* was positively associated with accuracy from intermediate proficiency levels onward.

<COMP: Place Figure 2 near here>

<COMP: Place Figure 3 near here>

<A>Discussion

<TXT>

In the present study, we investigated whether L2 learners tend to use grammatical morphemes more accurately when the inflected form is more available and reliable as well as when it occurs in a more formulaic context. Our corpus analyses of learners' free written production data demonstrated that reliability, or the contingency between the inflected form and the lemma, was consistently positively associated with higher accuracy for each of the four target grammatical morphemes (i.e., past tense *-ed*, progressive *-ing*, third person -*s*, plural -*s*). However, we could not identify a clear association between frequency and accuracy, except for progressive *-ing* (across proficiency levels) and for third person *-s* (intermediate- to high-proficiency levels). Nor did we detect a consistent pattern between formulaicity and accuracy.

Some of these findings replicate those of Guo and Ellis (2021), who investigated these effects in an EIT; some do not. Our discussion will therefore be structured as follows. We analyze how the language processing requirements in free written production might be affected or not by contingency, availability, and formulaicity. Because contingency is by far the most potent variable in our study, we relate it to work on the aspect hypothesis but also emphasize that it happens with all of the morphemes investigated here, including plural *-s*. We speculate that it might be a more general phenomenon that reflects the semantic and functional motivations of cognitive grammar. We then discuss interactions with learner proficiency. Finally, we consider the advantages and the limitations of large-scale learner corpus investigations such as our study.

<B>Pervasive Effects of Contingency

<TXT>

In our study, higher contingency between the inflected form and the lemma is consistently positively associated with higher accuracy of morpheme provision in students' writing. This finding is consistent with findings reported in the previous literature on usage-based theories that have repeatedly demonstrated the facilitative role of contingency in language acquisition, processing, and use (e.g., N. C. Ellis et al., 2014; Matthews & Theakston, 2006). It suggests that the lemma, which arguably reflects the semantic meaning of the target word, functions as a cue for its inflected forms and that L2 learners use such contingency in processing inflectional morphemes.

Contingency of cue-outcome mapping is a driving force of all associative learning, human and animal alike, to the degree that the field in which it is studied has become known as contingency learning (Shanks, 1995). Contingency of association is similarly key in cognitive-linguistic, corpus-based, and statistical models of language structure like collostructional analysis (Stefanowitsch & Gries, 2003; Gries & Stefanowitsch, 2004). Cognitive linguistic theories of construction grammar focus on lexical, morphological, and syntactic forms as form-function pairings (Goldberg, 1995, 2006; Trousdale & Hoffmann, 2013). Collostructional analysis focuses more on form-form reliability in measuring the degree of attraction or repulsion that words exhibit to constructions. These measures of association have been found to be better predictors of interpretation than have been measures of availability (N. C. Ellis et al., 2016; Gries & Ellis, 2015; Gries, 2015).

We have already described how, in L1 acquisition, the relative frequency of different forms of the same word have been found to predict the usage and error patterns in morpheme acquisition (Bybee, 1985; Hay, 2001; Matthews & Theakston, 2006). There is parallel L2 research showing the importance of contingency in production. Sugaya and Shirai (2009) described the case study of a native Russian speaker learning Japanese over the course of 10 months. They found that verbs that are more consistently conjugated in a certain common form compared to other possible forms, such as *siru* "come to know" with the imperfective aspect morpheme -*te-i-(ru),* were produced exclusively in the common form early in the learning trajectory, although this preferential bias was not observed for verbs that do not have a common form. In a recent study on Japanese L1 acquisition, Tatsumi, Ambridge, and Pine (2018) investigated 3- to 5-year-olds' productive use of different forms (simple past tense vs. completive past tense) of verbs in a primed elicited production paradigm in which the children described actions in line-drawings after hearing the experimenter describing the previous drawing using a verb in the uncommon completive past-tense form. It was found that the children's choice between the simple and completive forms for each verb reflected the relative frequency of the two forms in corpus data. Although the simple form was generally favored, verbs that have a higher completive past-tense to simple past-tense ratio were more likely to be successfully primed by the experimenter's use of the completive form compared to other verbs.

These studies have raised larger questions that we will return to in the section Why Reliability, Especially?, questions of why some verbs are more consistently conjugated in a certain common form compared to other possible forms, and, more generally, why some lemmas are more reliably found associated with particular morphological inflections. Whatever the answers to these questions concerning the quotidian motivations of usage, our findings about L2 acquisition and processing are clear: Highly reliable morphemes (i.e., exemplars involving lemmas more consistently conjugated in the form containing this morpheme) are more readily acquired and produced accurately. Learners' experience of these patterns of usage has tuned their language systems so that they are sensitive to the frequency distribution of inflectional forms within each lemma.

<B>Uneven Effects of Availability

<TXT>

In contrast to the effects of contingency, our study showed mixed and less compelling evidence of associations between availability and accuracy. Although there appeared to be positive associations in Figure 1 for progressive -*ing*, third person -*s*, and plural -*s*, in the Bayesian statistical modelling, when the other predictors were accounted for, the 95% credible intervals of the frequency parameter only excluded 0 for progressive -*ing*, suggesting that the effects of token frequency could not be reliably identified in our data.

The absence of the noticeable effects of token frequency was unexpected given the pervasive impact of frequency documented more generally in the usage-based literature (e.g., N. C. Ellis, 2002) and more specifically in Guo and Ellis (2021). Various explanations come to mind.

One answer lies in the differences between EITs and free written production. EITs require accurate online perception of an inflected word form with little contextual support and subsequent accurate reproduction for the inflected word, albeit in writing in Guo and Ellis (2021). In contrast, in creative written production, in response to the demands of the prompt and the current state of exposition, various ideas come to the minds of learners, and they respond with whatever available and relevant language forms that are associated with these ideas. Availability entails that higher proficiency learners will have at their disposal a wider range of items than will have low proficiency learners. But we did not score availability in this sense. We scored whether learners provided the correct inflection for the lemma that came to mind in that context; an outcome where contingency of association was much more pertinent than the effects of token frequency.

Another possible reason is that we only examined the reasonably frequent forms (i.e., inflected forms) in which frequency effects had presumably approached ceiling. In future research, it would be worth investigating the effects of token frequency in lower frequency items in order to elucidate the full scope of frequency (i.e., availability) effects.

<B>No Evidence for the Effects of Formulaicity

<TXT>

Our study similarly did not detect a consistent pattern between formulaicity and accuracy. This, again, was unexpected considering that previous usage-based research had shown general effects of formulaicity on language acquisition and processing (e.g., Bannard & Matthews, 2008) and, more specifically, the results of Guo and Ellis (2021) using an EIT. As with availability, there are possible explanations relating to the different task demands of EIT and writing; and there are possible explanations relating to operationalizations of formulaicity.

With respect to task differences, in the section Automatic Versus Controlled Language Processing, we explained how EITs prioritize online implicit and automatized processing running on the system of linguistic representations that have been tuned by prior usage, whereas writing, in contrast, involves more explicit, conscious processing. Conscious writing strategies may override the distributional norms that are offered up from the automatized/implicit system. Editing can disrupt what is first offered up to consciousness by the implicit systems, then rearranges and builds upon what is on the page, thus potentially moderating any immediate reliance upon formulaic knowledge and phrasal distributional norms.

The lack of clear formulaicity effects is interesting especially when it is contrasted with reliability. In our study, both reliability and ΔP could be considered as types of contingency indices and were measured in a similar manner. Reliability, however, was robustly associated with accuracy; ΔP was not. But there is a wide range of ways to measure formulaicity, ranging from string frequency through measures of association including mutual information, mutual expectation, ΔP (see examples such as Saito, 2020 or Tavakoli & Uchihara, 2020), and our field is still trying to understand the ramifications of these various operationalizations. To investigate formulaicity in their experiment, Guo and Ellis (2021) simply replaced pronoun contexts with proper-noun contexts, a blunt instrument that would have caused large reductions in formulaicity however it was measured.

In designing our study, we had to choose one measure of formulaicity, and we opted for ΔP given its theoretical motivations. We held firm to this choice rather than fishing with different measures. When we initially considered alternatives, simple string frequency came to mind, as did mutual expectation (Dias, Guilloré, Gabriel & Lopes, 1999). So too did the issue of separating between-lemma and within-lemma effects of formulaicity (see In'nami & Murakami, 2021). We also noted that word n-gram is a rather crude scope of formulaicity. It, for example, does not capture nonadjacent structural dependency of words (e.g., subject and verb) or association between abstract categories (e.g., part of speech). It is quite likely that learners are more sensitive to certain types of structural and semantic association than to others. The choice of an appropriate measure of formulaicity could, therefore, depend on the specific aspects of the construct that researchers wish to capture as well as on characteristics of the target language such as morphological typology (root, agglutinative, inflectional, or polysynthetic) and reliance on word order.

As long as it is acknowledged that any subsequent investigation is exploratory, there is scope and good reason to further investigate possible effects of these different types of formulaicity on morpheme provision, and this would be quite tractable using the corpus data here.

<B>Why Reliability, Especially?

<TXT>

Why is contingency of association more potent than availability? We can make sense of this from three perspectives: (a) associative learning theory, (b) cognitive linguistic theories of construction grammar, and (c) functional theories of SLA. Indeed, we see their confluence as an important theoretical triangulation where each theory informs and supports the other two.

Associative learning theory demonstrates that contingency of association trumps token frequency (as described in the section Contingency). In operationalizing contingency, we focused on how likely it is that a linguistic cue (a morpheme) reliably co-occurs with another (a lemma). But morphemes and lemmas go beyond being mere forms, they are linguistic constructions with particular functions and meanings; they are symbolic. Figure 4 illustrates the matrix of associations in de Saussure's (1916/1983) thought-sound relationship.

<COMP: Place Figure 4 near here>

Cognitive linguistic theories of construction grammar view lexical, morphological, and syntactic forms as symbolic form-function pairings and hold that people learn language from usage. Collostructional analysis focuses on form-form reliability in measuring the degree of attraction or repulsion that words exhibit to constructions. When learners are processing usage, they are tallying the associations among forms, among interpretations, and between forms and their interpretations. Verbs have interpretations and so do morphemes, and these can vary in their form-function contingency. Verbs and morphemes can be more or less reliably associated (form-form contingency).

Functional theories of SLA emphasize the interplay of form and meaning in acquisition. One much-researched example for morphology is the aspect hypothesis (Andersen & Shirai, 1994; Bardovi-Harlig, 2000; for a state-of-the-scholarship review of the last 20 years of research, see Bardovi-Harlig & Comajoan-Colomé, 2020). The aspect hypothesis builds on three main constructs: tense, grammatical aspect, and lexical aspect. Tense establishes the location of an event (or situation) in time with respect to the moment of speech or some other reference point. Grammatical aspect allows for "ways of viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3). For instance, in English, a contrast in grammatical aspect is found between simple past *John walked* and past progressive *John was walking*. In contrast, lexical aspect refers to semantic differences in verbs and their arguments (Dowty, 1979) such as whether a predicate has inherent duration (e.g., *walk*, *sleep*, and *kid* [v.]), or is punctual (e.g., *recognize*, *broke*, and *sigh*), or has elements of both duration and culmination (e.g., *walk a mile* and *paint a picture*). The aspect hypothesis predicts that "second language learners will initially be influenced by the inherent semantic aspect of verbs or predicates" (Andersen & Shirai, 1994, p. 533). Bardovi-Harlig and Comajoan-Colomé (2020) stated:

In its simplest form, the AH [aspect hypothesis] for SLA predicts that in the initial stages of the acquisition of tense-aspect morphology by adults, the acquisition of past morphology will be influenced by lexical aspectual categories. Namely, verbal morphology will be attracted to and will occur with predicates with similar semantics. Perfective past will occur with telic predicates (i.e., those with inherent endpoints), imperfective will occur with unbounded predicates, and progressive will occur with ongoing activities (p. 3).

Bardovi-Harlig and Comajoan-Colomé concluded from their review of about 30 studies that the aspect hypothesis accurately predicts the adult L2 acquisition of past morphology in a number of languages.

Associative learning theory explains that people learn contingencies that are relevant to their predictive processing. Cognitive linguistics explains why certain form-meaning relations are more contingent in usage. The aspect hypothesis has investigated these patterns in the initial acquisition of tense-aspect for past tense *-ed*, progressive *-ing*, and third person *-s*, and we have confirmed the within-lemma distributional effects of contingency in our study. In addition, our study showed these relations to hold for plural *-s* as well. We speculate that all of this might be part of a more general phenomenon that reflects the semantic and functional motivations of cognitive grammar. For noun number, we suspect that inherent number, *pluralia tantum*, and prototypically plural count nouns might lead the way. Form-form and form-function associations interact in various complex and adaptive ways in usage, and speakers' language systems reflect the history of their processing these associations. There is good reason and plenty of scope to study a broad range of morphology in this way.

<B>Learner Proficiency and Interactions With Availability and Contingency

<TXT>

The aspect hypothesis emphasizes these effects in the initial acquisition of tense-aspect. In our study, we examined whether distributional factors interact with the learners' proficiency and/or longitudinal development. We might, for instance, have expected that the contingency between the lemma and its inflected forms (i.e., reliability) would be attenuated as learners' proficiency increased and the overall accuracy improved, as Guo and Ellis (2021) had found. But our study did not provide support for this. In other words, reliability was associated with accuracy arguably to a similar degree regardless of L2 learners' development. Also, if anything, we might have expected frequency effects to attenuate at higher proficiency levels because these learners have presumably already been exposed to individual inflected forms a large number of times. As far as third person *-s* is concerned, however, we found the reverse pattern, thus again suggesting that the influence of distributional factors (or the lack thereof) can well remain throughout L2 learners' development.

<A>Strengths and Limitations of the Research and Future Directions

<TXT>

The present investigations would not have been possible without large corpora of language that reflect typical language exposure as well as large learner corpora covering tens of thousands of participants ranging widely in demographics. Only in the last few years has it become possible to bring together cognition, corpora, and computing at scale to triangulate research in usage-based language learning (N. C. Ellis, 2017). Some advantages of our research are that it is reasonably large in scale, that the range of learners is wider than that of typical experimental studies, and that our work is open and transparent as a result of our making available the R code and data.

A major limitation of our study is the lack of transparency of process while learners write. If only we had the same scope of learner data from online language processing tasks as we have in EFCAMDAT for writing.

<A>Conclusion

<TXT>

Using large-scale (learner) corpora, we investigated the relationship between usage-based distributional factors and the accuracy of inflectional morphemes in L2 English written production. Consistent with the previous literature, we found that contingency between inflected forms and their lemmas is robustly associated with morpheme accuracy, suggesting that L2 learners are sensitive to the within-lemma distribution of inflected forms in their input. Contrary to prior studies investigating speech and online processing, however, our study did not identify consistent effects of availability or formulaicity on morpheme accuracy. We also explored possible interactions between distributional factors and learners' proficiency levels or longitudinal development. We generally did not find interactions, pointing toward the possibility that learners are sensitive to certain distributional properties in their input throughout their development.

<A>Notes

[1]In this sense, ΔP is similar to reliability discussed earlier. ΔP is basically the reliability minus the probability of the outcome when the cue is absent. This means that ΔP applied to the relationship between a surface form as a cue and its lemma as an outcome reduces to reliability when an inflected form is associated with only one lemma. Also, if both indices are used as a measure of formulaicity, ΔP and reliability yield very similar values because the second term of the calculation of ΔP, that is, $c/(c + d)$ (discussed later), is often negligibly small when the reference corpus is large, resulting in a ΔP value dominated by the first term, that is, $a/(a + b)$, corresponding to reliability. Whereas both ΔP and reliability are contingency measures, the difference lies in the nature and scope of the cue (i.e., lemma vs. word forms surrounding the target inflected form), and this difference leads to the measurement of different distributional characteristics (i.e., distribution of surface forms within a lemma [reliability] vs. distribution of an inflected form across contexts [ΔP]).

[2]Englishtown has been replaced by English Live. The description here is of the version of Englishtown from which the corpus data were gathered.

[3]Because teacher corrections in EFCAMDAT were not intended to be error annotation, they are not necessarily exhaustive (e.g., Mita, Kiyono, Kaneko, Suzuki, & Inui, 2020). Murakami (2014), however, showed that the results of the accuracy analysis of grammatical morphemes in EFCAMDAT largely match those based on the Cambridge Learner Corpus, which has been manually error-tagged for the sake of error annotation (Nicholls, 2003). Insofar as we focused on the general pattern of the accuracy of grammatical morphemes, therefore, the noise in teacher corrections should not have impacted the conclusions that we have drawn.

[4]One often-raised concern in automatically annotating part-of-speech (POS) tags on learner language is their accuracy. Huang et al. (2018) evaluated the accuracy of automated syntactic parsing and POS tagging in EFCAMDAT and found that accuracy is generally high at around 95% in POS tagging (cf. 96% for native speaker data). It is, therefore, fair to assume that automated POS tagging in EFCAMDAT is generally accurate. Also, and more importantly, we manually calculated the precision and recall of the identification of errors and accurate uses (see Appendix S1 in the online Supporting Information). Because we examined the accuracy at a lower end of the data retrieval process, the accuracy of POS tagging per se is not relevant.

[5]The same standardization procedure was used in the analysis of COCA.

[6]Because the word *lots* was often erroneously tagged as a pronoun in COCA, its frequency was computed irrespective of the part-of-speech tags annotated.

[7]To avoid log-transforming negative values, we added the maximum absolute value of the negative ΔP values plus half of the minimum absolute value of ΔP to all the ΔP values prior to log-transformation.

[8]Murakami and Alexopoulou (2016) also found a strong influence of L1 on progressive *-ing*, a finding inconsistent with the present study. The reason for the discrepancy is unclear.

References

<bib id="bib1" label="Aguado-Orea, 2004" type="Book">Aguado-Orea, J. J. (2004). *The acquisition of morpho-syntax in Spanish: Implications for current theories of development*. (Unpublished doctoral dissertation). University of Nottingham, Nottingham, UK.</bib>

<bib id="bib2" label="Allan, 1980" type="Periodical">Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society, 15*, 147–149. https://doi.org/10.3758/BF03334492</bib>

<bib id="bib3" label="Alexopoulou et al., 2017" type="Other">Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning*, *67*(S1), 180–208. https://doi.org/10.1111/lang.12232</bib>

<bib id="bib4" label="Ambridge et al., 2015" type="Periodical">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*(2), 239–273. https://doi.org/10.1017/S030500091400049X</bib>

<bib id="bib5" label="Andersen and Shirai, 1994" type="Other">Andersen, R. W., & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition*, *16*, 133–156. https://doi.org/10.1017/S0272263100012845</bib>

<bib id="bib6" label="Arnon and Clark, 2011" type="Periodical">Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth: Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7, 107–129. https://doi.org/10.1080/15475441.2010.505489</bib>

<bib id="bib7" label="Bannard and Matthews, 2008" type="Periodical">Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x</bib>

<bib id="bib8" label="Bardovi-Harlig, 2000" type="Book">Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford, UK: Blackwell.</bib>

<bib id="bib9" label="Bardovi-Harlig and Comajoan-Colomé, 2020" type="Other">Bardovi-Harlig, K., & Comajoan-Colomé, L. (2020). The aspect hypothesis and the acquisition of L2 past morphology in the last 20 years: A state of the scholarship review. *Studies in Second Language Acquisition*, *42*, 1137–1167. https://doi.org/10.1017/S0272263120000194.</bib>

<bib id="bib10" label="Barr et al., 2013" type="Periodical">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</bib>

<bib id="bib11" label="Beckner et al., 2009" type="Periodical">Beckner, C., Blythe, R. A., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., Schoenemann, T. (2009). Language is a complex adaptive system. Position paper. *Language Learning*, 59(Supplement 1), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x</bib>

<bib id="bib12" label="Braine, 1976" type="Other">Braine, M. D. S. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, *41*(1), 1--104. https://doi.org/10.2307/1165959</bib>

<bib id="bib13" label="Braine et al., 1990" type="Periodical">Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language 29*, 591–610. https://doi.org/10.1016/0749-596X(90)90054-4</bib>

<bib id="bib14" label="Bürkner, 2017" type="Periodical">Bürkner, P.-C. (2017). brms: An R package for Bayesian generalized linear mixed models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01</bib>

<bib id="bib15" label="Bybee, 1985" type="Book">Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam, The Netherlands: John Benjamins.</bib>

<bib id="bib16" label="Bybee and Hopper, 2001" type="Book">Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.</bib>

<bib id="bib17" label="Bod and Hay, 2003" type="Book">Bod, R., Hay, J., & Jannedy, S. (Eds). (2003). *Probabilistic Linguistics*. Cambridge, MA: MIT Press.</bib>

<bib id="bib18" label="Bowles, 2011" type="Other">Bowles, M. (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</bib>

<bib id="bib19" label="Carpenter et al., 2017" type="Periodical">Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01</bib>

<bib id="bib20" label="Childers and Tomasello, 2001" type="Periodical">Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology 37*, 739–48. https://doi.org/10.1037/0012-1649.37.6.739.</bib>

<bib id="bib21" label="Christiansen and Chater, 2016" type="Periodical">Christiansen, M., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, E62. https://doi.org/10.1017/S0140525×1500031X</bib>

<bib id="bib22" label="Cintrón-Valentín and Ellis, 2016" type="Periodical">Cintrón-Valentín, M., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in Psychology*, *7*, 1284. https://doi.org/10.3389/fpsyg.2016.01284</bib>

<bib id="bib23" label="Comrie, 1976" type="Other">Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.</bib>

<bib id="bib24" type="Book">Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.</bib>

<bib id="bib25" label="Cumming, 1989" type="Other">Cumming, A. (1989). Writing expertise and second-language proficiency. *Language Learning, 39*, 81–135. https://doi.org/10.1111/j.1467-1770.1989.tb00592.x</bib>

<bib id="bib26" type="URL">Davies, M. (2008-). The Corpus of Contemporary American English (COCA). Retrieved from https://www.english-corpora.org/coca/</bib>

<bib id="bib27" label="Dias et al., 1999" type="URL">Dias, G., Guilloré, S., & Pereira Lopes, J. G. (1999). Language independent automatic acquisition of rigid multiword units from unrestricted text corpora. In *Actes de la 6ième Conférence sur le Traitement Automatique des Langues Naturelles* (Proceedings of the 6th Conference on Natural Language Processing). Cargèse, France: Association pour le Traitement Automatique des Langues. Retrieved from http://talnarchives.atala.org/TALN/TALN-1999/taln-1999-poster-005.pdf</bib>

<bib id="bib28" label="de Saussure and Reidlinger, 1916" type="Book">de Saussure, F. (1983). *Course in General Linguistics* (C. Bally & A. Sechehaye, Eds.; A. Reidlinger, Coll.; R. Harris, Trans.). London, UK: Duckworth. (Original work published 1916).</bib>

<bib id="bib29" label="Dowty, 1979" type="Other">Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.</bib>

<bib id="bib30" label="Dulay and Burt, 1973" type="Other">Dulay, H. C., & Burt, M. K. (1973). Should we teach children syntax? *Language Learning*, *23*, 245–258. https://doi.org/10.1111/j.1467-1770.1973.tb00659.x</bib>

<bib id="bib31" label="Dulay and Burt, 1974" type="Other">Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning*, *24*, 37–53. https://doi.org/10.1111/j.1467-1770.1974.tb00234.x</bib>

<bib id="bib32" label="Dunn, 2018" type="Periodical">Dunn, J. (2018). Multi-unit directional measures of association: Moving beyond pairs of words. *International Journal of Corpus Linguistics*, *23*, 183–215. https://doi.org/10.1075/ijcl.16098.dun</bib>

<bib id="bib33" label="Elder and Ellis, 2009" type="Book">Elder, C., & Ellis, R. (2009). Implicit and explicit knowledge of an L2 and language proficiency. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.). *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 167–193). Bristol, UK: Multilingual Matters.</bib>

<bib id="bib34" label="Ellis, 2002" type="Periodical">Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188. https://doi.org/10.1017/S0272263102002024</bib>

<bib id="bib35" label="Ellis, 2006" type="Other">Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24. https://doi.org/10.1093/applin/ami038</bib>

<bib id="bib36" label="Ellis, 2012" type="Periodical">Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17--44.</bib>

<bib id="bib37" label="Ellis, 2015" type="Book">Ellis, N. C. (2015). Implicit *and* explicit learning: Their dynamic interface and complexity. In P. Rebushat (Ed.), *Implicit and explicit learning of languages* (pp. 3–23). Amsterdam, The Netherlands: John Benjamins.</bib>

<bib id="bib38" label="Ellis, 2017" type="Other">Ellis, N. C. (2017). Cognition, corpora, and computing: Triangulating research in usage-based language learning. *Language Learning*, *67* (S1), 40–65. *https://doi.org/10.1111/lang.12215*</bib>

<bib id="bib39" label="Ellis et al., 2014" type="Other">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</bib>

<bib id="bib40" label="Ellis et al., 2016" type="Periodical">Ellis, N. C., Römer, U., & O'Donnell, M. B. (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*, *Language Learning Monograph Series*. Chichester, UK: Wiley.</bib>

<bib id="bib41" label="Ellis and Sagarra, 2011" type="Periodical">Ellis, N. C., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study and meta-analysis. *Studies in Second Language Acquisition*, *33*, 589–624. https://doi.org/10.1017/S0272263111000325</bib>

<bib id="bib42" label="Ellis, 2005" type="Other">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</bib>

<bib id="bib43" label="Ellis et al., 2006" type="Other">Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–368. https://doi.org/10.1017/S0272263106060141</bib>

<bib id="bib44" label="Elman, 1990" type="Periodical">Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1</bib>

<bib id="bib45" label="Erlam, 2006" type="Other">Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464--491. https://doi.org/10.1093/applin/aml001</bib>

<bib id="bib46" label="Erman and Warren, 2000" type="Periodical">Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*, 29–62. https://doi.org/10.1515/text.1.2000.20.1.29</bib>

<bib id="bib47" label="Finley, 2018" type="Periodical">Finley, S. (2018). Cognitive and linguistic biases in morphology learning. *Wiley Interdisciplinary Reviews: Cognitive Science* *9*(5): e1467. https://doi.org/10.1002/wcs.1467</bib>

<bib id="bib48" label="Geertzen et al., 2014" type="Proceeding">Geertzen, J., Alexopoulou, T., & Korhonen, A. (2014). Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). In R. T. Millar, K. I. Martin, C. M. Eddington, A. Henery, N. M. Miguel, A. Tseng, … D. Walter (Eds.), *Selected proceedings of the 2012 Second Language Research Forum. Building bridges between disciplines* (pp. 240–254). Somerville, MA: Cascadilla.</bib>

<bib id="bib49" label="Development Team, 2020" type="URL">Stan Development Team (2020, April 17). Prior choice recommendations [Web log post]. Retrieved from https://web.archive.org/web/20210414053443/https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations</bib>

<bib id="bib50" label="Goldberg, 1995" type="Book">Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument atructure*. Chicago, IL: University of Chicago Press.</bib>

<bib id="bib51" label="Goldberg, 2006" type="Book">Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford, UK: Oxford University Press.</bib>

<bib id="bib52" label="Goldschneider and DeKeyser, 2001" type="Other">Goldschneider, J. M., & DeKeyser, R. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1–50. https://doi.org/10.1111/1467-9922.00147</bib>

<bib id="bib53" label="Gries, 2013" type="Periodical">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</bib>

<bib id="bib54" label="Gries, 2015" type="Periodical">Gries, S. Th. (2015). More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26, 505–536. https://doi.org/10.1515/cog-2014-0092</bib>

<bib id="bib55" label="Gries and Ellis, 2015" type="Periodical">Gries, S. Th., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics: Statistical measures for usage-based linguistics. *Language Learning* 65(S1), 228–255. https://doi.org/10.1111/lang.12119</bib>

<bib id="bib56" label="Gries and Stefanowitsch, 2004" type="Periodical">Gries, S. Th., & A. Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9, 97–129. https://doi.org/10.1075/ijcl.9.1.06gri</bib>

<bib id="bib57" label="Gudmestad et al., 2013" type="Other">Gudmestad, A., House, L., & Geeslin, K. L. (2013). What a Bayesian analysis can do for SLA: New tools for the sociolinguistic study of subject expression in L2 Spanish. *Language Learning*, *63*, 371–399. https://doi.org/10.1111/lang.12006</bib>

<bib id="bib58" label="Guo and Ellis, 2021" type="Periodical">Guo, R., & Ellis, N. C. (2021). Language usage and second language morphosyntax: Effects of availability, reliability, and formulaicity. *Frontiers in Psychology*, *12*, 582259. https://doi.org/10.3389/fpsyg.2021.582259</bib>

<bib id="bib59" label="Hanks, 2013" type="Book">Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. Cambridge, MA: MIT Press.</bib>

<bib id="bib60" label="Harrell, 2015" type="Book">Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis* (2nd ed.). New York, NY: Springer.</bib>

<bib id="bib61" label="Hay, 2001" type="Periodical">Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics*, *39*, 1041–1070. https://doi.org/10.1515/ling.2001.041</bib>

<bib id="bib62" label="Huang et al., 2018" type="Periodical">Huang, Y., Murakami, A., Alexopoulou, T., & Korhonen, A. (2018). Dependency parsing of learner English. *International Journal of Corpus Linguistics*, *23*, 28–54. https://doi.org/10.1075/ijcl.16080.hua</bib>

<bib id="bib63" label="Hulstijn, 2015" type="Other">Hulstijn, J.H. (2015). Discussion: How different can perspectives on L2 development be? *Language Learning*, *65*, 210–232. https://doi.org/10.1111/lang.12096</bib>

<bib id="bib64" label="Hulstijn et al., 2015" type="Other">Hulstijn, J.H., Ellis, R., & Eskildsen, S.W. (2015). Orders and sequences in the acquisition of L2 morphosyntax, 40 years on: An

introduction to the special issue. *Language Learning*, *65*, 1–5. https://doi.org/10.1111/lang.12097</bib>

<bib id="bib65" label="In'nami and Murakami, 2021" type="Book">In'nami, Y., & Murakami, A. (2021). Statistical modelling of SLA theories: Connecting test performance to construct. In P. Winke & T. Brunfaut (Eds.), *The Routledge handbook of second language acquisition and language testing* (pp. 445–456). New York, NY: Routledge.</bib>

<bib id="bib66" label="Jarvis and Pavlenko, 2007" type="Book">Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition*. New York, NY: Routledge.</bib>

<bib id="bib67" label="Jia and Fuse, 2007" type="Periodical">Jia, G., & Fuse, K. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research*, *50*, 1280–1299. https://doi.org/10.1044/1092-4388(2007/090)</bib>

<bib id="bib68" label="Johnson and Newport, 1989" type="Periodical">Johnson, J., & Newport, E. L. (1989). Critical period effects in second language learning and the influence of the maturational state on the acquisition of ESL. *Cognitive Psychology*, *21*, 215–258. https://doi.org/10.1016/0010-0285(89)90003-0</bib>

<bib id="bib69" label="Kruschke, 2014" type="Book">Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London, UK: Academic Press.</bib>

<bib id="bib70" label="Lambert, 2018" type="Book">Lambert, B. (2018). *A student's guide to Bayesian statistics*. Los Angeles, CA: Sage.</bib>

<bib id="bib71" label="Levelt, 1989" type="Book">Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.</bib>

<bib id="bib72" label="MacWhinney, 1987" type="Periodical">MacWhinney, B. (1987). Applying the competition model to bilingualism. *Applied Psycholinguistics*, *8*, 315–327. https://doi.org/10.1017/S0142716400000357</bib>

<bib id="bib73" label="MacWhinney, 1997" type="Book">MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–142). Mahwah, NJ: Lawrence Erlbaum.</bib>

<bib id="bib74" label="MacWhinney, 2008" type="Book">MacWhinney, B. (2008). A unified model. In N. C. Ellis, & P. Robinson (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 341–372). New York: Erlbaum.</bib>

<bib id="bib75" label="MacWhinney and Bates, 1989" type="Book">MacWhinney, B., & Bates, E.(Eds.) (1989). *The crosslinguistic study of sentence processing*. Cambridge, UK: Cambridge University Press.</bib>

<bib id="bib76" label="MacWhinney and O'Grady, 2015" type="Book">MacWhinney, B., & O'Grady, W. (Eds.) (2015). *The handbook of language emergence*. Hoboken, NJ: Wiley-Blackwelle</bib>

<bib id="bib77" label="Marchman, 1997" type="Periodical">Marchman, V. (1997). Children's productivity in the English past tense: The role of frequency, phonology, and neighborhood structure. *Cognitive Science 21*, 283–303. https://doi.org/10.1016/S0364-0213(99)80025-1</bib>

<bib id="bib78" label="Matthews and Theakston, 2006" type="Periodical">Matthews, D. E., & Theakston, A. L. (2006). Errors of omission in English-speaking children's production of plurals and the past tense: The effects of frequency, phonology, and competition. *Cognitive Science*, *30*, 1027–1052. https://doi.org/10.1207/s15516709cog0000_66</bib>

<bib id="bib79" label="McElreath, 2020" type="Book">McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Boca Raton, FL: CRC Press</bib>

<bib id="bib80" label="Michel et al., 2019" type="Periodical">Michel, M., Murakami, A., Alexopoulou, T., & Meurers, D. (2019). Effects of task type on morphosyntactic complexity across proficiency: Evidence from a large learner corpus of A1 to C2 writings. *Instructed Second Language Acquisition*, *3*, 124–152. https://doi.org/10.1558/isla.38248</bib>

<bib id="bib81" label="Mintz, 2003" type="Periodical">Mintz, T. H. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition 90*, 91–117. https://doi.org/10.1016/S0010-0277(03)00140-9.</bib>

<bib id="bib82" label="Mintz et al., 2014" type="Periodical">Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. *Cognitive Psychology*, *75*, 1–27. https://doi.org/10.1016/j.cogpsych.2014.07.003</bib>

<bib id="bib83" label="Mita et al., 2020" type="Book">Mita, M., Kiyono, S., Kaneko, M., Suzuki, J., & Inui, K. (2020). A self-refinement strategy for noise reduction in grammatical error correction. In T. Cohn, Y. He., & Y. Liu (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020* (pp. 267–280). https://doi.org/10.18653/v1/2020.findings-emnlp.26</bib>

<bib id="bib84" label="Monteiro et al., 2020" type="Other">Monteiro, K. R., Crossley, S. A., & Kyle, K. (2020). In search of new benchmarks: Using L2 lexical frequency and contextual diversity indices to assess second language writing. *Applied Linguistics*, *41*, 280–300. https://doi.org/10.1093/applin/amy056</bib>

<bib id="bib85" label="Mudhsh, 2021" type="Periodical">Mudhsh, B. A. D. M. (2021). A comparative study of tense and aspect categories in Arabic and English. *Cogent Arts & Humanities*, *8*(1), 1899568. https://doi.org/10.1080/23311983.2021.1899568</bib>

<bib id="bib86" label="Murakami, 2014" type="Book">Murakami, A. (2014). *Individual variation and the role of L1 in the L2 development of English grammatical morphemes: Insights from learner corpora* (Unpublished doctoral dissertation). University of Cambridge, Cambridge, UK</bib>

<bib id="bib87" label="Murakami, 2016" type="Other">Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, *66*, 834–871. https://doi.org/10.1111/lang.12166</bib>

<bib id="bib88" label="Murakami and Alexopoulou, 2016" type="Other">Murakami, A., & Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes: A learner corpus study. *Studies in Second Language Acquisition*, *38*, 365–401. https://doi.org/10.1017/S0272263115000352</bib>

<bib id="bib89" label="Murakami and Ellis, 2022" type="Dataset">Murakami, A., & Ellis, N. C. (2022). *Corpus data and R code. Datasets from "Effects of availability, contingency, and formulaicity on the accuracy of English grammatical morphemes in second language writing"* . IRIS Database, University of York, UK. https://doi.org/10.48316/9amf-cz93 </bib>

<bib id="bib90" label="Nicholls, 2003" type="Proceeding">Nicholls, D. (2003). The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference*, *Vol. 16* (pp. 572–581). Lancaster, UK: Lancaster University.</bib>

<bib id="bib91" label="Norouzian et al., 2018" type="Other">Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, *68*, 1032–1075. https://doi.org/10.1111/lang.12310</bib>

<bib id="bib92" label="O'Grady, 2015" type="Other">O'Grady, W. (2015). Processing determinism. *Language Learning*, *65*, 6–32. https://doi.org/10.1111/lang.12091</bib>

<bib id="bib93" label="Pawley and Syder, 1983" type="Book">Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–225). London, UK: Longman.</bib>

<bib id="bib94" label="Core Team, 2020" type="Book">R Core Team. (2020). R: A language and environment for statistical computing (Version 4.0.0) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. www.r-project.org.</bib>

<bib id="bib95" label="Redington et al., 1998" type="Periodical">Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*, 425–469. https://doi.org/10.1207/s15516709cog2204_2</bib>

<bib id="bib96" label="Robinson, 2008" type="Book">Robinson, P., & Ellis, N. C. (Eds.) (2008). *Handbook of cognitive linguistics and second language acquisition*. London, UK: Routledge.</bib>

<bib id="bib97" label="Ryding, 2014" type="Book">Ryding, K. (2014). *Arabic: A linguistic introduction*. Cambridge, UK: Cambridge University Press</bib>

<bib id="bib98" label="Saito, 2020" type="Other">Saito, K. (2020). Multi- or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*, *70*, 548–588. https://doi.org/10.1111/lang.12387</bib>

<bib id="bib99" label="Schad et al., 2021" type="Periodical">Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*, 103–126. https://doi.org/10.1037/met0000275</bib>

<bib id="bib100" label="Schmid, 1994" type="Proceeding">Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing* (pp. 44–49). Manchester, UK: University of Manchester Institute of Science and Technology, Centre for Computational Linguistics.</bib>

<bib id="bib101" label="Schmidt, 1984" type="Periodical">Schmidt, R. (1984). The strengths and limitations of acquisition: A case study of an untutored language learner. *Language, Learning, and Communication*, *3*, 1–16.</bib>

<bib id="bib102" label="Shanks, 1995" type="Book">Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge, UK: Cambridge University Press.</bib>

<bib id="bib103" label="Shatz, 2020" type="Periodical">Shatz, I. (2020). Refining and modifying the EFCAMDAT: Lessons from creating a new corpus from an existing large-scale English learner language database. *International Journal of Learner Corpus Research*, 6, 220–236. https://doi.org/10.1075/ijlcr.20009.sha</bib>

<bib id="bib104" type="Book">Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2018). Understanding formulaic language: A second language acquisition perspective. London, UK: Routledge. https://doi.org/10.4324/9781315206615</bib>

<bib id="bib105" label="Siyanova-Chanturia and Spina, 2020" type="Other">Siyanova-Chanturia, A. & Spina, S. (2020). Multi-word expressions in second language writing: A large-scale longitudinal learner corpus study. *Language Learning*, *70*, 420–463. https://doi.org/10.1111/lang.12383</bib>

<bib id="bib106" label="Spada and Shiu, 2015" type="Other">Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*, 723–751. https://doi.org/10.1111/lang.12129</bib>

<bib id="bib107" label="Stefanowitsch and Gries, 2003" type="Periodical">Stefanowitsch, A., & Gries, S. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8, 209–43. https://doi.org/10.1075/ijcl.8.2.03ste</bib>

<bib id="bib108" label="Sugaya and Shirai, 2009" type="Book">Sugaya, N., & Shirai, Y. (2009). Can L2 learners productively use Japanese tense-aspect markers? A usage-based approach. In R.

Corrigan, E. A. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Typological studies in language*, (pp. 383--423). Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/tsl.83.10sug</bib>

<bib id="bib109" label="Suzuki and DeKeyser, 2015" type="Other">Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*, 860–895. https://doi.org/10.1111/lang.12138</bib>

<bib id="bib110" label="Tatsumi et al., 2018" type="Periodical">Tatsumi, T., Ambridge, B., & Pine, J. M. (2018). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production study of Japanese. *Cognitive Science*, *42*, 555–77. https://doi.org/10.1111/cogs.12554</bib>

<bib id="bib111" label="Tavakoli and Uchihara, 2020" type="Other">Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning, 70*, 506–547. https://doi.org/10.1111/lang.12384</bib>

<bib id="bib112" label="Tremblay et al., 2011" type="Other">Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*, 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x</bib>

<bib id="bib113" label="Trousdale and Hoffmann, 2013" type="Book">Trousdale, G., & Hoffmann, T.(Eds). (2013). *The Oxford handbook of construction grammar*. Oxford, UK: Oxford University Press.</bib>

<bib id="bib114" label="Wolter and Yamashita, 2018" type="Other">Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, *40*, 395–416. https://doi.org/10.1017/S0272263117000237</bib>

<bib id="bib115" label="Wulff and Ellis, 2018" type="Book">Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Studies in Bilingualism*, *54*, 37–56. Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/sibil.54.03wul</bib>

<A>Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Details and Accuracy Evaluation of Data Retrieval.

Appendix S2. Calculation of ΔP.

Appendix S3. Model Validation.

Appendix S4. Model Diagnostics.

Appendix S5. Numerical Summary of Posterior Distributions.

<A>Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

<B>Distributional characteristics of accurate second language English inflectional morpheme use

<C>*What This Research Was About And Why It Is Important*

<TXT>

Usage-based theories hold that (second) language acquisition is influenced by the distributional characteristics of learners' input including such factors as frequency and formulaicity. Our study empirically tested this claim by examining the accuracy of four inflectional morphemes (past tense *-ed*, progressive *-ing*, third person *-s*, plural *-s*) in a large-scale corpus of learner writing. The study demonstrated that second language learners use the morphemes more accurately in reliable forms, i.e. when the frequency of the inflected word form (e.g., *arrived*) relative to the frequency of the lemma of the form (e.g., *arrive*) is high. Contrary to a previous study that investigated the issue experimentally in laboratory online processing tasks, however, our study did not consistently identify the effects of the frequency of the inflected form itself or the formulaicity of the words surrounding the inflected form on morpheme accuracy.

<C>*What the Researchers Did*

<TXT>

- We drew data from a large-scale corpus of learner writing (EF-Cambridge Open Language Database [EFCAMDAT]).

- Based on the coding of learners' errors (available in the corpus), we identified both the errors and accurate uses of each target morpheme in each sample of writing.
- The information about how these morphemes and lemmas are typically distributed in the language as a whole (e.g., frequency of the inflected form, formulaicity of the context in which the inflection occurred) was calculated based on a large-scale corpus of American English (Corpus of Contemporary American English).

<C>*What the Researchers Found*

<TXT>

- We found a consistent positive association between reliability (i.e., the frequency of the inflected form relative to the frequency of the lemma) and morpheme accuracy.
- This suggests that the lemma functions as a cue for its inflected forms and that learners use (tally) such an association in processing inflectional morphemes when they read or hear input.
- On the other hand, the study did not find a consistent pattern between the raw frequency of inflected forms and morpheme accuracy or between the formulaicity of the context and morpheme accuracy.
- The association between the distributional characteristics and morpheme accuracy did not differ much across learners' proficiency levels or their longitudinal development.

<C>*Things to Consider*

<TXT>

- Both our study and the previous experimental study on the same topic found a consistent effect of reliability, indicating the robust effect of lemma-morpheme association strength on morpheme accuracy.
- This robust effect can be interpreted as evidence for the phenomenon known as associative learning, a theory of learning that can explain (at least some parts of) second language acquisition.
- Contrary to the previous experimental study, our study did not find the effect of the raw frequency of the inflected form or formulaicity on morpheme accuracy.
- This difference is possibly due to methodological differences between the experimental task (elicited imitation, demanding online processing of predetermined stimuli) and the more conscious, written compositions in EFCAMDAT analyzed for the current study, in which learners decide which language they need, and so the language may be less prone to effects of formulaicity or the frequency of inflected forms.

Materials, data, open access article: Analysis code and data are publicly available at IRIS (www.iris-database.org) and OSF (https://osf.io/ba8mf/).

How to cite this summary: Murakami, A., & Ellis, N. C. (2022). Distributional characteristics of accurate second language English inflectional morpheme use. *OASIS Summary* of Murakami and Ellis (2022) in *Language Learning*. https://oasis-database.org

Table 1 Numbers of learners, writings, word types, and obligatory contexts for each morpheme

Morpheme	Learners	Writings	Word types	Obligatory contexts
Past tense *-ed*	30,955	83,001	87	151,979
Progressive *-ing*	39,744	88,010	94	123,869
Third person *-s*	39,961	78,667	124	113,298
Plural *-s*	2,633	59,759	89	136,601

Table 2 Moving-window calculation of ΔP

N-gram	L4	L3	L2	L1	Node	R1	R2	R3	R4	ΔP	Standardized log-transformed ΔP
3-gram											
			with	all	kinds					0.01	
				all	kinds	of				<u>0.19</u>	1.51

					kinds	of	people			0.01	
4-gram											
		deal	with	all	kinds					0.03	
			with	all	kinds	of				0.33	
				all	kinds	of	people			0.54	1.61
					kinds	of	people	and		0.02	
5-gram											
	can	deal	with	all	kinds					0.05	
		deal	with	all	kinds	of				0.38	
			with	all	kinds	of	people			0.43	
				all	kinds	of	people	and		0.49	1.54
					kinds	of	people	and	i	0.02	

Note. Underling in the ΔP column indicates the largest ΔP value for each n-gram. Underlining in the rightmost column indicates the largest standardized log-transformed ΔP across the three n-grams.

Table 3 First-language type of each morpheme in each nationality group

Nationality	Past tense *-ed*	Progressive *-ing*	Third person *-s*	Plural *-s*
Brazil	1	1	1	1

China / Taiwan	0	0	0	0
France	1	0	1	1
Germany	1	0	1	1
Italy	1	1	1	1
Japan	1	1	0	0
Mexico	1	1	1	1
Russia	1	0	1	1
Saudi Arabia	1	0	1	1
Turkey	1	1	0	1

Note. The assumed L1 of each nationality group was: Brazil = Brazilian Portuguese; China and Taiwan = Mandarin Chinese; France = French; Germany = German; Italy = Italian; Japan = Japanese; Mexico = Spanish; Russia = Russian; Saudi Arabia = Arabic; and Turkey = Turkish. 1 indicates that an equivalent feature is obligatorily marked in the L1 (referred to as the PRESENT group), and 0 indicates otherwise (ABSENT group). For the references supporting the coding of Arabic, see Mudhsh (2021) for progressive *-ing* and Ryding (2014) for the other morphemes. For the remaining nationality groups, see Murakami (2014, Table 21) and the references therein.

Figure 1 Morpheme accuracy across predictor values. The trend lines in quantitative variables are the fitted values of the frequentist binary logistic regression models predicting accuracy as a function of the predictor in concern. The shaded areas are their 95% confidence intervals. The minimum value of the horizontal axis of the maximum standardized log-transformed ΔP was set to –3 to allow the inspection of densest areas. Further, only the word sequences that occurred 100 or more times were included in the panels of ΔP. In the panels of L1 type, small dots represent the mean accuracy of individual nationality groups, and what looks like larger dots represents the 95% confidence intervals of the overall mean accuracy.

Figure 2 Posterior distribution of fixed-effects parameters. The points are the mean posterior values and the error bars represent 95% credible intervals. All the quantitative predictor variables were standardized to a mean of 0 and a standard deviation of 1.

Figure 3 Interaction of the mean Englishtown level and frequency in third person *-s.* Each CEFR level is represented by the middle Englishtown level of the three levels that constitute the CEFR level. For instance, A1 is represented by Englishtown Level 2 because Englishtown Levels 1 through 3 correspond to A1. The line in each panel represents the posterior mean, and the shaded area is its 95% credible interval. The accuracy of individual third person *-s* forms is observed values.

Figure 4 Illustration of the matrix of de Saussure's thought-sound relationship. Adapted from de Saussure (1916/1983, p. 111).