



# Language Usage and Second Language Morphosyntax: Effects of Availability, Reliability, and Formulaicity

Rundi Guo\* and Nick C. Ellis

Language Learning Laboratory, Department of Psychology, University of Michigan, Ann Arbor, MI, United States

## OPEN ACCESS

### Edited by:

Brian MacWhinney,  
Carnegie Mellon University,  
United States

### Reviewed by:

Patricia J. Brooks,  
College of Staten Island,  
United States  
Helen Zhao,  
The University of Melbourne, Australia

### \*Correspondence:

Rundi Guo  
guorundi@umich.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Psychology

**Received:** 11 July 2020

**Accepted:** 23 March 2021

**Published:** 29 April 2021

### Citation:

Guo R and Ellis NC (2021)  
Language Usage and Second  
Language Morphosyntax: Effects  
of Availability, Reliability,  
and Formulaicity.  
Front. Psychol. 12:582259.  
doi: 10.3389/fpsyg.2021.582259

A large body of psycholinguistic research demonstrates that both language processing and language acquisition are sensitive to the distributions of linguistic constructions in usage. Here we investigate how statistical distributions at different linguistic levels – morphological and lexical (Experiments 1 and 2), and phrasal (Experiment 2) – contribute to the ease with which morphosyntax is processed and produced by second language learners. We analyze Chinese ESL learners' knowledge of four English inflectional morphemes: *-ed*, *-ing*, and third-person *-s* on verbs, and plural *-s* on nouns. In Elicited Imitation Tasks, participants listened to length- and difficulty-matched sentences each containing one target morpheme and typed the whole sentence as accurately as they could after a short delay. Experiment 1 investigated lexical and morphemic levels, testing the hypotheses that a morpheme is expected to be more easily processed when it is (1) highly *available* (i.e., occurring in frequent word-forms), and (2) highly *reliable* (i.e., occurring in lemma words that are consistently conjugated in the form containing this morpheme). Thirty sentences were made for each morpheme, divided into three Availability-Reliability Distribution (ARD) groups on the basis of corpus analysis in the Corpus of Contemporary American English (COCA; Davies, 2008-): 10 target words high in availability, 10 high in reliability, and 10 low in both reliability and availability. Responses were scored on whether the target morpheme was accurately reproduced given the provision of the correct lemma. A generalized linear mixed-effects logit model (GLMM) revealed fixed effects of morpheme type, availability, and reliability on the accuracy of morpheme provision. There were no effects of lemma frequency. Experiment 2 successfully replicated these results and extended the investigation to explore phrasal formulaicity by manipulating the frequency of the four-word strings in which the morpheme was embedded. GLMMs replicated the effects of word-form availability and reliability and additionally revealed independent *phrase-superiority effects* where morphemes were better reproduced in contexts of higher string-frequency. Taken together, these findings demonstrate that morpheme acquisition reflects the distributional properties of learners' experience and the mappings

therein between lexis, morphology, phraseology, and semantics. These conclusions support an emergentist view of the statistical symbolic learning of morphology where language acquisition involves the satisfaction of competing constraints across multiple grain-sizes of units.

**Keywords:** SLA, morphosyntax, availability, reliability, formulaicity, phrase-superiority effects, elicited imitation task

## INTRODUCTION

Novice and intermediate learners of English as a second language (ESL) are far from consistent in their production of inflectional morphemes, such as regular past-tense *-ed*, or third person singular present-tense *-s*. Jia and Fuse (2007) show that the acquisition of a morpheme such as the third-person singular *-s* can take 5 years or more to go from 0 to 80% provision in obligatory contexts for ESL children. Five years of English usage involves many thousands of receptive experiences of high frequency functional morphemes, and many thousands of contexts requiring their productive use, yet provision is variable. This suggests that the system is learned incrementally, and that regularities/generalization/productivity emerge from the combined experience of usage. But is it the case that, for any given morpheme, some exemplars are more easily recognized in the input and produced earlier in acquisition? If so, what are these exemplars that are more likely to be preferentially processed? And why these bellwethers? Are they special in their distributional statistics, for example, in terms of their frequency, or their form-function contingency, or their formulaicity? These are the questions that motivate our research here. How does second language (L2) morphological ability depend upon usage?

Usage-based theories hold that domain-general cognitive mechanisms drive the learning of linguistic constructions and the emergence of generalizations (e.g., Goldberg, 2006; Beckner et al., 2009; MacWhinney and O'Grady, 2014; Wulff and Ellis, 2018). They proposed that acquisition is modulated by factors affecting attention and memory, such as exemplar type- and token-frequency, contingency of form-function mapping, salience of form and of function, paradigm complexity, neighborhood effects and the proportion of friends to enemies<sup>1</sup> in quasi-regular domains, etc. (e.g., Marchman, 1997; MacWhinney, 2001; Bybee, 2006; Ellis, 2006a,b; Seidenberg and Plaut, 2014). For the case of morphology, we might ask then, in the 5 years during which L2 learners are learning to produce third-person singular *-s*, do experiences of particular *-s* inflected verbs play a role in the acquisition of the system more than others? Likewise, for the even more extended period during which L2 learners are learning to produce regular past-tense *-ed*, are particular *-ed* inflected verbs more potent exemplars than others? And so on. From studies of children (Brown, 1973; Braine et al., 1990; Ambridge et al., 2015; Finley, 2018) and of adults (e.g., Seidenberg and Plaut, 2014; Pollatsek et al., 2015), there is good reason to

suspect that distributional factors affect L1 and L2 morpheme acquisition and processing.

Linguistic constructions vary in frequency and they distribute across usage in complex probabilistic patterns. Psycholinguistics research has established several important aspects of these distribution patterns. The most studied parameter is *availability*, which concerns how often a language learner experiences a given form in their usage history. Availability is estimated as the normalized token frequency of a specific word-form in representative corpora. For example, the availability of the word-form *depends* is the overall probability of encountering the word-form *depends* in English usage, i.e.,  $P(\textit{depends})$ . The effects of availability on the development and processing of L1 and L2 has been well-established. For example, words high in frequency are named faster (Forster and Chambers, 1973; Seidenberg and McClelland, 1989), judged faster in lexical decision tasks (Yap and Balota, 2014), fixated for shorter durations in reading (McDonald and Shillcock, 2003), recognized more easily in speech (Luce, 1986), and spelled more accurately (Barry and Seymour, 1988). More generally, language learners are sensitive to the frequency of linguistic cues across a wide range of linguistic domains and levels of representation, including phonology and phonotactics, lexis, reading and spelling, morphosyntax, sentence comprehension, etc. (Bybee and Hopper, 2001; Ellis, 2002; Bod et al., 2003).

For the particular case of morphology, words inflected in a form that is high in token frequency are produced earlier and more accurately in that form compared to in other forms and compared to other words that are inflected in low token frequency forms. Such token frequency effects of word-forms have been reported in the acquisition of L1 (Marchman, 1997; Ambridge et al., 2015; Räsänen et al., 2016), L2 (Larsen-Freeman, 1976; Goldschneider and DeKeyser, 2001; Jia and Fuse, 2007), and artificial grammars (Braine et al., 1990; Finley, 2015). Notably, the frequency of word lemmas plays a lesser role in the accurate retrieval of inflected word-forms as compared to the token frequency of the inflected word-form itself - a key finding that has important implications for emergentist approaches which posit chunk-based learning from usage, construction grammar, and linguistic structure as processing history.

Another important distribution parameter is *reliability*, i.e., how likely it is that a linguistic cue reliably co-occurs either with another construction, or with a particular interpretation. Measuring reliability entails the statistical estimation of some form of contingency (MacWhinney, 2001; Ellis, 2006a; Gries and Ellis, 2015). In the context of morpheme acquisition, reliability can be understood as the relative frequency of different word-forms of a lemma, for example, the reliability of the lemma [depend] occurring in its *-s* morpheme inflected form *depends*

<sup>1</sup> Friends share mappings from stem to tense (e.g. *walk-walked*, *talk-talked*) and pull together; enemies use different mappings (e.g., *throw-threw*) and pull apart (Marchman et al., 1999).

can be calculated as the number of occurrences of the word-form *depends* divided by the number of occurrences of all possible word-forms of the lemma [depend] such as *depend*, *depending*, and *depended*, i.e.,  $P(\textit{depends} \mid \textit{depend})$ . To the native ear, *depends* might well sound more natural than *depended*, perhaps due to the fact that in an English-speaking environment, when the word *depend* is used, it is most often conjugated in its third-person singular form. In other words, the high reliability of *depends* might facilitate its processing in this form, regardless of the overall frequency of occurrence of *depends* in the entire environment. As a result, *depend* might become implicitly more associated with the morpheme *-s* and with the present than with the other tenses. Psychological research into animal and human learning alike demonstrates profound and ubiquitous impacts of contingency in the learning of cue-outcome associations (Shanks, 1995).

The relative frequency of different morphological forms of the same word have been found to predict usage, language change, accuracy, and error patterns in language processing and acquisition (Bybee, 1985; Hay, 2001; Matthews and Theakston, 2006; Sugaya and Shirai, 2009; Tatsumi et al., 2018). Hay's (2001) study of relative frequency in derivational morphology, which follows proposals on the structure of paradigms (groups of inflectionally related words with a common lexical stem) in Bybee (1985, Chapter 3), demonstrates that the more frequent member of a paradigm is more accessible and less compositional. Paradigms consist of words of different frequency/accessibility levels, the high frequency words are dominant, and the others are dependent upon them. In studies of language change, paradigms are more likely to be re-made on the basis of the highest frequency form (Bybee, 1985). In studies of L1 acquisition, when acquiring irregular plural forms, English speaking children tend to erroneously produce phrases like *\*two mouse* much more frequently than phrases like *\*two tooth*, likely because *mouse* is a much more reliable form for the lemma [mouse] than *tooth* is for [tooth]: the word-form *mouse* occurs seven times more than the word-form *mice*, whereas the word-form *tooth* occurs only one sixth as often as does *teeth* (Matthews and Theakston, 2006).

The third distribution parameter to be considered here is *formulaicity*, i.e., the frequency of the multi-word strings in which a morpheme-inflected word-form is embedded<sup>2</sup>. Consider how you might more naturally end the phrase "you've got to be . . ." with "kidding" than with "playing." According to the Corpus of Contemporary American English (Davies, 2008-), the multi-word string "you've got to be kidding" is more frequently used than "you've got to be playing"; i.e.,  $P(\textit{you've got to be kidding}) > P(\textit{you've got to be playing})$ . Note that this is an effect of string frequency, since  $P(\textit{kidding}) < P(\textit{playing})$ .

High-frequency phrases, idioms, and formulaic sequences (Sinclair, 1995; Wray, 2002) are processed more fluently than matched low-frequency strings. For example, in phrasal decision

tasks (Bod, 2001; Jiang and Nekrasova, 2007; Arnon and Snider, 2010), high frequency phrasal constituents or short sentences (e.g., *don't have to worry*; *I like it*) are judged to be grammatical phrases faster than less frequent controls composed of frequency-matched component words (*don't have to wait*; *I keep it*). Formulaicity effects have likewise been demonstrated in L1 acquisition by Bannard and Matthews (2008) who showed that 2-3-year-old English speaking children were quicker and more accurate at repeating frequently occurring multi-word strings (e.g., "sit in your chair") sampled from a large child-directed speech corpus, compared to matched infrequent strings ("sit in your truck"). High-frequency slot-and-frame patterns (Braine, 1976) or frames (Mintz, 2003) can strongly constrain the nature of the slot-filler, e.g., a frame like "to \_ it" is highly predictive of verbs<sup>3</sup>. Such distributional information can be potent in the acquisition of both the grammatical and the semantic properties of the slot-filler (Elman, 1990; Redington et al., 1998). Mintz et al. (2014) compared training situations in which target words (such as *lowfa*) occurred surrounded by two-word frames (such as *swetch\_klide*) that frequently co-occurred, against situations in which target words occurred in simpler bigram contexts (such as *swetch\_lowfa* or *lowfa\_klide*) where only an immediately adjacent word provides the context for categorization). They found that learners categorized words together when they occurred in similar frame contexts, but not when they occurred in similar bigram contexts. In a study of L1 English-speaking 2 1/2-year-olds, Childers and Tomasello (2001) found that a nonce verb was better acquired so to be subsequently used creatively in a transitive utterance when it was surrounded by pronouns than when surrounded by proper nouns or names, suggesting that the child's transitive schema may start out with pronouns in pre-/post-verbal positions (i.e., pronoun V pronoun) rather than being fully general. In other words, frequent formulaic frames can positively promote the processing and productivity of their subcomponent words.

Together, these studies demonstrate that the three distributional factors of *availability*, *reliability*, and *formulaicity* pervasively affect language acquisition and processing. In the current study, we are concerned with their roles in L2 morphology, and whether particular exemplars are more easily recognized in the input and correctly produced because of their privileged distributions in the language.

We examine L2 knowledge of four common English inflectional morphemes (the regular past-tense ending *-ed*, the progressive marker *-ing*, the third-person singular present-tense ending *-s*, and the nominal plural marker *-s*). We target ESL learners whose native language is Mandarin Chinese because this population has been shown to experience greater challenges in acquiring L2 inflections due to the fact that Mandarin Chinese has minimal verb-tense and noun morphology (Yeh et al., 2015). None of the four English morphemes included in our study has a direct morphological equivalence in Mandarin Chinese, although some of them can be expressed with non-inflectional grammatical

<sup>2</sup>There are alternative operationalizations available for formulaicity involving mutual information or other measures of conditional probability rather than string frequency, and these various measures can affect acquisition and processing in different ways (e.g., Ellis, 2012b). Here we adopt simple string frequency because it is the most basic and widely used metric for formulaic language, multi-word expressions, and lexical bundles (Wray, 2002; Biber, 2004; Gries and Ellis, 2015).

<sup>3</sup>Searching for this frame in COCA returns more than 1600 verb types, in decreasing order: *do*, *make*, *get*, *see*, *keep*, *take*, *put*, *say*, . . . following a Zipfian frequency distribution).

cues, e.g., certain classifiers that can express plurality, certain aspectual markers (e.g., V-le) that arguably possess properties of a tense marker (Ross, 1995), and lexical cues such as numbers and adverbs. We aim to assess how much ESL morpheme processing and production depends upon their English usage distributions.

Experiment 1 investigates distributions at lexical and morphological levels. Experiment 2 extends the study to include the effects of the distributions of larger phraseological constructions on the processing of embedded morphemes.

## EXPERIMENT 1

Experiment 1 investigates the effects of availability and reliability of the word-forms containing target morphemes on the production accuracy of the target morphemes. We hypothesize that a morpheme is more easily processed (1) when it occurs in a word-form that is highly frequent in usage (i.e., highly available), and (2) when it is attached to a word that is more consistently conjugated in the form containing this morpheme compared to other forms of the same word lemma (i.e., highly reliable).

Many studies of morpheme acquisition, following Brown (1973), assess spontaneous production of target morphemes in obligatory contexts, i.e., where the morpheme would be obligatory in a native-speaking adult's speech either because of the pragmatic context of discourse (e.g., describing something happened in the past calls for use of past-tense verbs) or the syntactic structure of the utterance (e.g., "Yesterday, I walk\_\_ to the store" requires the regular past-tense morpheme *-ed*). Here, instead, we use an Elicited Imitation Task (EIT) with morphemes in non-obligatory contexts of the sort used by Marchman (1997) who investigated the production of the past-tense *-ed* morpheme in L1 English-speaking children.

### Elicited Imitation Tasks

Elicited Imitation Tasks have been widely used in studies of L2 processing and have been shown to have high validity and reliability (Ortega et al., 2002; Erlam, 2006; Gaillard and Tremblay, 2016). In one version of EIT, the participant hears a sentence and is asked to repeat the exact sentence after a short delay. Unlike production in uncontrolled spontaneous speech (e.g., Jia and Fuse, 2007), EIT allows controlled examination of morpheme production in contexts that are matched in important respects, thus isolating the effects of the property of the morphemes themselves from that of their context (Erlam, 2006). The use of predetermined sentence stimuli allows the control of important potential confounds such as the presence of an adverbial tense cues, the frequency of the strings of words that contain the morpheme, grammatical complexity, memory load, etc.

For present purposes, we modified the EIT design to require participants to 'repeat' the sentence by typing the written form rather than speaking the oral form. This modification circumvents accent-induced transcription ambiguity, facilitates data collection and analysis, and is less threatening to our Chinese ESL participants who reportedly experience considerable discrepancy between their proficiency in spoken and written

forms of English as a result of classroom pedagogical practices in China, which commonly deemphasize oral English instruction (Ren, 2011).

## A Process Analysis of EIT as It Relates to Morpheme Production

Each of the 120 randomized trials of our EIT involved listening to a single sentence out of context and then, after a short delay during which the participant rates it for how sensible it is, repeating it verbatim, in all of its parts, as accurately as possible. What processes might be involved in the successful repetition of a target morpheme in such a task? The following sketch is informed by proposals in usage-based linguistics, construction grammar, the psycholinguistics of sentence processing, predictive processing, and first and second language acquisition [see, particularly, Christiansen and Chater (2016) on "Chunk-and-Pass" processing, and, more generally for review of language emergence, MacWhinney and O'Grady, 2014]:

The perception and comprehension encoding stages in EIT involve three parts: (1) taking in word-forms into an auditory/lexical buffer, (2) linking lexical items syntactically, and (3) constructing a meaningful interpretation of the sentence. Based on the psycholinguistic research which has shown a variety of frequency effects in the perception and processing of words, morphemes, multi-word chunks, and syntactic constructions, we propose that the initial recognition and preservation of the correct form of target words is likely influenced by the forces of availability, reliability, and formulaicity in terms of storage in an auditory/lexical buffer. Then, the language system rapidly integrates all available incoming information, interactively satisfying multiple 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 contextual information to recognize words and morphemes, which 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 potential influences of syntactic integrity, e.g., auxiliary [be] impacting particularly progressive *-ing*, and of contextual support where context could influence the encoding of the past *-ed*. The encoding of third person present *-s* and plural *-s* on subjects should also be under the influence of syntactic integrity, although in English, agreement processing is generally less obligatory than processing for tense and aspect (MacWhinney, 1997, 2001). The final stage of EIT, (4) production, is also expected to be sensitive to frequency effects and sequential probabilities at word, morpheme, and particularly phrasal levels: a well-entrenched formulaic phrase will support provision of its component morphemes whether they are analyzed or not.

A relevant process analysis of the imitative written production of a recently heard message 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, accessing their orthographic codes, and creating and executing motor programs for skilled typing. In such imitative redintegration, we might expect probabilistic effects to be at their strongest. Formulaic language is more common in speech than in writing (Erman and 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.

## Method

### Participants

Participants were Chinese native speakers ( $n = 22$ ) who were international students at a major university in the United States. They were either sampled from the Subject Pool of the Psychology Department and participated for course credit ( $n = 1$ ) or recruited through recruitment posters around the campus and paid \$15 for their participation ( $n = 21$ ). The sample were unintentionally female-dominant ( $n = 18$ ). Participants were between 18 and 28 years old (Mean = 22,  $SD = 2.64$ ). All but one of them had lived in an English immersed environment for some time<sup>4</sup>. Excluding this participant, the length of residence in an English-speaking country was between 6 and 84 months (Mean = 31.33,  $SD = 21.69$ ). All participants had a high-level English proficiency sufficient to permit them to follow the English instructions and complete the language task entirely in English. Their proficiency in English was assessed by self-ratings and self-reported TOEFL scores. One participant was excluded from analysis due to excessive missing data. Summary of participant characteristics is reported in Section 1.1 of **Supplementary Data Sheet 1**.

### Materials

#### *Elicited imitation task*

We followed the EIT design features outlined by Ortega et al. (2002). All sentences were within the recommended syllable length of 10–17 syllables to ensure optimal difficulty; the words that contain the target morpheme were placed in the middle of sentences, with filler words at the very beginning and ends of the sentences to reduce primacy and recency effects. So, in the stimulus sentence “*Late Wednesday evening I thanked him for the lovely flowers,*” the target morpheme is the *-ed* in *thanked* in the middle of the sentence, the controlled four-word context was *I thanked him for*, the primacy filler was a randomly selected three word phrase *Late Wednesday evening*, and the recency filler was *the lovely flowers*. More detail on how these sentences were constructed is described in the following section. To reduce the

impact of phonological rehearsal in short term memory, a 3–5 s distraction task was set up in-between the stimulus and response for each sentence, during which the participants had to judge whether the sentence seemed sensible to them, thus reducing the opportunity for rehearsal. This semantic judgment task helped to ensure that participants actively engaged in semantic processing of the sentences rather than simply trying to encode and retain their acoustic forms. The following section describes the procedure for how the sentences were created.

#### *Item development*

The study targeted four of the most studied inflectional morphemes in English verbs and nouns: the regular past-tense ending *-ed*, the progressive marker *-ing*, the third-person singular present-tense ending *-s*, and nominal plural marker *-s*. Thirty sentences were made for each morpheme, which were divided into three Availability-Reliability Distribution (ARD) groups on the basis of corpus analysis in the Corpus of Contemporary American English (COCA) (Davies, 2008)<sup>5</sup>. COCA is widely used and frequently updated and contains over 520 million words (with 20 million words added each year from 1990 to 2015) coming from 220 thousand text sources that are equally divided among different genres of American English such as spoken, news and magazines, academic texts, fiction, etc., making it the only large and balanced corpus of English used in the United States.

The ARD groups for each morpheme were determined on the basis of their carrier words. The three groups are: (1) Top 10 Availability, (2) Top 10 Reliability, and (3) Bottom 10 Reliability. We first assessed the lemma frequency and the inflected word-form frequency by conducting searches for the top 1000 most frequent content verbs ( $[vv^*]$ ) and the top 1000 content nouns ( $[nn^*]$ ) and recording the frequency counts in Excel. The Top 10 Availability group consists of the top 10 most frequent inflected word-forms exemplifying each of the four target morphemes. The search commands were as follows: regular past-tense verbs (*-ed*:  $*ed.[vv^*d]$ ), third-person singular present-tense verbs (*-s*:  $*s.[vv^*z]$ ), progressive verbs (*-ing*:  $[vv^*g]$ ), and regular plural nouns (*-s*:  $*s.[*nn2^*]$ )<sup>6</sup>. Morpheme reliability was operationalized as the proportion of times that the lemma occurred with that specific morpheme by dividing the word-form frequency (i.e., the frequency of the word-form inflected with the target morpheme) by the lemma frequency (i.e., frequency of all possible word-forms of the lemma). For each morpheme, we ranked the items by reliability and took the top 10 of these to form the Top 10 Reliability group, unless the item had already been included in the Top 10 Availability group. Lastly, the Bottom 10 Reliability group were the items lowest in reliability of expression of the embedded morpheme. It was formed from the bottom results of the proportion rankings that were also relatively low in word-form frequency. Where there was room for choice between exemplars, we favored the alternative with the highest lemma frequency. We also tried to match the Top 10 Reliability and Bottom 10 Reliability items for word-form frequency. The frequency and

<sup>5</sup><https://corpus.byu.edu/coca/>

<sup>6</sup>More information regarding the syntax can be found at the BYU corpus portal ([corpus.byu.edu](https://corpus.byu.edu)) and from the “insert POS” tab in search page of COCA (<https://corpus-byu-edu.proxy.lib.umich.edu/coca/>)

<sup>4</sup>A visiting student who had just arrived in the US for the first time was included in the study because of her demonstrable fluency in English and the fact that she had been receiving part-time English-immersed education in China.

**TABLE 1** | The mean lemma frequency, inflected word-form frequency (availability), and word-form:lemma proportion (reliability) of the carrier words in the stimulus sample by ARD group and by morpheme.

Group and morpheme	Lemma frequency (Mean)	Inflected word-form frequency (Mean)	Reliability <sup>1</sup> (Proportion) (Mean)
Top 10 Availability	744806	121345	0.32
Past-tense -ed	372980	96811	0.29
Third-person -s	884702	86873	0.13
Progressive -ing	1462384	139036	0.20
Plural -s	259157	162659	0.66
Top 10 Reliability	28908	20042	0.64
Past-tense -ed	16206	10967	0.68
Third-person -s	29185	11750	0.45
Progressive -ing	12426	6597	0.56
Plural -s	57813	50856	0.88
Bottom 10 Reliability	815824	16912	0.03
Past-tense -ed	64125	1615	0.03
Third-person -s	97507	1929	0.02
Progressive -ing	2904224	53365	0.02
Plural -s	197441	10741	0.06

<sup>1</sup> Reliability = Word-form frequency/Lemma frequency.

reliability characteristics of the stimulus sample are summarized in **Table 1**. **Figure 1** illustrates the distribution of the stimuli belonging to each group for each morpheme within the top 1000 frequent content verbs or nouns accordingly. We show the Top 10 Availability group in blue and illustrate with the leading exemplar (e.g., *students* for plural -s), the Top 10 Reliability group in green (*participants*), and the Bottom 10 Reliability group in red (*gods*).

To build the sentence contexts for these carrier words, we first conducted n-gram searches for possible three-word strings with the target word in the middle (e.g., [*\* wanted \**]) and then selected the top frequent results for each. These results (e.g., [*you wanted to*]) were then fed into searches for possible four-word strings with an extra slot at the end (i.e., [*you wanted to \**]). Then, a three-word random time filler phrase (e.g., *On Wednesday morning. . .*) was put at the beginning of each sentence. We only included tense-neutral time phrases — those that do not provide any lexical time cue — so that the target inflectional morpheme would be the only indicator of the tense, i.e., the morphemes would be in non-obligatory contexts. The sentences were then completed to a length of 14–17 syllables by adding a random possible phrase at the end that would make the whole sentence grammatical, logical, and relatively sensible. All filler words were checked with a lexical range breakdown using the computer program VocabProfile from the Compleat Lexical Tutor website (Cobb, accessed 8/2017) so they are roughly in the same frequency band, mostly the top 1000 frequent words. All the finalized sentences were manually checked by two native Chinese speakers to make sure they were sufficiently comprehensible for an average Chinese ESL

learner. All sentences were recorded in Audacity<sup>7</sup>. Each of them was spoken twice by a male native speaker of American English and was evaluated by another native speaker to select the best version. Sample sentences can be found in **Table 2**. The complete stimuli set including the full list of words with their lemma frequency, word-form frequency, and the calculated reliability along with their carrier sentences is available in **Supplementary Data Sheet 2**.

## Procedure

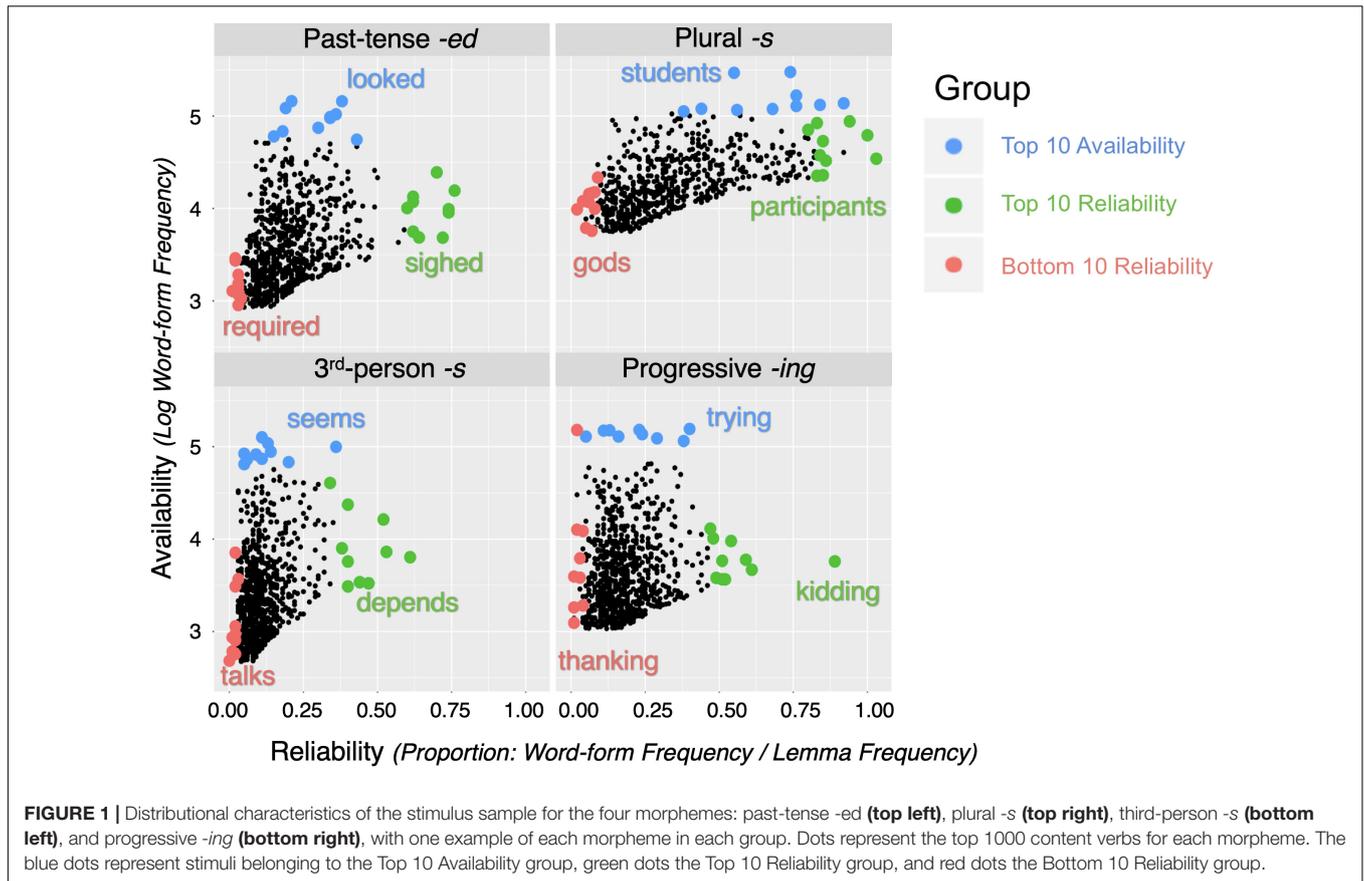
The EIT task was administered individually through a PsychoPy program (PsychoPy, RRID:SCR\_006571, Peirce, 2007) running on an iMac computer equipped with headphones located in an experimental booth. The total duration of the experiment was approximately 80 min. After providing informed consent, participants received brief oral instructions from the experimenter. The program began with an instruction screen that explained how each sentence would be presented and what their task was, followed by a practice session of five sentences. Participants proceeded to the experimental trials if no further questions arose. The experimenter remained available to aid them as needed.

All participants listened to the 120 sentences. The presentation sequence was individually randomized. On each trial, they first heard a spoken sentence, such as “*Late Wednesday evening I thanked him for the lovely flowers.*” Immediately after the audio ended, the screen displayed instructions for the participant to judge how much sense the sentence made to them by rating it on a sliding scale of 1–7 using the mouse. Once this rating had been completed, the participant was asked to type out the complete and correct sentence to the best of their ability. Participants decided when the next trial should start by pressing the spacebar. They were notified at the midpoint of the experiment and allowed to take a short break if desired. Their reproduction of the sentences was recorded in csv files.

After the experiment, participants completed a 5-min language history questionnaire (**Supplementary Data Sheet 1, 2.2**) adapted from Lim and Godfroid (2015). This included questions on general demographics such as gender and age, as well as language background including previous and current exposure and usage of English, English proficiency test scores, self-rated general proficiency in English, and self-rated proficiency on different aspects of using English (reading, writing, speaking, and listening).

Elicited imitation responses were scored for accuracy of production of the one target word that contained the target morpheme in each sentence in the following steps: First, using a string search command in Excel, we screened whether each response contained the exact match of the target word (marked as 1) or not (marked as 0). “Exact matches” were also automatically marked 1 for “correct lemma” and 1 for “correct morpheme.” Second, we manually checked responses marked as 0s for “exact match” looking for typos and spelling errors, as well as irregularities in the inflectional paradigms of certain words, to decide whether a reasonable

<sup>7</sup><https://www.audacityteam.org/>



**FIGURE 1 |** Distributional characteristics of the stimulus sample for the four morphemes: past-tense *-ed* (top left), plural *-s* (top right), third-person *-s* (bottom left), and progressive *-ing* (bottom right), with one example of each morpheme in each group. Dots represent the top 1000 content verbs for each morpheme. The blue dots represent stimuli belonging to the Top 10 Availability group, green dots the Top 10 Reliability group, and red dots the Bottom 10 Reliability group.

**TABLE 2 |** Sample sentences for selected morpheme in each group for Experiment 1.

Morpheme frequency group	Target morpheme	Sentence	Word-form frequency	Lemma frequency	Proportion (reliability)
Top 10 availability	<i>-ed</i>	On Monday afternoon he <b>looked</b> at me for a long moment	121996	652141	0.19
Top 10 reliability	<i>-ing</i>	Late Thursday evening I was <b>kidding</b> about his hair and beard	5734	6410	0.89
Bottom 10 reliability	3sg <i>-s</i>	On Friday afternoon he <b>talks</b> about the latest developments	7125	304560	0.02
Bottom 10 reliability	Plural <i>-s</i>	On Wednesday she hears about the <b>gods</b> of the new religion	6156	125937	0.05

Target word in each sentence is bolded.

*attempt* at the target word was present. In cases where the attempted target word reasonably resembled any form of the lemma (e.g., *\*glansed* for *glanced*), they were given a “correct lemma” score of 1. Likewise, if its ending reasonably resembled the target morpheme (e.g., *\*lookign* for *looking*), or if its form resembled the tense or number indicated by the target morpheme (e.g., *\*drooling* for *drilling*), they were given a “correct morpheme” score of 1. Finally, using Excel commands again, we identified whether the correct morpheme was present in the target word *given* the presence of a correct lemma (“correct morpheme *given* correct lemma”). To sum up, each typed response was either marked as 1 (for “correct morpheme *given* correct lemma”), 0 (for “incorrect or absent morpheme *given* correct lemma”), or N/A (for cases where the lemma is absent). The lemma-absent cases constituted 8.91% of the responses and were excluded from further analysis. The

scoring method is illustrated in **Table 3** with two examples of each scenario.

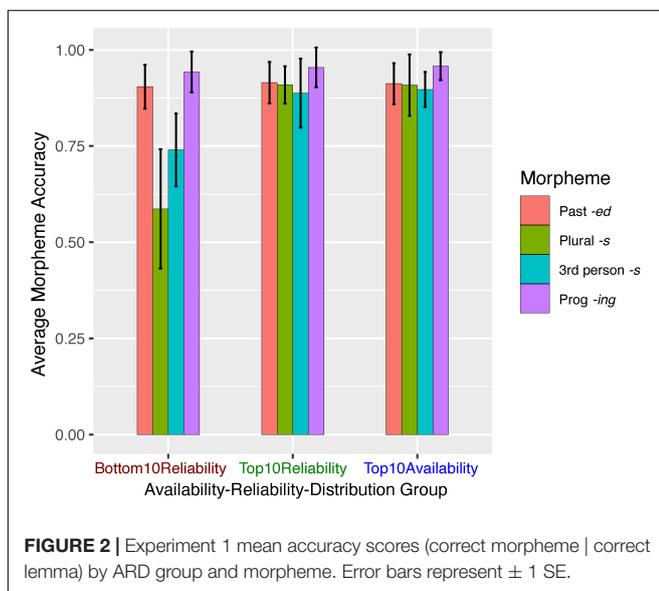
### Results

The accuracy scores of sentences for each morpheme in each ARD group are shown in **Figure 2**. To examine the effects of the two distributional factors, availability and reliability, on the production accuracy of the morphemes, we used generalized linear mixed-effect models using the “lme4” package (R package: lme4, RRID:SCR\_015654, version 1.1-13, Bates et al., 2015) in R (version 3.3.3, R Core Team, 2017). The models were fit by maximum likelihood (Laplace Approximation), with random effects specified for subjects and items. Because the four target morphemes are inherently stratified in frequency in the corpus, e.g., past-tense *-ed* verbs are generally used more frequently than third-person *-s* verbs, the distributional factors

**TABLE 3** | Scoring method with sample sentences.

<b>Correct morpheme given correct lemma</b>	<b>Score: 1</b>
Examples: (1a) Late Wednesday evening, I <b>thanked</b> him for the lovely flowers. (2a) On Thursday, she knew about the <b>fathers</b> of the fat children.	
<b>Incorrect or absent morpheme given correct lemma</b>	<b>Score: 0</b>
Examples: (1b) Late Wednesday evening, I <b>thanks</b> him for the lovely flowers. (2b) On Thursday, she knew about the <b>father</b> of the fat children.	
<b>Incorrect or absent lemma</b>	<b>Score: N/A (excluded)</b>
Examples: (1c) Late Wednesday evening, I <b>think</b> him for the lovely flowers. (2c) On Thursday, she knew about the <b>mothers</b> of the fat children.	

Target word in each sentence is bolded.

**FIGURE 2** | Experiment 1 mean accuracy scores (correct morpheme | correct lemma) by ARD group and morpheme. Error bars represent  $\pm 1$  SE.

are correlated with morpheme type. To reduce multicollinearity and to account for the between-morpheme differences, we first ran a mixed-effect model with morpheme type as the only fixed-effect predictor to serve as a baseline model which parses out the differences between the four morphemes. From there, we built up the model by incrementally specifying other predictors one at a time to identify the unique contributions of each. To determine which subject-level random effects to include, specifically whether or not to include the random slopes for subjects for each fixed-effect predictor, we ran two versions for each model, one with only random intercepts, and one further adding random slopes. We report the model with random intercepts for subjects unless adding random slopes significantly improves model fit, in which case we report the latter. The preliminary steps (testing morpheme type alone; morpheme type + morpheme reliability; and morpheme type + morpheme availability) are detailed in **Supplementary Data Sheet 1, 1.2**. We describe here the complete Model 1 involving all three fixed effects.

### Model 1: Morpheme Type + Morpheme Reliability + Morpheme Availability

Model 1, which included morpheme type, reliability, availability (i.e., log word-form frequency) as fixed-effect predictors, and random intercepts for subjects and items), is detailed in **Table 4**. Stimulus sentence length in syllables was included as a fixed predictor to control for any stimulus length effects. Each participant's stimulus sense rating for each sentence was also included as a potential predictor.

There were effects of morpheme type: *-ing* had significantly higher accuracy than *-ed* (estimate = 0.963,  $SE = 0.435$ ,  $z = 2.21$ ,  $p = 0.027$ ); Plural *-s* had significantly lower accuracy than *-ed* (estimate =  $-1.480$ ,  $SE = 0.424$ ,  $z = -3.49$ ,  $p = 0.000$ ). The difference between the third person present tense *-s* and *-ed* was not significant (estimate =  $-0.347$ ,  $SE = 0.392$ ,  $z = -0.88$ ,  $p = 0.37$ ). The effect of morpheme reliability was highly significant (estimate = 1.937,  $SE = 0.537$ ,  $z = 3.61$ ,  $p = 0.000$ ). Additionally, availability had significant but smaller effects (estimate = 0.419,  $SE = 0.201$ ,  $z = 2.08$ ,  $p = 0.037$ ). Stimulus sentence length was non-significant (estimate = 0.035,  $SE = 0.207$ ,  $z = 0.17$ ,  $p = 0.866$ ). Stimulus sense rating was also non-significant (estimate =  $-0.009$ ,  $SE = 0.059$ ,  $z = -0.16$ ,  $p = 0.874$ ).

Analysis of Deviance using Type III Wald chi-square tests showed that morpheme type, reliability, and availability were all significant predictors of accuracy [morpheme type:  $\chi^2(df = 3) = 31.595$ ,  $p = 0.000$ ; reliability:  $\chi^2(df = 1) = 13.015$ ,  $p = 0.000$ ; availability:  $\chi^2(df = 1) = 4.332$ ,  $p = 0.04$ ], confirming their individual unique contributions to production accuracy. Stimulus sentence length was not a significant predictor  $\chi^2(df = 1) = 0.055$ ,  $p = 0.814$ , nor was stimulus sense  $\chi^2(df = 1) = 0.025$ ,  $p = 0.874$ . **Figure 3** separately plots the effects of Morpheme (3a), reliability (3b), and availability (3c).

Model 1b (mid panel of **Table 4**) investigated the interaction between morpheme type and reliability. Past-tense *-ed* was the reference level for type. Allowing for the interaction removes any overall effect of reliability (estimate = 0.554,  $SE = 1.026$ ,  $z = -0.54$ ,  $p = 0.589$ ). However, there remains a significant effect of reliability on Plural *-s* (estimate = 3.902,  $SE = 1.231$ ,  $z = 3.17$ ,  $p = 0.002$ ) and a marginal one on third person present tense *-s* (estimate = 2.952,  $SE = 1.638$ ,  $z = 1.80$ ,  $p = 0.072$ ).

Model 1c (lower panel of **Table 4**) investigated the interaction between morpheme type and availability, again with past-tense *-ed* as the reference level. Allowing for the interaction removes any overall effect of availability (estimate = 0.221,  $SE = 0.355$ ,  $z = 0.62$ ,  $p = 0.534$ ); although there remains a substantial effect of availability upon Plural *-s* (estimate = 1.943,  $SE = 0.604$ ,  $z = 3.22$ ,  $p = 0.001$ ).

### Exploring Log Lemma Frequency

To examine any effects of lemma frequency (rather than the frequency of the inflected form) alongside morpheme type, we ran a model which included morpheme type and log lemma frequency as fixed-effect predictors and random intercepts for subjects and items. Morpheme type showed consistent effects: *-ing* had significantly higher accuracy than *-ed* (estimate = 1.031,  $SE = 0.481$ ,  $z = 2.14$ ,  $p = 0.03$ ); Plural *-s* had significantly lower accuracy than *-ed* (estimate =  $-0.875$ ,  $SE = 0.431$ ,  $z = -2.03$ ,

**TABLE 4 |** Experiment 1 results from the mixed effects model including fixed effects of morpheme type, morpheme reliability (proportion), morpheme availability (log word-form frequency), stimulus sentence length, stimulus sense rating, and random effects of subject and item.

**Model 1: no interactions.**

Parameters	Fixed effects				Random effects		
	Estimate	SE	z	p	By subject	By item	
					SD	SD	
Intercept	0.145	3.424	0.04	0.966	1.04	1.24	
Morpheme type <sup>1</sup>	Plural -s	-1.480	0.424	-3.49	0.000***		
	Third-person -s	-0.347	0.391	-0.88	0.376		
	Progressive -ing	0.963	0.435	2.21	0.027*		
Morpheme reliability	1.937	0.536	3.61	0.000***			
Morpheme availability <sup>2</sup>	0.420	0.202	2.08	0.037*			
Sentence length (syllables)	0.035	0.207	0.17	0.866			
Sense rating	-0.009	0.059	-0.16	0.874			

<sup>1</sup>Past-tense -ed is the reference level. <sup>2</sup>Word-form frequency was logarithmically transformed. Model formula: accuracy~morpheme + reliability + availability + length + sense + (1| subject) + (1| item). \*\*\*p = 0; \*p < 0.01; p < 0.1.

**Model 1b: Interactions between morpheme type and reliability.**

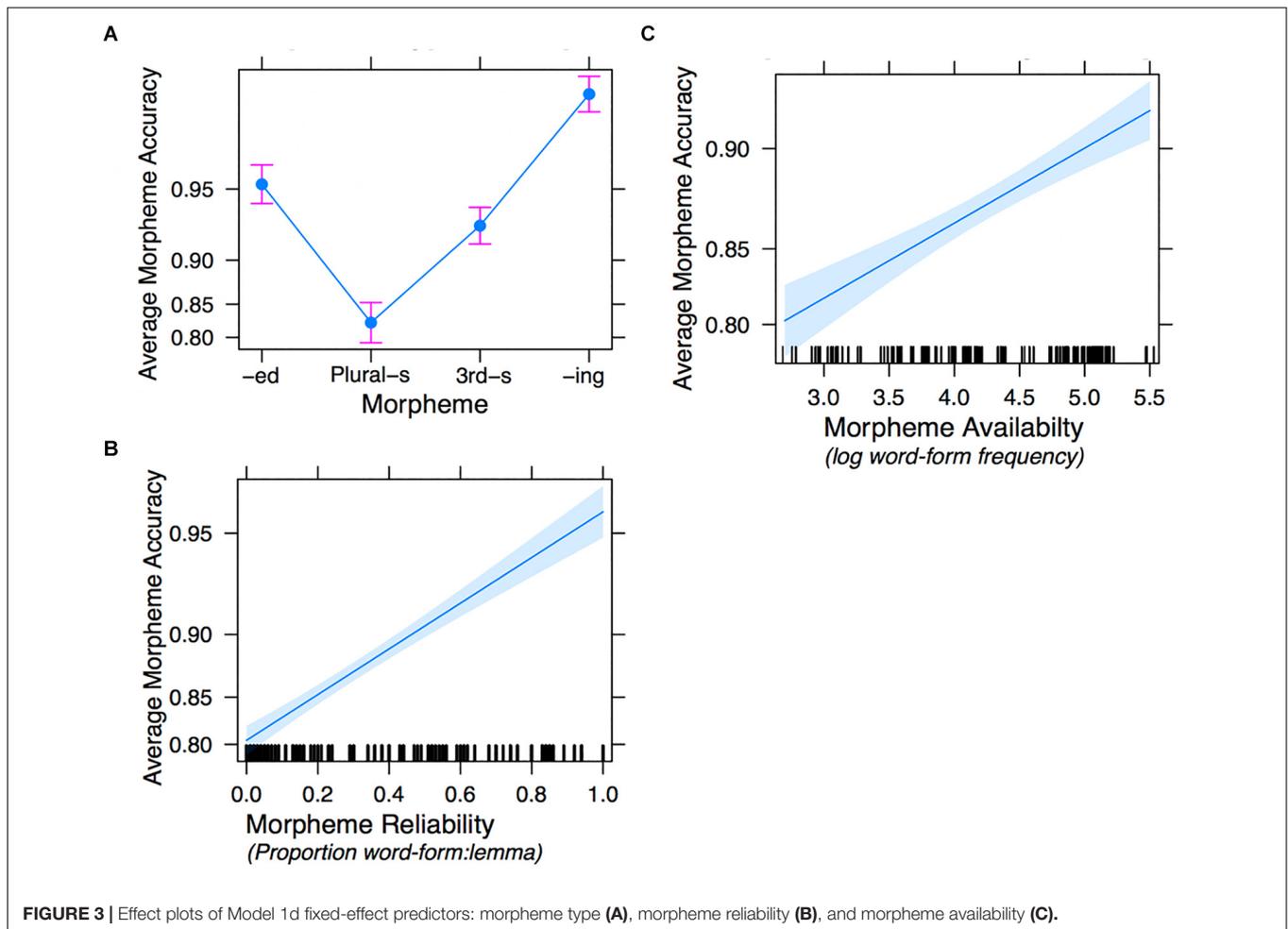
Parameters	Fixed effects				Random effects		
	Estimate	SE	z	p	By subject	By item	
					SD	SD	
Intercept	3.491	2.940	1.19	0.235	1.10	1.11	
Morpheme type <sup>1</sup>	Plural -s	-2.667	0.593	-4.50	0.000***		
	Third-person -s	-1.257	0.536	-2.35	0.019*		
	Progressive -ing	0.534	0.594	0.90	0.368		
Morpheme reliability	0.554	1.026	-0.54	0.589			
Sentence length (syllables)	-0.028	0.192	-0.14	0.886			
MorphemePlural-s: Reliability	3.902	1.231	3.17	0.002**			
MorphemePres-s: Reliability	2.952	1.638	1.80	0.072.			
MorphemeProg-ing: Reliability	1.626	1.647	0.99	0.323			

<sup>1</sup>Past-tense -ed is the reference level. Model formula: accuracy~morpheme\*reliability + length + (1| subject) + (1| item). \*\*\*p = 0; \*\*p < 0.001; \*p < 0.01; p < 0.05; p < 0.1.

**Model 1c: interactions between morpheme type and availability.**

Parameters	Fixed effects				Random effects		
	Estimate	SE	z	p	By Subject	By Item	
					SD	SD	
Intercept	1.765	3.546	0.50	0.619	1.10	1.15	
Morpheme type <sup>1</sup>	Plural -s	-9.827	2.673	-3.68	0.000***		
	Third-person -s	-2.660	1.869	-1.42	0.155		
	Progressive -ing	1.779	2.246	0.79	0.428		
Morpheme availability <sup>2</sup>	0.221	0.355	0.62	0.534			
Sentence length (syllables)	0.017	0.200	0.08	0.934			
MorphemePlural-s: Availability	1.943	0.604	3.22	0.001**			
MorphemePres-s: Availability	0.517	0.458	1.13	0.259			
MorphemeProg-ing: Availability	-0.201	0.527	-0.38	0.703			

<sup>1</sup>Past-tense -ed is the reference level. <sup>2</sup>Word-form frequency was logarithmically transformed. Model formula: accuracy~morpheme\*availability + length + (1| subject) + (1| item). \*\*\*p = 0; \*\*p < 0.001; p < 0.1.



**FIGURE 3 |** Effect plots of Model 1d fixed-effect predictors: morpheme type (A), morpheme reliability (B), and morpheme availability (C).

$p = 0.04$ ). However, the effect of log lemma frequency was negligible (estimate =  $-0.007$ ,  $SE = 0.223$ ,  $z = -0.03$ ,  $p = 0.98$ ). Analysis of Deviance using Type III Wald chi-square tests confirmed the morpheme effect [ $\chi^2(df = 3) = 21.231$ ,  $p = 0.000$ ], and revealed that lemma frequency was not a significant predictor of accuracy when morpheme type was taken into account [ $\chi^2(df = 1) = 0.001$ ,  $p = 0.97$ ].

### Post hoc Explorations of Effects of n-Gram Frequency

As a *post hoc* analysis to explore potential effects of phrasal frequency, we investigated whether log frequency of the three-word string (e.g., *you wanted to*, see section “Item Development”) explained significant additional variance alongside morpheme type. Morpheme type showed consistent effects: *-ing* had significantly higher accuracy than *-ed* (estimate =  $1.131$ ,  $SE = 0.445$ ,  $z = 2.54$ ,  $p = 0.01$ ); Plural *-s* had significantly lower accuracy than *-ed* (estimate =  $-0.980$ ,  $SE = 0.401$ ,  $z = -2.44$ ,  $p = 0.02$ ). The effect of log 3-gram frequency was also highly significant (estimate =  $0.526$ ,  $SE = 0.165$ ,  $z = 3.18$ ,  $p = 0.001$ ). Analysis of Deviance using Type III Wald chi-square tests showed that in addition to morpheme type, log 3-gram frequency was also significantly predictive of accuracy

[morpheme type:  $\chi^2(df = 3) = 26.66$ ,  $p = 0.000$ ; log 3-gram frequency:  $\chi^2(df = 1) = 10.14$ ,  $p = 0.001$ ].

To try to see whether availability (i.e., log word-form frequency) or log 3-gram frequency had independent effects, and which was the greater contributor, we tried models which included both as potential contributors. However, because log word-form frequency and log 3-gram frequency were inherently highly correlated ( $r = 0.810$ ), they pull against each other and neither ends up as significant: availability (estimate =  $0.025$ ,  $SE = 0.345$ ,  $z = 0.07$ ,  $p = 0.941$ ), log 3-gram frequency (estimate =  $0.425$ ,  $SE = 0.274$ ,  $z = 1.55$ ,  $p = 0.122$ ). This is to be further investigated in Experiment 2.

### Results Summary

In sum, these analyses revealed independent effects on production accuracy of morpheme availability and reliability. The interactions of these factors with morpheme type revealed a significant effect of reliability on Plural *-s* and a marginal effect of third person present tense *-s*, and significant effects of availability on Plural *-s*. In contrast to availability of the inflected form, there were no effects of log lemma frequency. Neither sentence length nor sense rating had any effect on morpheme provision. Post-hoc exploratory analyses showed that

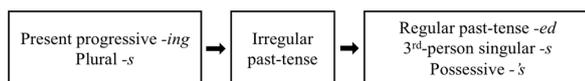
the frequency of the three-word strings also positively predicted accurate provision of the embedded morpheme. However, we had not planned this analysis and had not systematically manipulated the 3-gram frequency in the stimulus materials or controlled for the inherently high correlation between the frequency of the three-word string and the frequency of the word-form inside the string. More careful controls of string frequency are therefore needed to confirm this tentative conclusion.

## Discussion of Experiment 1

As predicted, both availability and reliability of the morphemes were positively associated with morpheme production accuracy in the EIT. A morpheme (e.g., plural-*s*) in a word-form (e.g., *participant-s*) is more easily recognized and produced when the word-form is high in token frequency and when it is the more reliable form of the lemma ([PARTICIPANT]). The effects of reliability were numerically greater than those of availability.

The participants showed a greater sensitivity to the distribution of the morphologically complex surface forms of the words than to the distribution of the underlying lemmas. This finding supports those of Bybee (1985) and Hay (2001) on the importance of relative frequency in derivational morphology described in the introduction. Similar patterns were also observed in Sereno and Jongman's (1997) lexical decision task, in which words were presented in singular (e.g., *car*), or in plural (*cars*) to native speakers. It was found that the difference in reaction times were predicted only by how frequent the specific surface form was presented, whether singular or plural, but not by the total frequency of both forms (i.e., the lemma frequency). Sereno and Jongman took this as evidence against rule-based processing models of inflectional morphology.

The rank order difficulty of the target morphemes (*-ing* > *-ed* > third person present tense *-s* > plural-*s*) was generally consistent with the common order reported in prior SLA morpheme studies (Krashen et al., 1977; Goldschneider and DeKeyser, 2001):



with the exception of the plural *-s*, which was previously reported to be among the earliest to be acquired and processed by L1 and L2 learners of English (Brown, 1973; Krashen et al., 1977). Due to the limited sample size, we refrain from further interpreting this pattern unless it is replicated in Experiment 2. Note also that our stimuli involve a systematically factored selection of 30 exemplars of each type rather than a representationally random sample as used in previous studies, and this might have led to the deviation from the common order.

*Post hoc* exploratory analyses involving the three-word string suggested that frequency beyond the lexical level could also have affected the production accuracy of the embedded morpheme. As previously discussed, facilitation effects of string frequency (formulaicity) have been observed in the processing of phrasal expressions and non-phrasal “lexical bundles” (Arnon

and Snider, 2010; Tremblay et al., 2011), and high frequency frames can facilitate the acquisition and processing of individual component words (Childers and Tomasello, 2001). However, formulaicity research has primarily focused on the facilitation effects on the processing and acquisition of lexical items (Ellis, 2012b; Siyanova-Chanturia and Pellicer-Sanchez, 2018) rather than morphology. The demonstration of effects of formulaicity upon L2 morpheme processing requires more formal control and investigation in a design with greater power than the post-hoc explorations we report here – hence Experiment 2.

## EXPERIMENT 2

Here we aimed to replicate Experiment 1's findings on the pattern of morpheme acquisition, and the facilitation effects of morpheme availability and reliability, with a larger sample of participants, with improved stimulus materials, and with a new speaker for the stimulus recordings. Importantly, it extended to the investigation of the effects on morpheme processing and production of frequency at a phrasal level, i.e., the frequency of the four-word strings that contained the target morpheme. To achieve this, we included the same morphemes as those in Experiment 1 but embedded them in high- and low- frequency four-word strings in the sentences for elicited imitation. Motivated by existing literature on formulaicity and the preliminary results from Experiment 1, we predicted that besides the frequency of the word-forms inflected with the target morpheme, the frequency of the four-word strings in which the morpheme-carrying word-form are embedded would also positively predict the morpheme production accuracy in the elicited imitation of sentences.

## Method

### Participants

Forty-nine native Mandarin Chinese speakers who did not participate in Experiment 1 were recruited for Experiment 2. They were sampled from the same population as the participants in Experiment 1 and were recruited with the same poster. They were paid \$15 for their participation. Data from four participants were excluded from the analysis due to computer malfunction ( $N = 3$ ) or incompleteness of the task ( $N = 1$ ). The remaining 45 participants were predominantly female ( $N = 29$ ), and were between 18 and 28 years old (Mean = 22.02,  $SD = 3.20$ ). All of them have been exposed to an English immersion environment. The length of residence in an English-speaking country was between 0.5 and 192 months (Mean = 44.28,  $SD = 48.15$ ). All participants were sufficiently proficient to complete the task in English. Their English proficiency was assessed by self-ratings and self-reported TOEFL scores using the questionnaire of Experiment 1. Participant characteristics for Experiment 2 are reported in Section 1.3 of **Supplementary Data Sheet 1**.

### Materials

The materials used in the EIT were similar to those used in Experiment 1. All the morphemes and target words remained the same. Experiment 2 had a 4 (morphemes) \* 3 (morpheme

ARD groups) \* 2 (string frequency groups) design. The string frequency grouping factor was operationalized using the COCA-based frequency of the four-word string in which the target morpheme embedded. For example, for sentence 1a in **Table 3**, where “*thanked*” is the target word for the target morpheme *-ed*, the four-word string is “*I thanked him for*,” which occurred 56 times in COCA. Thus, the string frequency for this item is 56. The 120 stimulus sentences in Experiment 1 had been selected for high 4-gram frequency (i.e., the frequency of the four-word string) by design. Thus, we included the Experiment 1 sentences as the *a priori* high string frequency group<sup>8</sup>.

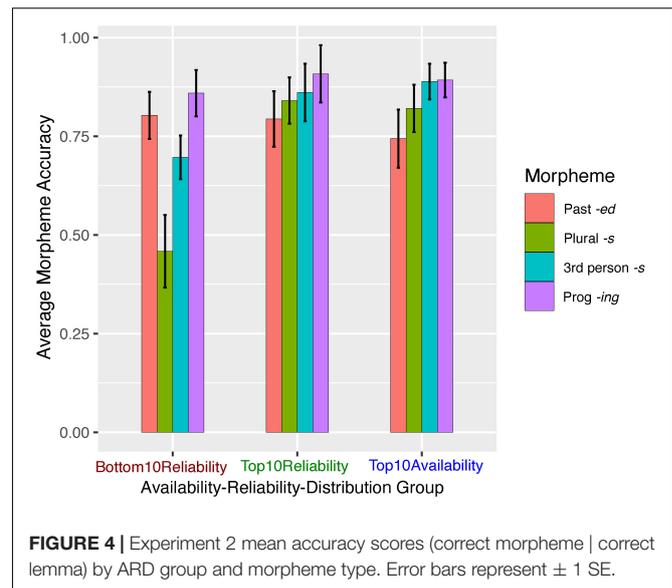
To form the low string frequency group, 120 additional sentences were created by modifying the four-word strings that contain the target morphemes in the 120 existing ones. For the three verbal morphemes (*-ed*, *-ing*, and third-person *-s*), we adopted a modified version of the manipulation of Childers and Tomasello (2001) and constructed the low string frequency version by substituting any pronouns in the high frequency string with person names, e.g., changing “*I thanked him for*” into “*Ashley thanked Steven for*.” The resulting low string frequency sentences contained either one or two person names depending on the transitivity of the target verb. All person names were highly familiar<sup>9</sup> and were between 1 and 4 syllables long. Nevertheless, the resulting low string frequency 4-grams typically do not occur in COCA, resulting in a 4-gram frequency of 0.

For the sentences containing the noun plural *-s*, in which the high frequency four-word strings did not typically contain pronouns, the low string frequency versions were constructed by inserting a familiar adjective before the target word that contained the morpheme. In the example sentence 2a given in **Table 3**, in which “*the fathers of the*” was the high frequency four-word string containing the plural *-s*, we inserted the adjective “*real*” before “*fathers*” so that the low string frequency four-word string became “*real fathers of the*.” We made sure that the adjective-noun pairs were high in bigram frequency to avoid any syntactic violations and/or semantic oddity. We also made sure that all the inserted adjectives were highly frequent words themselves and were all 1–2 syllables long so that the two versions of the sentences were comparable in length. On average, the low string frequency sentences were 1.175 syllables longer than their high frequency counterparts.

All 240 sentences were between 14 and 17 syllables. They were recorded by a female native speaker of American English in a noise-proof recording booth. In order to avoid familiarity effects, we made two counterbalanced versions of the stimuli: the first had odd items (1, 3, 5, ...) of high frequency and even items (2, 4, 6, ...) of low frequency, and vice versa for the second version. In this way, subjects never encountered both versions of a matched pair (For full stimuli list, see **Supplementary Data Sheet 2**).

<sup>8</sup>Most of the Experiment 1 sentences remained exactly the same except 24 sentences. The four-word string of 13 sentences were changed to include at least one animate pronoun (e.g., changing “*it seemed to be*” to “*he seemed to be*”). Eleven sentences were slightly shortened so that both the high formulaicity and the low formulaicity versions were between 14 and 17 syllables long.

<sup>9</sup>All the person names were selected from the top 100 female and top 100 male names used in the United States over the last 100 years according to the Social Security census results (available at <https://www.ssa.gov/oact/babynames/decades/century.html>).



**FIGURE 4** | Experiment 2 mean accuracy scores (correct morpheme | correct lemma) by ARD group and morpheme type. Error bars represent  $\pm 1$  SE.

## Procedure

The procedure remained the same as in Experiment 1. The total duration of Experiment 2 was approximately 80 min. For each participant, the 120 sentences were presented in an individually randomized sequence. The two versions were interspersed in participants to ensure equal number of participants completing each version. The typed responses were recorded in .csv files for analysis. The scoring method also remained the same: correct morpheme given the presence of a correct lemma was marked as 1; incorrect or absent morpheme given correct lemma was marked as 0; lemma-absent cases were marked as N/A and were excluded from further analysis (17.19%).

## Results

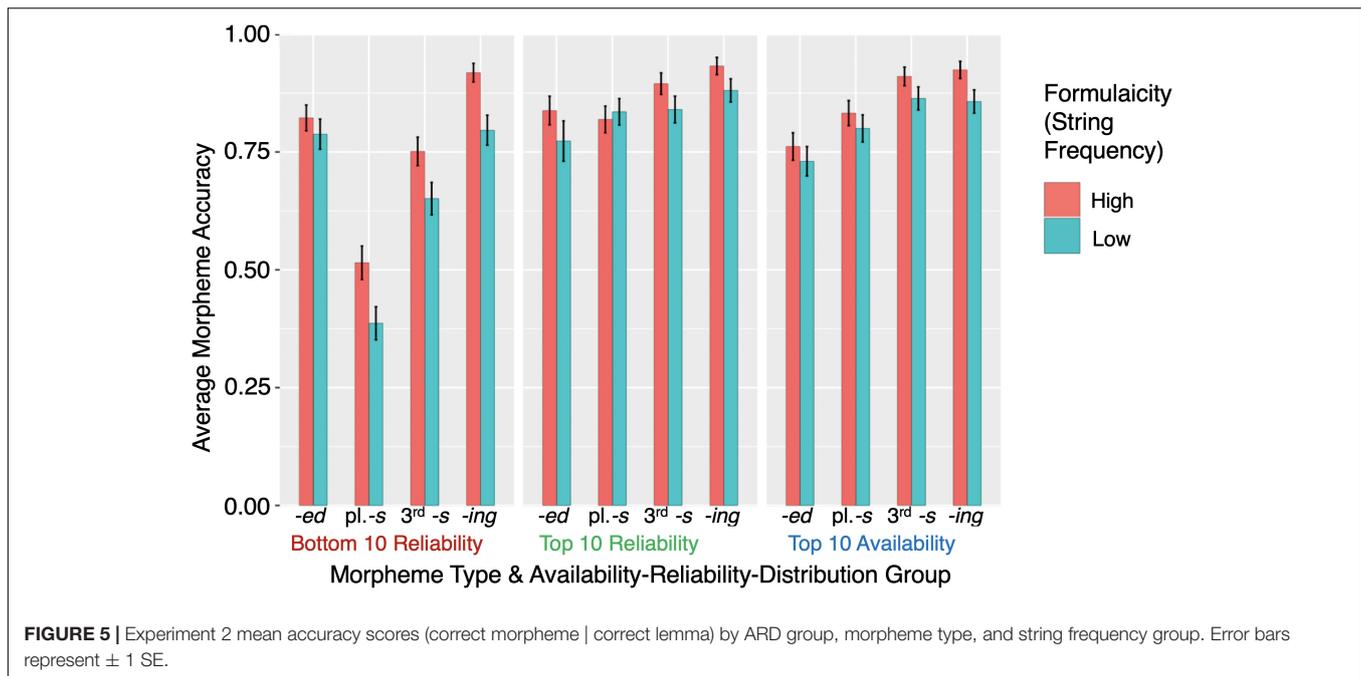
### Model 1: Morpheme Type + Morpheme Reliability + Morpheme Availability

The accuracy scores for sentences by morpheme and ARD group are shown in **Figure 4**. To replicate Experiment 1, we first conducted GLMM analyses without formulaicity as a factor. The results closely mirrored those of Experiment 1 Model 1, demonstrating significant effects of morpheme type, reliability and availability (see **Supplementary Data Sheet 1, 1.4**). In this sample there was reason to include subject random slopes for both reliability [ $\chi^2(df = 2) = 7.754, p = 0.02$ ] and availability [ $\chi^2(df = 3) = 314.7, p = 0.000$ ]<sup>10</sup>.

### Model 2: Morpheme Type + Morpheme Reliability + Morpheme Availability + Formulaicity

The accuracy scores for sentences by morpheme, ARD group, and formulaicity are shown in **Figure 5**. To examine any

<sup>10</sup>Like Experiment 1, in order to determine which subject-level random slopes to include, we ran each of the models in Experiment 2 both with and without subject random slopes and compare the fit of the two models using a likelihood ratio test. We included the subject random slope for a fixed-effect predictor only when this addition significantly improves model fit. Otherwise, we report the model without the subject random slope (i.e., the one with random intercept only).



additional effects of formulaicity, we added string frequency as another fixed-effect predictor and conducted incremental model comparisons to determine a maximal model which included fixed-effects of morpheme type, morpheme reliability, morpheme availability, and formulaicity, random intercepts for subjects and items, and random slopes by subject for morpheme reliability and morpheme availability. In creating the sentence stimuli for Experiment 2, formulaicity was inevitably correlated with sentence length (on average, the low string frequency sentences were 1.175 syllables longer than their high frequency counterparts). To rule out the possibility that sentence length is what drives any observed formulaicity effect, as we did in Experiment 1 (where sentence length was non-significant), we included sentence length (i.e., total number of syllables) as a fixed-effect predictor of morpheme production accuracy. Each participant's stimulus sense rating for each sentence was also included as a potential predictor; model comparison using a likelihood ratio test showed no need to include subject random slopes for sense rating [ $\chi^2(df = 6) = 0.911, p = 0.99$ ]. The final model is summarized in **Table 5**. To see the random slopes for effects of availability and reliability in each subject are shown, refer to **Supplementary Data Sheet 1, 3.1**.

There were effects of morpheme type: *-ing* had significantly higher accuracy than *-ed* (estimate = 1.293,  $SE = 0.246, z = 5.25, p = 0.000$ ); Plural *-s* had significantly lower accuracy than *-ed* (estimate =  $-0.875, SE = 0.239, z = -3.66, p = 0.000$ ). The difference between the third person present tense *-s* and *-ed* was also significant (estimate = 0.809,  $SE = 0.236, z = 3.42, p = 0.001$ ). The effect of morpheme reliability was highly significant (estimate = 1.742,  $SE = 0.348, z = 5.01, p = 0.000$ ). Availability had significant but smaller effects (estimate = 0.301,  $SE = 0.126, z = 2.39, p = 0.017$ ). Formulaicity was significant (estimate =  $-0.619, SE = 0.223, z = -2.77, p = 0.006$ ). Stimulus sentence length was non-significant (estimate =  $-0.014,$

$SE = 0.129, z = -0.11, p = 0.92$ ). Stimulus sense rating was significant (estimate = 0.126,  $SE = 0.035, z = 3.65, p = 0.000$ ).

Analysis of Deviance using Type III Wald chi-square tests shows that morpheme type, morpheme reliability (proportion), morpheme availability (log word-form frequency), formulaicity (string frequency), and sentence sense rating were all significant predictors of accuracy [morpheme type:  $\chi^2(df = 3) = 81.940, p = 0.000$ ; morpheme reliability:  $\chi^2(df = 1) = 25.111, p = 0.000$ ; morpheme availability:  $\chi^2(df = 1) = 5.724, p = 0.017$ ; formulaicity:  $\chi^2(df = 1) = 7.688, p = 0.006$ ; sentence sense rating:  $\chi^2(df = 1) = 13.321, p = 0.000$ ]. Stimulus length was non-significant:  $\chi^2(df = 1) = 0.012, p = 0.912$ .

Model 2b (mid panel of **Table 5**) investigated the interaction between morpheme type and reliability. Past-tense *-ed* was the reference level for type. Allowing for the interaction removes any overall effect of reliability (estimate =  $-0.604, SE = 0.598, z = -1.01, p = 0.312$ ). However, there remain significant effect of reliability on Plural *-s* (estimate = 3.995,  $SE = 0.722, z = 5.54, p = 0.000$ ), third person present tense *-s* (estimate = 2.781,  $SE = 1.010, z = 2.75, p = 0.006$ ), and Prog-*ing* (estimate = 1.881,  $SE = 0.946, z = 1.99, p = 0.047$ ).

Model 2c (lower panel of **Table 5**) investigated the interaction between morpheme type and availability, again with past-tense *-ed* as the reference level. Allowing for the interaction removes any overall effect of availability (estimate =  $-0.158, SE = 0.209, z = -0.76, p = 0.448$ ); although there remains a substantial effect of availability upon Plural *-s* (estimate = 1.971,  $SE = 0.369, z = 5.34, p = .000$ ) and third person present tense *-s* (estimate = 1.051,  $SE = 0.281, z = 3.74, p = 0.000$ ).

### Model 3: Exploring Effects of Proficiency

To test whether individual differences in English proficiency were reflected in participants' morpheme production accuracy, we first included self-rated general proficiency scores as a fixed-effect

**TABLE 5 |** Experiment 2 including fixed-effects of morpheme type, morpheme reliability (proportion), morpheme availability (log word-form Frequency), formulaicity (string frequency), length (syllables), and stimulus sense rating, and random effects of subject and item.

**Model 2: no interactions.**

Parameters	Fixed effects					Random effects	
	Estimate	SE	z	p	By subject	By item	
					SD	SD	
Intercept	0.484	2.135	-0.23	0.820	3.027	1.048	
Morpheme type <sup>1</sup>	Plural -s	-0.875	0.239	-3.66	0.000***		
	Third-person -s	0.809	0.236	3.42	0.001***		
	Progressive -ing	1.293	0.246	5.25	0.000***		
Morpheme reliability	1.742	0.348	5.01	0.000***	0.786		
Morpheme availability <sup>2</sup>	0.301	0.126	2.39	0.017*	0.026		
Formulaicity (low) <sup>3</sup>	-0.619	0.223	-2.77	0.006**			
Sentence length (syllables)	-0.014	0.129	-0.11	0.921			
Sense rating	-0.126	0.035	-3.65	0.000***			

<sup>1</sup>The Past-tense -ed is the reference level. <sup>2</sup>Word-form frequency was logarithmically transformed. <sup>3</sup>High formulaicity was the reference level. Model formula: accuracy~morpheme + reliability + availability + formulaicity + length + sense + (1 + reliability + availability| subject) + (1| item). \*\*\*p = 0; \*\*p < 0.001; \*p < 0.01; p < 0.1.

**Model 2b: Interactions between morpheme type and reliability.**

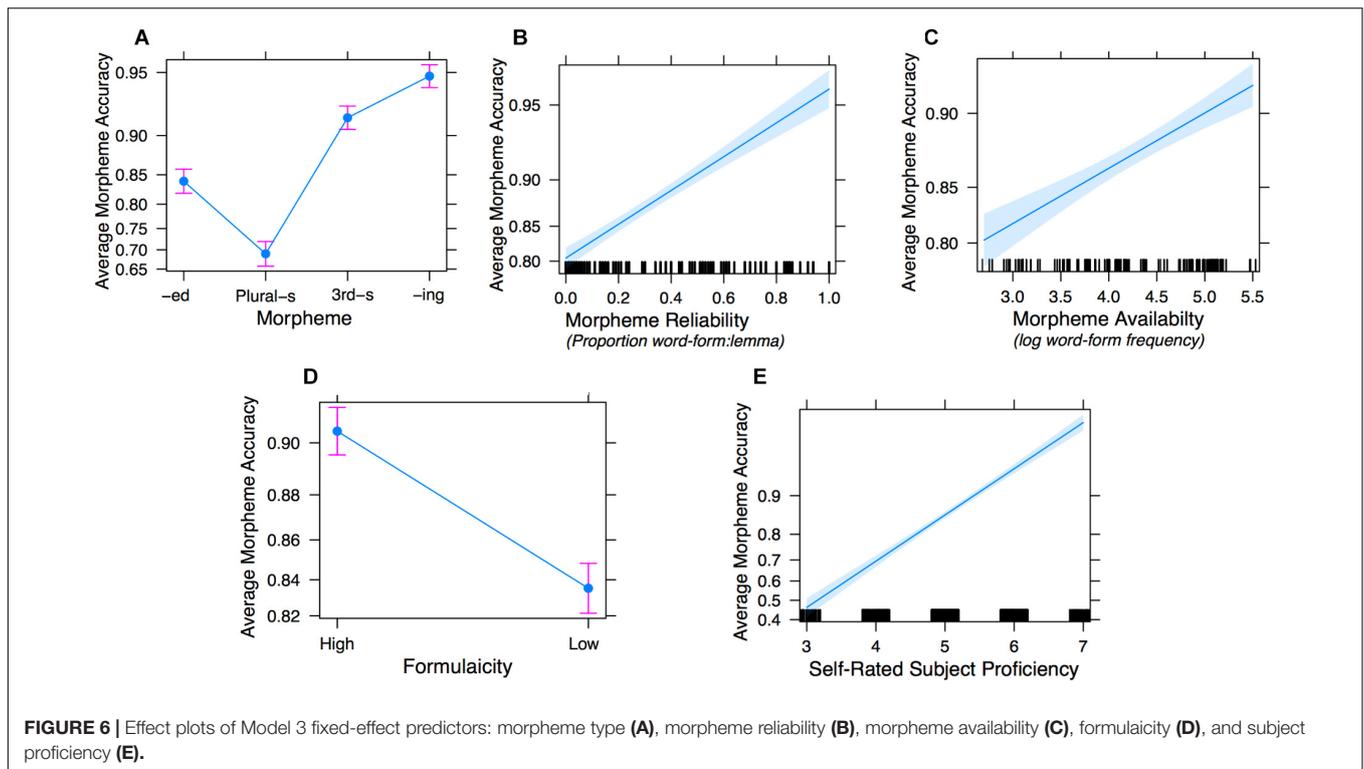
Parameters	Fixed effects					Random effects	
	Estimate	SE	z	p	By subject	By item	
					SD	SD	
Intercept	4.294	1.886	2.28	0.023*	1.288	0.955	
Morpheme type <sup>1</sup>	Plural -s	-2.249	0.352	-6.38	0.000***		
	Third-person -s	-0.059	0.32	-0.18	0.853		
	Progressive -ing	0.743	0.34	2.19	0.029*		
Morpheme reliability	-0.604	0.598	-1.01	0.312	0.960		
Formulaicity (low) <sup>2</sup>	-0.462	0.211	-2.19	0.028*			
Sentence length (syllables)	-0.153	0.122	-1.25	0.210			
MorphemePlural-s: Reliability	3.995	0.722	5.54	0.000***			
MorphemePres-s: Reliability	2.781	1.01	2.75	0.006**			
MorphemeProg-ing: Reliability	1.881	0.946	1.99	0.047*			

<sup>1</sup>Past-tense -ed is the reference level. <sup>2</sup>High formulaicity was the reference level. Model formula: accuracy~morpheme\*reliability + formulaicity + length + (1 + reliability| subject) + (1| item). \*\*\*p = 0; \*\*p < 0.001; \*p < 0.01; p < 0.1.

**Model 2c: interactions between morpheme type and availability.**

Parameters	Fixed effects					Random effects	
	Estimate	SE	z	p	By Subject	By Item	
					SD	SD	
Intercept	2.453	2.209	1.110	0.267	1.943	0.984	
Morpheme type <sup>1</sup>	Plural -s	-9.346	1.655	-5.650	0.000***		
	Third-person -s	-3.651	1.147	-3.180	0.001**		
	Progressive -ing	0.526	1.292	0.410	0.684		
Morpheme availability	-0.158	0.209	-0.760	0.448	0.269		
Formulaicity (low) <sup>2</sup>	-0.638	0.217	-2.940	0.003**			
Sentence Length (syllables)	-0.002	0.127	-0.020	0.985			
MorphemePlural-s: Availability	1.971	0.369	5.340	0.000***			
MorphemePres-s: Availability	1.051	0.281	3.740	0.000***			
MorphemeProg-ing: Availability	0.169	0.302	0.560	0.575			

<sup>1</sup>Past-tense -ed is the reference level. <sup>2</sup>High formulaicity was the reference level. Model formula: accuracy~morpheme\*availability + formulaicity + length + (1 + availability| subject) + (1| item). \*\*\*p = 0; \*\*p < 0.001; p < 0.1.



predictor (in addition to morpheme type, reliability, availability, and formulaicity) without any interaction terms. As shown in **Figure 6**, the effects of morpheme type were consistent with previous models: *-ing* had significantly higher accuracy than *-ed* (estimate = 1.260,  $SE = 0.245$ ,  $z = 5.15$ ,  $p = 0.000$ ); the Plural *-s* had significantly lower accuracy than *-ed* (estimate =  $-0.857$ ,  $SE = 0.238$ ,  $z = -3.60$ ,  $p = 0.000$ ); the third-person *-s* had significantly higher accuracy than *-ed* (estimate = 0.762,  $SE = 0.234$ ,  $z = 3.26$ ,  $p = 0.001$ ). The significant effects of morpheme reliability (estimate = 1.742,  $SE = 0.341$ ,  $z = 5.10$ ,  $p = 0.000$ ), availability (estimate = 0.326,  $SE = 0.122$ ,  $z = 2.67$ ,  $p = 0.008$ ), and formulaicity (estimate =  $-0.628$ ,  $SE = 0.165$ ,  $z = -3.81$ ,  $p = 0.000$ ) all remained the same as in Model 2. Importantly, subject proficiency also positively predicted morpheme accuracy (estimate = 0.974,  $SE = 0.162$ ,  $z = 6.02$ ,  $p = 0.000$ ), and the addition of subject proficiency as a fixed-effect predictor significantly improved model fit from Model 2 [ $\chi^2(df = 1) = 27.24$ ,  $p = 0.000$ ].

The individualized subject slopes in **Figure 7** suggest that participants who have lower proficiency show greater effects of reliability. This interaction seems less marked for effects of availability. To further investigate whether the effects of morpheme reliability, morpheme availability, and formulaicity varied as a function of proficiency, we added the interaction terms into the GLMM (see **Table 6** Model 3).

Model comparisons using a likelihood ratio test revealed that adding the interaction terms significantly improved model fit [ $\chi^2(df = 3) = 8.20$ ,  $p = 0.04$ ]. The morpheme type effects were consistent with previous models. The positive effect of morpheme reliability (estimate = 3.946,  $SE = 1.142$ ,  $z = 3.46$ ,

$p = 0.001$ ), formulaicity (estimate =  $-1.437$ ,  $SE = 0.614$ ,  $z = -2.34$ ,  $p = 0.02$ ), and proficiency (estimate = 1.234,  $SE = 0.378$ ,  $z = 3.26$ ,  $p = 0.001$ ) on morpheme accuracy all remained significant, but the effect of morpheme availability was no longer significant. Notably, there was a significant interaction between proficiency and morpheme reliability (estimate =  $-0.455$ ,  $SE = 0.220$ ,  $z = -2.06$ ,  $p = 0.04$ ): morpheme reliability effects on production accuracy are greater at lower levels of proficiency (**Figure 7A**). Proficiency did not interact with the other two distribution factors: the effects of morpheme availability and formulaicity stay the same across different levels of proficiency (**Figures 7B,C**).

### Summary of Experiment 2 Results

Experiment 2 replicated the pattern of the rank order difficulty of the four target morphemes: three of the four morphemes conformed to the natural order reported in prior morpheme studies, in that the *-ing* had higher accuracy than the *-ed* and the third-person *-s*. The plural *-s* was again found to be more difficult than what the common order would have predicted. The effects of morpheme availability and reliability were consistent with the results from Experiment 1 and with previous studies. The effects of reliability were greater than those of availability. The interactions with morpheme type confirmed significant reliability effects on Plural *-s*, third person present tense *-s*, and Prog-*ing*. Effects of reliability were greater at lower levels of proficiency; there was no such interaction between availability and proficiency. There were Phrase-Superiority Effects whereby higher frequency four-word strings were associated with increased accuracy of production of

morphemes embedded therein. These formulaicity effects were not explicable in terms of sentence length. Participants showed greater accuracy of morpheme provision in sentences they rated as making more sense, though this correlation says nothing about the direction of causation.

## GENERAL DISCUSSION

This study investigated ESL learner's productive morphology in non-obligatory contexts using elicited imitation of sentences containing corpus-sampled morpheme exemplars varying across three probabilistic distribution patterns of item-level features: availability, reliability, and formulaicity. We found that a morpheme is better perceived and more accurately reproduced when (1) it occurs as a word-form that is frequent in usage (i.e., highly available), (2) it is attached to a word that is more consistently conjugated in the form containing this morpheme compared to other forms (i.e., highly reliable), and (3) when the word-form containing it is embedded in a frequent four-word string (i.e., highly formulaic).

### Availability

The facilitation effect of morpheme availability was consistent with the results of prior studies on L1 and L2 acquisition of English morphemes (Larsen-Freeman, 1976; Marchman, 1997; Goldschneider and DeKeyser, 2001; Jia and Fuse, 2007; Räsänen et al., 2016). For example, Marchman (1997) investigated the productive use of English past-tense morphology in children in an elicited-production task and found that errors on regular and irregular verbs (e.g., zero-marking, over-regularization, etc.) were all predicted by item frequency among other factors. Using a similar paradigm, Räsänen et al. (2016) targeted the elicited production of inflectional morphology in L1 Finnish children and found that person/number marked verbs were

produced faster and with less errors if they were high-frequency word-forms. Naigles and Hoff-Ginsberg's (1998) corpus analyses on mother-child conversations collected in naturalistic settings also revealed that the frequency with which verbs appeared in the child-directed speech was significantly predictive of how often and how flexibly the child produced the verbs 10 weeks later. Using production-based measures, Braine et al. (1990) investigate child and adult learning of inflected word-forms of an artificial language and found effects of item and pattern frequency. These include: (1) high frequency morphemes were learned faster and were favored over less frequent ones; (2) item frequency was a significant factor in learning idiosyncratic irregulars; (3) production of a morpheme appropriate to a noun in a generalization test was often affected by whether or not their pairing had been presented previously. Likewise, in adult language acquisition, Finley (2015) trained English speakers on an artificial grammar involving words containing suffixes varying in frequencies. Subsequent testing showed effects upon both memory and generalization: (1) words with high-frequency suffixes were judged to be more acceptable than those with low-frequency suffixes in grammaticality judgment tests where they were to be distinguished from ungrammatical forms, and (2) novel items containing high-frequency suffixes were more likely to be accepted as grammatical compared to those containing low-frequency suffixes. These studies, together with our results of advanced ESL speakers here, all manifest the effects of prior language experience upon processing and cognitive representation. More available forms in usage become entrenched and are more readily perceived and produced by learners (Ellis, 2011).

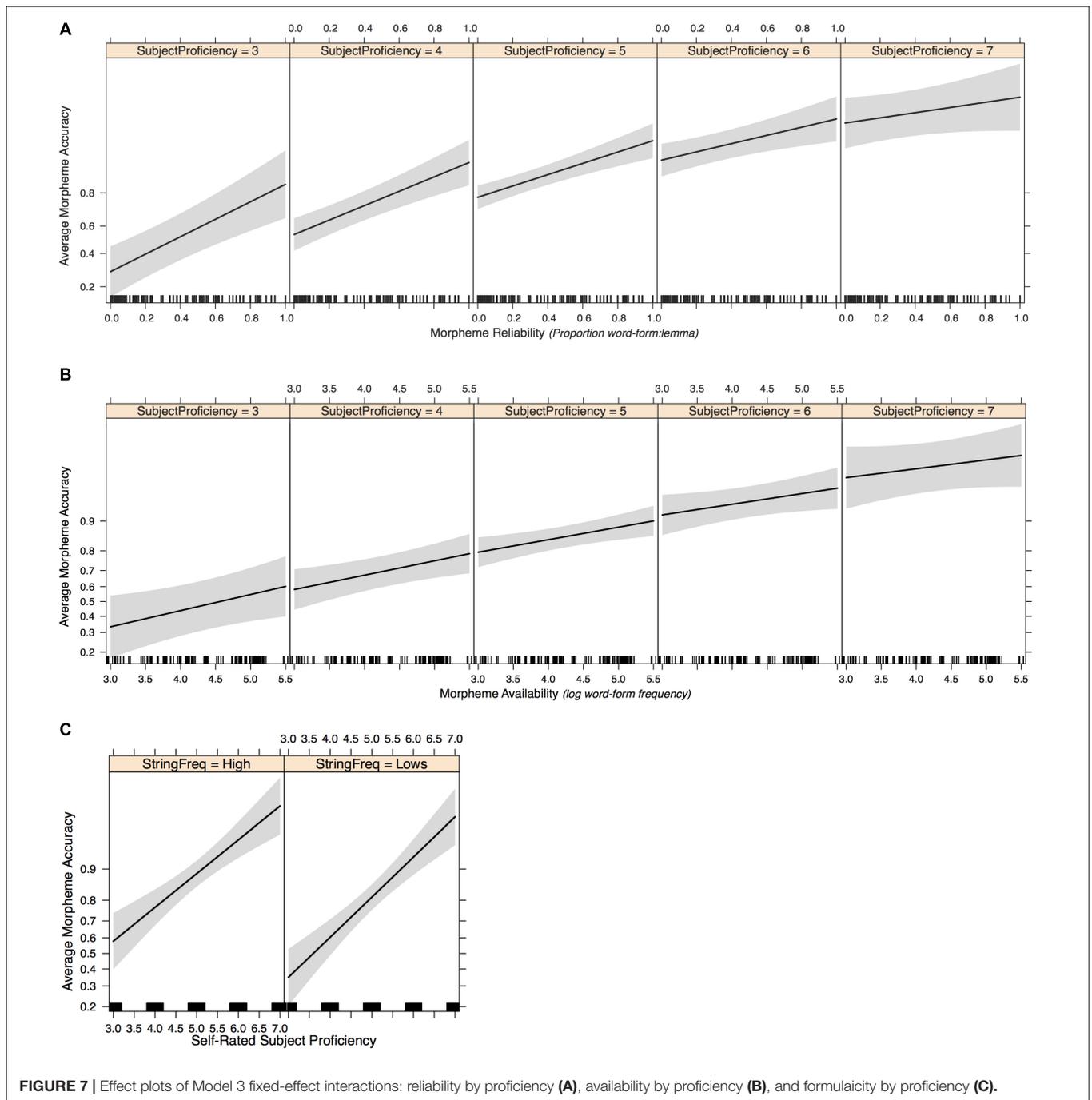
### Reliability

Usage-based studies of acquisition and processing show that there are a range of different frequencies beyond mere availability

**TABLE 6** | Experiment 2 Model 3 results from the mixed effects model with proficiency and interactions between proficiency and the distribution factors.

Parameters	Fixed effects					Random effects	
	Estimate	SE	z	p	By subject <sup>4</sup>	By item	
Intercept	-6.366	1.939	-3.28	0.001**	0.82	1.00	
Morpheme Type <sup>1</sup>							
Plural -s	-0.888	0.235	-3.78	0.000***			
Third-person -s	0.746	0.231	3.23	0.001**			
Progressive -ing	1.242	0.242	5.14	0.000***			
Morpheme reliability	3.946	1.142	3.46	0.001***			
Morpheme availability <sup>2</sup>	0.591	0.433	1.36	0.172			
Formulaicity (Low) <sup>3</sup>	-1.437	0.614	-2.34	0.019*			
Proficiency	1.234	0.378	3.26	0.001**			
Proficiency*Morpheme reliability	-0.455	0.220	-2.06	0.039*			
Proficiency*Morpheme availability	-0.050	0.085	-0.58	0.559			
Proficiency*Formulaicity	0.165	0.120	1.38	0.168			

<sup>1</sup>Past-tense -ed is the reference level. <sup>2</sup>Word-form frequency was logarithmically transformed. <sup>3</sup>High formulaicity was the reference level. Model formula: accuracy ~ morpheme + proficiency \* (reliability + availability + formulaicity) + (1|subject) + (1|item). <sup>4</sup>Random slopes by subject are not included in this model because adding them did not significantly improve model fit [ $\chi^2(df=5) = 10.30, p = 0.07$ ]. \*\*\* $p = 0$ ; \*\* $p < 0.001$ ; \* $p < 0.01$ ;  $p < 0.1$ .



that are important in driving the acquisition of linguistic constructions (Ellis, 2012a). The automatic computation that underpins implicit learning does not just tally individual forms or functions, it also automatically learns associations between forms and functions, between forms and other forms, and between forms and contexts (Ellis et al., 2016).

Psychological research into associative learning has long recognized that while frequency of form is important, more so still is contingency of cue-outcome or form-function mapping (Shanks, 1995; MacWhinney, 2001; Ellis, 2006a). Cues with

multiple interpretations (i.e., low-contingency) are ambiguous and thus hard to resolve; whereas cue-outcome associations of high-contingency are reliable in their interpretation and readily processed. Consider how, in the learning of the category of birds, while eyes and wings are equally frequently experienced features in the exemplars, it is wings which are distinctive in differentiating birds from other animals. Wings are important features to learning the category of birds because they are reliably associated with class membership while being absent from outsiders. Raw frequency of occurrence is therefore less

important than the contingency between cue and interpretation. Reliability of form–function mapping is a driving force of all associative learning, human and animal alike, to the degree that the field of its study has become known as ‘contingency learning.’ This is well recognized in second language acquisition research. For example, Andersen’s (1984) ‘One to One Principle’ of interlanguage construction specifies that an interlanguage system should be constructed in such a way that an intended underlying meaning is expressed with one clear invariant surface form or construction. Contingency learning is also central to the Competition Model, a psycholinguistic theory of language acquisition and sentence processing (MacWhinney, 1987; MacWhinney and Bates, 1989), as well as to other psycholinguistic models of construction learning as the rational learning of form–function contingencies (Ellis, 2006a; Xu and Tenenbaum, 2007).

The competition model focuses upon the various morphological, syntactic, and semantic linguistic cues contained in a sentence – e.g., case marking, word order, and semantic characteristics such as animacy – which people use to interpret the meaning of the sentence. Each cue is probabilistically associated with a particular interpretation, and the cue-weights combine in allowing the learner to choose the interpretation with the highest likelihood. Learners assign cue-weights inductively over their history of experience and usage. Cue-weights differ between languages as different languages use different cues to signal meanings. Thus, second and foreign language learners must learn which cues are important in which languages. To do this, they begin with cues that are more available in the input, 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, 2001).

Reliability of association is similarly key in cognitive-linguistic, corpus-based and statistical models of language structure like collocation analysis (Stefanowitsch and Gries, 2003; Gries and Stefanowitsch, 2004; Gries, 2009). Cognitive linguistic theories of construction grammar focus upon lexical, morphological, and syntactic forms as form–function pairings (Goldberg, 1995, 2006; Hoffmann and Trousdale, 2013). Collocation analysis focuses more upon form–form reliability in measuring the degree of attraction or repulsion that words exhibit to constructions. It comprises three different methods: (1) collexeme analysis, which measures the degree of attraction/repulsion of a lemma to a slot in one particular construction; (2) distinctive collexeme analysis, which measures the preference of a lemma to one particular construction over another, functionally similar construction; (3) covarying collexeme analysis, which measures the degree of attraction of lemmas in one slot of a construction to lemmas in another slot of the same construction. Collocation analysis differs from raw frequency counts by providing not only observed co-occurrence frequencies of words and constructions, but also a comparison of the observed frequency to the frequency expected by chance, so as to measure the attraction and repulsion of words and constructions. These measures of association, contingency, and reliability are found to be better predictors of interpretation than

are measures of availability (Gries, 2015; Gries and Ellis, 2015; Ellis et al., 2016).

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 and Theakston, 2006). There is parallel L2 research showing the importance of reliability. 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, while this preferential bias was not observed in verbs that do not have a common form. In a recent study on Japanese L1 acquisition, Tatsumi et al. (2018) investigated 3–5-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 children’s choice between simple and completive form 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: simple past-tense ratio were more likely to be successfully primed by the experimenter’s use of the completive form, compared to other verbs.

Our present findings further contribute to this growing literature on contingency effects in language processing and production: highly *reliable* morphemes (i.e., exemplars involving lemmas more consistently conjugated in the form containing this morpheme) are more readily acquired and processed.

## Formulaicity and Phrase-Superiority Effects

The results in Experiment 2 demonstrated clear effects of string frequency. There is substantial evidence of chunking and formulaicity effects in first and second language processing and acquisition and in language change (see, for reviews: Ellis, 1996, 2012b, 2017; Wray, 2002; Schmitt, 2004; Siyanova-Chanturia and Pellicer-Sanchez, 2018). For example, Janssen and Barber (2012) demonstrated language learners’ sensitivity to the frequency distributions of multi-word units by having native Spanish speakers produce three types of noun phrases (noun + adjective, noun + noun, determiner + noun + adjective) elicited by line drawings. They found that naming latencies were inversely related to the frequency of the noun phrase but were unrelated to the frequency of the individual words in the phrase. Notably, such formulaicity effects are not restricted to constituents and can span across traditionally defined syntactic boundaries. Tremblay et al. (2011) investigated the processing of non-phraseal sentence segments in a self-paced reading task and found that the frequently occurring “lexical bundles” such as *in the middle of the* were read faster than matched control segments like *in the front of the*. Conklin and Schmitt (2008) embedded such formulaic

sequences in short stories and found that they were read faster than matched non-formulaic sequences by both English native speakers and by proficient non-natives. Recent models of sentence processing like the proposals of Christiansen and Chater (2016) on “Chunk-and-Pass” processing have chunking and prediction at their core – we used this model in section “A process analysis of EIT as it relates to morpheme production” to guide our process analysis of EIT.

The fact that the perception and production of a chunk is more affected by the frequency of the chunk than by the frequencies of its component parts suggests that chunks are not fully analyzed into or assembled from their component parts. Bybee (2010) argues that users do not assemble word-forms from their component parts, but rather they store and access them as wholes. Ellis (1996, p. 111) has likewise suggested that formulas might be stored like single “big words.” But there is longstanding debate about whether such formulaic strings are stored as a unit or simply processed preferentially due to context effects and prediction. Likewise in the SLA literature, there are longstanding discussions about whether formulas and idioms are essential parts of the acquisition process or instead are islands of exception, divorced from the language system (Ellis, 2012b; Wulff, 2019).

Our results here show clear effects whereby higher frequency multi-word strings facilitated elicited imitation of the embedded morphemes. We think of these as Phrase-Superiority Effects, the phrasal equivalents of Word-Superiority Effects (WSE, Reicher, 1969; Wheeler, 1970) whereby recognition of a letter is more accurate when it is part of a meaningful word than when it is alone. In WSE experiments, a string of four or five letters is flashed for a few milliseconds onto a screen. Readers are then asked to choose which of two letters had been in the flashed string. For example, if “WORD” had been flashed, a reader might have to decide whether “K” or “D” had been in the final letter position. A WSE arises when subjects choose the correct letter more consistently when letter strings are real words rather than non-words (e.g., “OWRD”) or single letters presented alone (e.g., “\_D”). Performance on a forced-choice letter detection task averaged 10% better when the stimuli were four-letter English words than when the stimuli were single letters appearing alone (Wheeler, 1970).

The WSE was a milestone observation in cognitive models of word-recognition and led to the development of the interactive activation model of word recognition (McClelland and Rumelhart, 1981), itself a milestone in the development of connectionist (Parallel Distributed Processing, PDP) models of language. According to this model, when a reader is presented with a word, each letter in parallel either stimulates or inhibits different *feature* detectors (e.g., a curved shape for “C,” horizontal and vertical bars for “T,” etc.). Those feature detectors then stimulate or inhibit different *letter* detectors at a higher level, which finally stimulate or inhibit different *word* detectors at the top-most level. Each activated connection in the large parallel network of connections carries a different weight, and the activation is propagated across levels to give the word detector for “WORD” (in the example above) more activation than any other detector, making “WORD” what the reader eventually

recognizes. So far, so bottom-up good. But there are top-down connections too: word detectors pass excitation down to the letter detectors for the letters they contain and inhibit letters they do not; letter detectors pass excitation down to feature detectors for the features they contain and inhibit features they do not. Finally, the model includes inhibitory connections within levels, so that the activation of “WORD” inhibits that of other words, like “WORK,” “WORM,” “LORD,” etc.

The interactive-activation model is a working computational model. It both simulates and explains the WSE as follows: When the target letter is presented within a word, the feature detectors, letter detectors and word detectors are all activated, adding weight to the final recognition of the WORD stimulus, and this in turn sends activation down to its component letters and features. Thus, recognizing the “D” in “WORD” results from activation from both the feature detectors and from the word detectors. However, when recognizing “D” with only the letter presented, there is only the bottom-up activations to the letter detector level. Therefore, perceiving a presented word allows more accurate identification of its component letters, as observed in the WSE.

The WSE demonstrates how frequency and activation at one level of representation (words) may affect the processing and acquisition of linguistic stimuli at another level (letters). Yet ubiquitously, linguistic constructions are inherently nested across various overlapping levels (e.g., morphemes within a word, words within a phrase, phrases within a sentence, all of which can be decomposed into phonemes, etc.). Thus it is likely that there are effects of frequency and contingency across many different grain-sizes of construction, all of which might overlap and interactively activate in intricate ways (Ellis, 2012a; Gries and Ellis, 2015). The Phrase-Superiority effects show how frequency of phrases percolates down to affect the processing of the embedded words and morphemes.

## Between-Morpheme Comparisons

As predicted by the common ESL morpheme acquisition order (Krashen et al., 1977), the *-ing* morpheme was found to be the easiest to acquire by the Chinese participants in our study. This is likely a result from the presence of several facilitating properties such as high level of perceptual saliency, as proposed by Goldschneider and DeKeyser (2001). The morpheme *-ing* is both phonologically salient (because it is a syllabic vowel) and syntactically salient (because its usage is morphosyntactic in nature due to the required co-occurrence of auxiliary *be*). Since auxiliary *be* has been consistently found to be mastered quite well by ESL learners (Krashen et al., 1977), it might have served as an effective cue that have improved the memory for the co-occurring *-ing*. In addition, *-ing* is high in morphophonological regularity and low in semantic complexity, a 1:1 mapping which promotes acquisition.

The past-tense *-ed* morpheme was found to be second easiest to acquire. Besides saliency and regularity, the difference between the processing difficulty of *-ed* vs. *-ing* among Chinese speakers could also be due to typological differences between English and Chinese: namely, English is a “tense-prominent” language whereas Chinese is “aspect-prominent.” It has been reported in the L1 transfer literature that ESL learners who have

“aspect-prominent” L1s such as Punjabi use fewer English tense markers such as *-ed* and more aspect markers such as *-ing*, compared to ESL learners whose L1 is the “tense-prominent” Italian (Slobin, 1996). In other words, when exposed to English, Chinese native speakers might habitually pay more attention to aspectually marked verbs than tense-marked verbs.

Interestingly, the plural *-s* morpheme was the lowest in production accuracy in our Chinese speaker sample, which deviates from Krashen et al.’s (1977) common order. This finding is also inconsistent with Tsimplici and Dimitrakopoulou’s (2007) interpretability hypothesis, according to which morphological features related to “interpretable” universal semantic concepts – e.g., plurality – should be easier to acquire than purely grammatical and language-dependent features that lack semantic significance – e.g. the verb agreement on third person singular subjects. One possible explanation why the plural *-s* is difficult for Chinese native speakers lies in the typological differences in how the concept of plurality is expressed between English and Chinese (Jarvis and Pavlenko, 2008; Murakami and Alexopoulou, 2016). In classifier languages like Chinese, plurality (as well as the concept of count/mass distinction) is typically expressed with stand-alone classifiers besides numbers, rather than morphological inflections. Thus, the property (e.g., plurality) of the nouns is entailed by the classifiers that they follow, and not by the nouns themselves. Jiang (2004) argues that Chinese speakers are morphologically insensitive to number information and the count/mass distinction in nouns, especially for those representing abstract concepts, because there is not a classifier that specifically expresses the abstract property of the noun and that is used to count the noun. This also explains why the plural *-s* in the Bottom 10 Reliability group, which consists mostly of abstract nouns, had especially low production accuracy.

Such L1 transfer effects and the resulting deviations from the common order have been demonstrated in previous studies. For example, Luk and Shirai (2009) found that L1 speakers of Japanese, Korean, and Chinese acquire the plural *-s* and articles later than as predicted by the natural order while acquiring the possessive *’s* much earlier. To examine the nature of L1 influence, Murakami and Alexopoulou (2016) conducted a corpus analysis on English morpheme acquisition by ESL learners from seven L1 backgrounds using a database consisting of written exam scripts drawn from the Cambridge Learner Corpus. Usage analyses of six morphemes (articles, past tense *-ed*, plural *-s*, possessive *’s*, progressive *-ing*, and third person *-s*) revealed significant between-L1 differences, such that L1 type, i.e., whether an equivalent form of the morpheme in English is present or absent in the L1, strongly predicted morpheme accuracy and the order of acquisition. For instance, Japanese L1 speakers tend to score higher on possessive *’s* and lower on articles than French L1 speakers do, as Japanese lacks the grammatical articles which French has, while French lacks a possessive morpheme which Japanese has. In fact, the lack of L1 equivalence resulted in an accuracy below 90% in almost all morphemes and L1s even among the highly proficient ESL speakers. In addition, the researchers also reported differential influence of L1 type on different morphemes, with articles and progressive *-ing* being the most sensitive to L1 influence,

plural *-s* mildly affected, and possessive *’s* and third-person *-s* the least affected.

Nevertheless, the findings of L1 transfer effects on morpheme acquisition do not deny universal tendencies in the order of morpheme acquisition that are driven by the L2 linguistic input. Corder (1967) noted that what ‘goes on’ in the environment does not equal what actually ‘goes in’ the learner and introduced the concept of ‘intake’ to represent the subset of the available ‘input’ that learners have attended. Second language learners come to the L2 input already “trained” by their prior language experiences to pay different kinds and degrees of attention to patterns in the L2 (Slobin, 1996; Ellis, 2006b). In other words, the between-L1 variance in English morpheme acquisition reflects how learners of different language backgrounds form different ‘focal sets’ through learned selective attention shaped by the nature of their L1, thus transforming different subsets of the L2 ‘input’ into the actual ‘intake’ (Ellis, 2006b). Such attentional bias was demonstrated by Ellis and Sagarra’s (2010) findings that ESL learners whose L1 lacks verb-tense morphology, such as Chinese, were biased to rely more on lexical (e.g., adverbial) cues than on morphological cues to extract temporal information in English. As a result, they experience greater difficulty in acquiring English tense morphemes compared to learners with morphologically rich L1s such as Spanish and Russian. Ellis (2006b) examined moderators of this blocking or “learned attention” bias and proposed that the contingency of form-function mapping in the L2, i.e., reliability, is a significant factor that determines whether input stimuli become the ‘intake,’ further lending support to the reliability effects on ESL morpheme acquisition.

It is important to note that although our design attempted to deny obligatory contexts for using the target morphemes, we achieved this goal with mixed success. The decontextualized nature of EIT denied extra-sentential cues from referential and pragmatic contexts, however, it was more difficult to remove relevant sentence-internal lexical and structural cues. We had greater success removing the cues for plural *-s* than we had for the verbal inflections, and this alone might explain learners’ unexpectedly low performance on this morpheme compared to the others. In contrast, Auxiliary [be] was always provided as a cue for progressive *-ing*, as needs must, and this may well give progressive *-ing* a processing advantage over the others. Even though we randomly allocated the primacy-denying three-word opening phrases such as *On Saturday morning*, *Late Wednesday evening*, etc., and these are theoretically tense-neutral, in fact they have tense and modality associations from usage. The *Late Wednesday evening* opening pulls for simple past tense if no auxiliary is provided, and progressive *-ing* if the auxiliary is present. Designating a day of the week and a time of day such as *Late Wednesday evening* implies a more punctual one-time occurrence more than less constrained openings like *On Saturday morning*, and the more specific implication is not always compatible with the Simple Present or Present Progressive. Likewise, the third-person Present *-s* Top 10 reliability set is skewed toward verbs with dummy subjects that only occur in the 3rd-person present in that construction, such as *it concerns*, *it implies*, *it consists*, *it sounds*, and these feel strange in punctual temporal contexts. This range of systemic biases may well have



interact in various complex and adaptive ways in usage, and a speaker's language system reflects 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.

## LIMITATIONS

There are various limitations to this study. (1) It is relatively small in size in terms of its sample of participants and its sample of stimuli<sup>11</sup>. (2) We are also concerned just how representative the corpus is for our participants. We drew our stimuli from the largest existing corpus of American English, COCA (Davies, 2008), assuming it to truly represent the English exposure and usage of the non-native English-speaking participants living in the US. This assumption can be problematic as the English used by the general population as the L1 might well vary from that used as an L2 in pedagogical settings. Despite our best efforts, participants might not be familiar with all the vocabulary in the sentences, and particularly, with the target words. If they did not know the target word, it seems likely that the whole word would be omitted in the response. Although we tried to mitigate the problem by only scoring the morpheme provision when the correct lemma is present, it still unavoidably resulted in an unequal number of missing cases across the ARD groups, with the highest exclusion rate in the bottom frequency and proportion group. (3) In creating the stimuli for Experiment 2, in order to control a range of important potential other factors, in making the low-frequency sentence strings, we replaced (high-frequency and shorter pronouns) with (lower frequency and longer proper nouns). This systematic confound introduces uncertainty into whether any effects of this manipulation result from sentence string frequency, length, or pronoun vs. noun. We were at least able to show that string frequency was a much more important effect than length. Nevertheless, if possible, it would be good to avoid such confounds in future research. (4) Our design attempted to provide non-obligatory contexts for using the target morphemes, but we achieved this goal with mixed success. The decontextualized nature of EIT denied extra-sentential cues from referential and pragmatic contexts, however, it was more difficult to remove relevant sentence-internal lexical and structural cues. For example, Auxiliary [be] was always provided as a cue for progressive -ing. This, and the range of other factors detailed at the end of section "Between-Morpheme Comparisons," introduce a range of factors that deny simple identification of the causes of between-morpheme differences. (5) Our quest for control and the matching of the stimuli in terms of several dimensions of corpus metrics resulted in stimulus sentences that are somewhat uneven in their approximation of naturally occurring English. (6) Our chosen experimental paradigm, the Elicited Imitation Test, targets decontextualized language repetition rather than situations of rich, meaningful communication where there is clearly more scope for the importance of word meanings and other form-function associations. (7) Adapting the EIT for typed

rather than spoken responding potentially allows more influence of considered explicit processing in the written responses, although the window for these influences comes after online listening, which we believe to be the rate-limiting step which maximizes demands for implicit or automatized processing. However, further research involving spoken responding would be useful for triangulation.

For future research, we encourage the analysis of large learner corpora (of the type exemplified by, e.g., Murakami and Alexopoulou, 2016) in order to broaden the investigation to many more learners, large amounts of more communicative natural language, a wider range of morphology, and a focus upon participant effects (including L1 transfer, longitudinal development, proficiency, etc.). Widening the range of languages studied is also a priority.

## CONCLUSION

We investigated usage-based effects of availability, reliability, and formulaicity in ESL acquisition of inflectional morphemes: *-ed*, *-ing*, and 3rd-person *-s* on verbs, and plural *-s* on nouns and showed using EIT that morphemes were more easily processed when they were (1) available (occurring in frequent word-forms), (2) reliable (occurring in lemmas consistently conjugated in this form), and (3) formulaic (embedded in high- vs. low-frequency phrases). Such conclusions support cognitive theories of the statistical symbolic learning of morphology. Language acquisition reflects the distributional properties of the linguistic input at multiple grain-sizes.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Michigan IRB Health Sciences and Behavioral Sciences (HSBS). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.582259/full#supplementary-material>

<sup>11</sup> Although, in defense, we point out that there is replication research here between Experiments 1 and 2, and that the stimuli were carefully chosen from a corpus of over 560 million words to lie at the interesting extremes to afford maximal power.

## REFERENCES

- Ambridge, B., Kidd, E., Rowland, C. F., and Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *J. Child Lang.* 42, 239–273. doi: 10.1017/S030500091400049X
- Andersen, R. W. (1984). The one-to-one principle of interlanguage construction. *Lang. Learn.* 34, 77–95. doi: 10.1111/j.1467-1770.1984.tb00353.x
- Andersen, R. W., and Shirai, Y. (1996). “The primacy of aspect in first and second language acquisition: the pidgin-creole connection,” in *Handbook of Second Language Acquisition*, eds W. Ritchie and T. Bhatia (Leiden: Brill), 527–570. doi: 10.1016/b978-012589042-7/50018-9
- Arnon, I., and Snider, N. (2010). More than words: frequency effects for multi-word phrases. *J. Mem. Lang.* 62, 67–82. doi: 10.1016/j.jml.2009.09.005
- Bannard, C., and Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on Children’s repetition of four-word combinations. *Psychol. Sci.* 19, 241–248. doi: 10.1111/j.1467-9280.2008.02075.x
- Bardovi-Harlig, K. (2000). *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Malden, MA: Blackwell.
- Bardovi-Harlig, K., and 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. *Stud. Second. Lang. Acquis.* 42, 1137–1167. doi: 10.1017/S0272263120000194
- Barry, C., and Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Q. J. Exp. Psychol. A* 40, 5–40. doi: 10.1080/14640748808402280
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:115553. doi: 10.18637/jss.v067.i01
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59, 1–26. doi: 10.1111/j.1467-9922.2009.00533.x
- Biber, D. (2004). If you look at: lexical bundles in university teaching and textbooks. *Appl. Ling.* 25, 371–405. doi: 10.1093/applin/25.3.371
- Bod, R. (2001). *Sentence Memory: Storage vs. Computation of Frequent Sentences. Paper Presented at the CUNY-2001*. Philadelphia, PA: University of Pennsylvania.
- Bod, R., Hay, J., and Jannedy, S. (2003). *Probabilistic Linguistics*. Cambridge, MA: MIT Press.
- Braine, M. D. S. (1976). Children’s first word combinations. *Monogr. Soc. Res. Child Dev.* 41:104.
- Braine, M. D. S., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., et al. (1990). Exploring language acquisition in children with a miniature artificial language: effects of item and pattern frequency, arbitrary subclasses, and correction. *J. Mem. Lang.* 29, 591–610. doi: 10.1016/0749-596X(90)90054-4
- Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Bybee, J. (2010). *Language, Usage, and Cognition*. Cambridge University Press.
- Bybee, J. L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins Publishing Company, doi: 10.1075/tsl.9
- Bybee, J. L. (2006). From usage to grammar: the mind’s response to repetition. *Language* 82, 711–733. doi: 10.1353/lan.2006.0186
- Bybee, J. L., and Hopper, P. J. (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins Publishing Company, doi: 10.1075/tsl.45
- Childers, J. B., and Tomasello, M. (2001). The role of pronouns in young children’s acquisition of the English transitive construction. *Dev. Psychol.* 37, 739–748. doi: 10.1037/0012-1649.37.6.739
- Christiansen, M. H., and Chater, N. (2016). The now-or-never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* 39:e62. doi: 10.1017/S0140525X1500031X
- Comrie, B. (1976). *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge, NY: Cambridge University Press.
- Conklin, K., and Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Appl. Ling.* 29, 72–89. doi: 10.1093/applin/amm022
- Corder, S. P. (1967). The significance of learner’s errors. *Int. Rev. Appl. Ling. Lang. Teach.* 5, 161–170. doi: 10.1515/iral.1967.5.1-4.161
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. Available online at: <https://www.english-corpora.org/coca/> (accessed December, 2017).
- Dowty, D. R. (1991). *Word meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague’s PTQ*. Dordrecht: Kluwer Academic.
- Ellis, N. C. (1996). Sequencing in SLA: phonological memory, chunking, and points of order. *Stud. Sec. Lang. Acq.* 18, 91–126. doi: 10.1017/S0272263100014698
- Ellis, N. C. (2002). Frequency effects in language processing. *Stud. Sec. Lang. Acq.* 24, 143–188. doi: 10.1017/S0272263102002024
- Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Appl. Ling.* 27, 1–24. doi: 10.1093/applin/ami038
- Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Appl. Ling.* 27, 164–194. doi: 10.1093/applin/aml015
- Ellis, N. C. (2011). “Frequency-based accounts of SLA,” in *Handbook of Second Language Acquisition*, eds S. Gass and A. Mackey (London: Routledge), 193–210. doi: 10.1515/9783110550443-022
- Ellis, N. C. (2012a). “What can we count in language, and what counts in language acquisition, cognition, and use?,” in *Frequency Effects in Language Learning and Processing*, eds S. T. Gries and D. Divjak (Berlin: De Gruyter), doi: 10.1515/9783110274059.7
- Ellis, N. C. (2012b). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annu. Rev. Appl. Ling.* 32, 17–44. doi: 10.1017/S0267190512000025
- Ellis, N. C. (2017). “Chunking,” in *The Changing English Language: Psycholinguistic Perspectives*, eds M. Hundt, S. Mollin, and S. Pfenninger (Cambridge: Cambridge University Press), 113–147.
- Ellis, N. C., Römer, U., and O’Donnell, M. B. (2016). *Usage-Based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Malden, MA: Wiley.
- Ellis, N. C., and Sagarra, N. (2010). The bounds of adult language acquisition. *Stud. Sec. Lang. Acq.* 32, 553–580. doi: 10.1017/S0272263110000264
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402\_1
- Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: an empirical validation study. *Appl. Ling.* 27, 464–491. doi: 10.1093/applin/aml001
- Erman, B., and Warren, B. (2000). The idiom principle and the open choice principle. *Text* 20, 29–62. doi: 10.1515/text.1.2000.20.1.29
- Finley, S. (2015). “Frequency effects in morpheme segmentation,” in *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, (Austin, TX: Cognitive Science Society), 704–709.
- Finley, S. (2018). Cognitive and linguistic biases in morphology learning. *WIREs Cogn. Sci.* 9:e1467. doi: 10.1002/wcs.1467
- Forster, K. I., and Chambers, S. M. (1973). Lexical access and naming time. *J. Verb. Learn. Verb. Behav.* 12, 627–635. doi: 10.1016/S0022-5371(73)80042-8
- Gaillard, S., and Tremblay, A. (2016). Linguistic proficiency assessment in second language acquisition research: the elicited imitation task. *Lang. Learn.* 66, 419–447. doi: 10.1111/lang.12157
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, CA: University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldschneider, J. M., and DeKeyser, R. M. (2001). Explaining the “Natural Order of L2 morpheme acquisition” in English: a meta-analysis of multiple determinants. *Lang. Learn.* 51, 1–50. doi: 10.1111/1467-9922.00147
- Gries, S. T. (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction. 1. publ.* New York, NY: Routledge.
- Gries, S. T. (2015). More (old and new) misunderstandings of collocation analysis: on Schmid and Küchenhoff (2013). *Cogn. Ling.* 26, 505–536. doi: 10.1515/cog-2014-0092
- Gries, S. T., and Ellis, N. C. (2015). Statistical measures for usage-based linguistics: statistical measures for usage-based linguistics. *Lang. Learn.* 65, 228–255. doi: 10.1111/lang.12119
- Gries, S. T., and Stefanowitsch, A. (2004). Extending collocation analysis: a corpus-based perspective on “alternations”. *Int. J. Corpus Ling.* 9, 97–129. doi: 10.1075/ijcl.9.1.06gri

- Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics* 39, 1041–1070. doi: 10.1515/ling.2001.041
- Hoffmann, T., and Trousdale, G. (eds). (2013). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780195396683.001.0001
- Janssen, N., and Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS One* 7:e33202. doi: 10.1371/journal.pone.0033202
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. Routledge.
- Jia, G., and Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: age-related differences. *J. Speech Lang. Hear. Res.* 50, 1280. doi: 10.1044/1092-4388(2007)090
- Jiang, N. (2004). Morphological insensitivity in second language processing. *Appl. Psycholing.* 25, 603–634. doi: 10.1017/S0142716404001298
- Jiang, N., and Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Mod. Lang. J.* 91, 433–445. doi: 10.1111/j.1540-4781.2007.00589.x
- Krashen, S., Houck, N., Giunchi, P., Bode, S., Birnbaum, R., and Strei, G. (1977). Difficulty order for grammatical morphemes for adult second language performers using free speech. *TESOL Q.* 11:338.
- Larsen-Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Lang. Learn.* 26, 125–134. doi: 10.1111/j.1467-1770.1976.tb00264.x
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Lim, H., and Godfroid, A. (2015). Automatization in second language sentence processing: a partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Appl. Psycholinguist.* 36, 1247–1282. doi: 10.1017/S0142716414000137
- Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Percept. Psychophys.* 39, 155–158. doi: 10.3758/BF03212485
- Luk, Z. P., and Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural -s, articles, and possessive 's. *Lang. Learn.* 59, 721–754. doi: 10.1111/j.1467-9922.2009.00524.x
- MacWhinney, B. (1987). Applying the competition model to bilingualism. *Appl. Psycholing.* 8, 315–327. doi: 10.1017/S0142716400000357
- MacWhinney, B. (1997). Second language acquisition and the competition model. *Tutor. Bilingual.* 113, 142.
- MacWhinney, B. (2001). "The competition model: the input, the context, and the brain," in *Cognition and Second Language Instruction*, ed. P. Robinson (Cambridge: Cambridge University Press), 69–90. doi: 10.1017/CBO9781139524780.005
- MacWhinney, B., and Bates, E. (1989). *The Crosslinguistic Study of Sentence Processing*. Cambridge: Cambridge University Press.
- MacWhinney, B., and O'Grady, W. (2014). *The Handbook of Language Emergence*. Hoboken, NJ: Wiley-Blackwell.
- Marchman, V. A. (1997). Children's productivity in the English past tense: the role of frequency, phonology, and neighborhood structure. *Cogn. Sci.* 21, 283–304. doi: 10.1207/s15516709cog2103\_2
- Marchman, V. A., Wulfeck, B., and Weismer, S. E. (1999). Morphological productivity in children with normal language and sli: a study of the English past tense. *J. Speech Lang. Hear. Res.* 42, 206–219. doi: 10.1044/jslhr.42.01.206
- Matthews, D. E., and 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. *Cogn. Sci.* 30, 1027–1052. doi: 10.1207/s15516709cog0000\_66
- McClelland, J. L., and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88, 375–407. doi: 10.1037/0033-295X.88.5.375
- McDonald, A. S., and Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychol. Sci.* 14, 648–652. doi: 10.1046/j.0956-7976.2003.psci.1480.x
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90, 91–117. doi: 10.1016/S0010-0277(03)00140-9
- Mintz, T. H., Wang, F. H., and Li, J. (2014). Word categorization from distributional information: frames confer more than the sum of their (Bigram) parts. *Cogn. Psychol.* 75, 1–27. doi: 10.1016/j.cogpsych.2014.07.003
- Murakami, A., and Alexopoulou, T. (2016). L1 influence on the acquisition order of English grammatical morphemes. *Stud. Sec. Lang. Acq.* 38, 365–401. doi: 10.1017/S0272263115000352
- Naigles, L. R., and Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *J. Child Lang.* 25, 95–120. doi: 10.1017/s0305000997003358
- Ortega, L., Iwashita, N., Norris, J. M., and Rabie, S. (2002). "An investigation of elicited imitation tasks in crosslinguistic SLA research," in *Paper presented at the Second Language Research Forum*, (Toronto, ON).
- Pawley, A., and Syder, F. (1983). "Two puzzles for linguistic theory: Nativelike selection and nativelike fluency," in *Language and Communication*, eds J. C. Richards and R. W. Schmidt (Milton Park: Taylor & Francis), 34.
- Peirce, J. W. (2007). PsychoPy: psychophysics software in python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Pollatsek, A., Treiman, R., Yap, M. J., and Balota, D. A. (2015). "Visual word recognition," in *The Oxford Handbook of Reading*, eds A. Pollatsek and R. Treiman (Oxford: Oxford University Press), doi: 10.1093/oxfordhb/9780199324576.013.4
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Räsänen, S. H. M., Ambridge, B., and Pine, J. M. (2016). An elicited-production study of inflectional verb morphology in child Finnish. *Cogn. Sci.* 40, 1704–1738. doi: 10.1111/cogs.12305
- Redington, M., Chater, N., and Finch, S. (1998). Distributional information: a powerful cue for acquiring syntactic categories. *Cogn. Sci.* 22, 425–469. doi: 10.1207/s15516709cog2204\_2
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *J. Exp. Psychol.* 81, 275–280. doi: 10.1037/h0027768
- Ren, Y. (2011). A study of the washback effects of the college English test (band 4) on teaching and learning English at tertiary level in China. *Int. J. Pedagogies Learn.* 6, 243–259. doi: 10.5172/ijpl.2011.6.3.243
- Ross, C. (1995). Temporal and aspectual reference in Mandarin Chinese. *J. Chin. Ling.* 23, 87–136.
- Schmid, H.-J. (2017). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge. 1*. Washington, DC: American Psychological Association.
- Schmid, H.-J. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. New York, NY: Oxford University Press.
- Schmitt, N. (2004). *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam: John Benjamins Publishing Company.
- Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychol. Rev.* 96, 523–568. doi: 10.1037/0033-295X.96.4.523
- Seidenberg, M. S., and Plaut, D. C. (2014). Quasiregularity and its discontents: the legacy of the past tense debate. *Cogn. Sci.* 38, 1190–1228. doi: 10.1111/cogs.12147
- Sereno, J. A., and Jongman, A. (1997). Processing of English inflectional morphology. *Mem. Cogn.* 25, 425–437. doi: 10.3758/BF03201119
- Shanks, D. R. (1995). *The Psychology of Associative Learning*. Cambridge: Cambridge University Press.
- Sinclair, J. (1995). *Corpus, Concordance, Collocation. 3. impr.* Oxford: Oxford University Press.
- Siyanova-Chanturia, A., and Pellicer-Sanchez, A. (2018). *Understanding Formulaic Language: A Second Language Acquisition Perspective*. New York, NY: Routledge.
- Slobin, D. I. (1996). "From 'thought and language' to 'thinking for speaking,'" in *Rethinking Linguistic Relativity*, eds J. Gumperz and S. Levinson (Cambridge: Cambridge University Press).
- Stefanowitsch, A., and Gries, S. T. (2003). Collocations: Investigating the interaction of words and constructions. *IJCL* 8, 209–243. doi: 10.1075/ijcl.8.2.03ste
- Sugaya, N., and Shirai, Y. (2009). "Can L2 learners productively use Japanese tense-aspect markers? A usage-based approach," in *Typological Studies in Language*,

- eds R. Corrigan, E. A. Moravcsik, H. Ouali, and K. Wheatley (Amsterdam: John Benjamins Publishing Company), 423. doi: 10.1075/tsl.83.10sug
- Tatsumi, T., Ambridge, B., and 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. *Cogn. Sci.* 42, 555–577. doi: 10.1111/cogs.12554
- Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks: lexical bundle processing. *Lang. Learn.* 61, 569–613. doi: 10.1111/j.1467-9922.2010.00622.x
- Tsimpli, I. M., and Dimitrakopoulou, M. (2007). The interpretability hypothesis: evidence from wh-interrogatives in second language acquisition. *Sec. Lang. Res.* 23, 215–242. doi: 10.1177/0267658307076546
- Wheeler, D. D. (1970). Processes in word recognition. *Cogn. Psychol.* 1, 59–85. doi: 10.1016/0010-0285(70)90005-8
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wulff, S. (2019). “Acquisition of formulaic language from a usage-based perspective,” in *Understanding Formulaic Language: A Second Language Acquisition Perspective*, eds A. Siyanova-Chanturia and A. Pellicer-Sánchez (New York, NY: Routledge).
- Wulff, S., and Ellis, N. C. (2018). “Usage-based approaches to second language acquisition,” in *Studies in Bilingualism*, eds D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Amsterdam: John Benjamins Publishing Company), 37–56. doi: 10.1075/sibil.54.03wul
- Xu, F., and Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychol. Rev.* 114, 245–272. doi: 10.1037/0033-295X.114.2.245
- Yap, M. J., and Balota, D. A. (2014). “Visual word recognition,” in *The Oxford Handbook of Reading*, eds A. Plllatsek and R. Treiman (Oxford University Press), 26–43.
- Yeh, Y.-F., Joshi, R. M., and Ji, X. R. (2015). The development of morphological awareness in Chinese ESL students. *Contem. Educ. Psychol.* 43, 51–60. doi: 10.1016/j.cedpsych.2015.10.002

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Guo and Ellis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.