To appear, *Language Learning*
RUNNING HEAD: L2 Learners and Competition Dynamics

Nonnative Speakers Do Not Take Competing Alternative Expressions into Account the Way Native Speakers Do

Clarice Robenalt and Adele Goldberg
Princeton University

When native speakers judge the acceptability of novel sentences, they appear to implicitly take competing alternative formulations into account, judging novel sentences with a readily available alternative formulation to be less acceptable than novel sentences with no competing alternative. Moreover, novel sentences with a competing alternative are more strongly dispreferred when they contain high frequency verbs than when they contain lower frequency verbs. The present study replicates these findings with a group of native speakers and extends the paradigm to second language (L2) users. Recent research suggests that L2 users are less able to generate online expectations during language processing, which in turn predicts a reduced ability to differentiate between novel sentences with and without a competing alternative. We test this prediction and confirm that while L2 speakers display evidence of learning from positive exemplars, they show no evidence of taking competing grammatical alternatives into account, except at the highest quartile of speaking proficiency, in which case L2 judgments align with native speakers.

Keywords
Indirect negative evidence; prediction; argument structure

Author Note
We are very grateful to Pavel Trofimovich for his advice as associate editor and also to three anonymous referees for this journal. We are also appreciative of Bodo Winter and Stefan Gries' statistical advice, and Einstein Foundation of Berlin for funding this project. The elicitation instruments used for the study can be accessed by readers in the IRIS digital repository (http://www.iris-database.org).

Correspondence concerning this article should be addressed to: Clarice Robenalt or Adele Goldberg, Princeton University, Department of Psychology, Peretsman-Scully Hall, Princeton, New Jersey United States 08542. E-mail: robenalt@princeton.edu; adele@princeton.edu.

Introduction

It is well known that adult second language (L2) learners rarely reach a level of speaking competence akin to that of native speakers (e.g., Birdsong & Molis, 2001; Flege, Yeni-Komshian, & Liu, 1999). One type of error that L2 learners commonly make involves producing utterances that are semantically sensible and make use of constructions that are licensed in the target language, but which are nonetheless avoided by native speakers (e.g., Bley-Vroman & Joo, 2001; Bley-Vroman & Yoshinaga, 1992; Hubbard & Hix, 1988; Inagaki, 1997; Oh, 2010). Examples of such sentences formulated by nonnative speakers of English are provided in (1)–(4). In these and following examples, underlined text identifies the phrase that is dispreferred by native speakers, and "??" is used to indicate that native speakers disfavor the formulation to some extent. We use question marks instead of the more traditional asterisk to indicate that judgments are gradient and subject to a number of factors, as quantified below.

Examples (1)-(4) are from Google.

(1) ?? Could you recommend me some English movies …
(2) ?? Can you explain me the lesson.
(3) ?? Have you ever considered to go climbing in Ecuador… we would love for you to come and discover it…
(4) ?? The afraid boy slowly took out the note.

Each of these errors involves using a predicate (verb or adjective) in a construction in which it does not normally occur. Some of these errors may be motivated by transfer from the speakers' native language (L1) (e.g., Ambridge & Brandt, 2013; Bley-Vroman & Joo, 2001; Rutherford, 1989). However, transfer effects are not necessarily responsible for this type of error, since these constructions are all amply attested in English. This is made clear by the minimally different but fully acceptable examples in (5)–(8) (see also Flett, Branigan, & Pickering, 2013):

(5) Could you give/offer/tell me some place to apply?
(6) Maybe it's better to tell/show me first.
(7) Have you ever wanted/hoped/tried/planned to go climbing in Ecuador?
(8) The astute/scared boy.

Moreover, since not everything from a L1 is transferred to a L2, it is reasonable to ask why these particular types of errors are so portable. As Hubbard and Hix (1988) note, "[intermediate and advanced ESL learners] have often mastered the majority of the major syntactic constructions in English. Yet, many grammatical errors persist… [involving] mistakes in using given lexical items in constructions they do not belong in" (p. 89).

In fact, the ill-formedness of the phrases like those in (1)–(4) have been argued to present a learnability paradox for L1 and L2 learners alike. Even when a native speaker has never encountered a particular utterance, he or she can instantly intuit whether the utterance is grammatically acceptable or not. This is especially interesting in cases where a non-occurring pattern is judged to be unacceptable despite being semantically sensible and involving a construction that is well attested in the target language (e.g., ??*explain me first*). Indeed, it has remained an outstanding challenge for linguists, despite more than four decades of research, to determine how learners come to know which combinations of verb and construction are acceptable and which are not (e.g., Ambridge, Pine, Rowland, & Young, 2008; Baker, 1979; Bowerman, 1988; Braine, 1971; Brown & Hanlon, 1970; Goldberg, 1995; Pinker, 1989; Gennari & Macdonald, 2008). Language input clearly does not come annotated with asterisks or question marks to indicate unacceptability. Furthermore, ill-formed utterances are not reliably corrected in naturalistic speech settings since people are generally more interested in the content of speech than in its form (Brown & Hanlon, 1970; Hirsh-Pasek, Treiman, & Schneiderman, 1984; Marcus, 1993). So how do people come to disfavor some—but not all—non-occurrences in their input? That is, how do learners come to know which novel combinations of verb and construction are acceptable and which are not?

Construction Learning from Language Input

A growing body of work indicates that speakers take advantage of the statistics in the input in order to learn language, both in L1 (e.g., Bybee, 1985, 1995, 2010; Langacker, 1988; Tomasello, 2003) and in L2 (e.g., Ellis, 2002; Ellis & Larsen-Freeman, 2006; Eskildsen, 2009; Trofimovich, Collins, Cardoso, White, & Horst, 2012; Tyler, 2008; Weinert, 1995). Such usage-based models assume that speakers retain memory traces of how verbs and other words have been heard used, generalizing these memory traces so that information about the frequencies of particular usage patterns constitute part of our knowledge of language. In the case of argument structure patterns, verbs are recognized to be associated with valency information (e.g., Herbst, 2014), while networks of individual verbs give rise to more general argument structure constructions (e.g., Goldberg, 1995, 2006; Tomasello, 2003). As there is ample evidence that frequency information plays an important role in language learning and language processing, we do not attempt to review it here. However, frequency information can play a more nuanced role than a simple counting of exemplars that occur, in that cases of *non*-occurrence can be informative to a language learner.

Of central interest in the present context is whether both L1 and L2 learners take advantage of a lack of positive exemplars during language learning. If learners implicitly take utterance types that they have not previously heard into account, it can provide a type of indirect negative evidence—that is, evidence of what is not acceptable in the language. In particular, a process of statistical preemption has been argued to play an important role in L1 learning to avoid certain formulations, such as those in (1)–(4). That is, the non-occurrence of a particular form can serve as evidence that the form is unacceptable if the form could reasonably have been expected to occur and did not. For example, if a learner repeatedly hears the formulation B in contexts where he or she might have expected to hear the semantically and pragmatically related alternative formulation A, he or she would come to recognize that B is the appropriate formulation in such a context and that A is less acceptable. This is statistical preemption (Bates & MacWhinney, 1987; Boyd & Goldberg, 2011; Clark, 1987; DiSciullo & Williams, 1987; Foraker, Regier, Khetarpal, Perfors, & Tenenbaum, 2007; Goldberg, 1993, 1995, 2006, 2011; Marcotte, 2005; Pinker, 1981): a type of error-driven learning (Rescorla & Wagner, 1972; Rumelhart, Hinton, & Williams, 1986).

A concrete example is useful for illustration. We know that different constructions typically have somewhat different semantics and/or distributions in discourse contexts. For example, the double-object construction (e.g., *tell me something*) tends to be used when the recipient argument is topical and pronominal, and the lexical theme argument is newer to the discourse; the to-dative allows the theme to be pronominal and the recipient to be newer to the discourse (e.g., *tell it to Zach*) (Arnold, Wasow, Losongco, & Ginstrom, 2000; Bresnan, Cueni, Nikitina, & Baayen, 2007; Collins, 1995; Erteschik-Shir, 1979; Goldberg, 1995, 2006; Green, 1974; Hovav & Levin, 2008; Oehrle, 1976; Thompson, 1995; Wasow, 2002). If, in these situations, the to-dative is consistently witnessed instead, a learner can learn that the double-object is not, after all, appropriate (Goldberg, 1995, 2006). For example, hearing *explain something to me* serves to statistically preempt *explain* in the double-object construction (e.g., ??*explain me something*). As Goldberg (2006) emphasizes, the process is necessarily statistical, because a single use of the alternative formulation could be due to some subtle difference in the functions of the two formulations that actually favors the alternative formulation in that context. It is also possible that a single use could simply be due to an error by the speaker. But if an alternative formulation is consistently heard when another formulation is expected, a process of statistical preemption predicts that speakers will learn to use the provided alternative.

Statistical preemption of phrasal forms has only received attention in the experimental literature fairly recently. Two studies have shown that witnessing novel intransitive verbs in periphrastic causative constructions significantly preempts children's use of those verbs in simple transitives (Brooks & Tomasello, 1999; Brooks & Zizak, 2002). For example, Brooks and Tomasello found that children aged six and seven were less than half as likely to productively use a novel verb causatively when the verb had been modeled in both the intransitive and in a periphrastic causative, compared to when it was only modeled in the intransitive form. For example, if the children had heard both *The cow is chamming* and *Ernie's making the cow cham*, they were less likely to respond to "what did Elmo do to the cow?" with *Ernie chammed the cow* (causative) than they were if only the intransitive construction had been witnessed. Hearing the novel verb used in the periphrastic causative construction provided a readily available alternative to the causative construction, statistically preempting the use of the latter (see also Tomasello, 2003). In another study, Boyd and Goldberg (2011) investigated a certain class of adjectives that begin with schwa and disfavor appearing prenominally (e.g., ??*the asleep boy*). Adult productions were elicited in a naturalistic situation, resulting in either a relative clause (*The fox that's sleepy/asleep/adax moved to the star*) or prenominal use of a targeted real or novel adjective (*The sleepy/*??*asleep/*?*adax fox moved to the star*). Results showed that witnessing a novel adjective used in a preemptive context (the relative clause) dramatically decreased prenominal uses.

A number of additional studies have found that, all other things being equal, lower frequency verbs are more available for creative extensions than are high frequency verbs. For example, both *disappear* and *vanish* are normally used intransitively, with *disappear* being more frequent than *vanish*. Native speakers are more resistant to extending *disappear* for use in a transitive construction than they are *vanish* (??*She disappeared something* vs. ?*She vanished something*). This is presumably because *disappear* has been witnessed in a competing alternative more often than *vanish*. More specifically, in a production study, Brooks, Tomasello, Dodson, and Lewis (1999) found that children aged 3–8 were more likely to extend a low frequency verb to a previously unwitnessed construction than they were to extend a high frequency verb. Similar results were found in a judgment task by Theakston (2004), who elicited acceptability ratings from adults and children aged 5 and 8. Across ages, novel sentences containing low frequency verbs were judged to be more acceptable than sentences containing high frequency verbs. For example, *She giggled her* was judged to be more acceptable than *She laughed her*, although presumably neither the low frequency *giggle* nor the high frequency *laugh* had been witnessed being used transitively. The same result was replicated in a judgment task by Ambridge et al. (2008) with children (aged 5–6 and 9–10) and adults. This effect is remarkable since, due to general fluency effects, sentences with higher frequency words can be expected to be judged more favorably than semantically equivalent sentences with lower frequency words (e.g., Bornstein & D'Agostino, 1992). For example, *Zach sat on the sofa* can be expected to be judged more acceptable than *Zach reclined on the divan*.

The idea that speakers may be capable of recognizing that certain constructions have not occurred with certain verbs raises the question of whether all novel pairings of verb and construction are equivalent. If there are independent reasons why a verb has not been witnessed in a particular construction, speakers should not necessarily assume that the verb and construction are incompatible. A primary reason why a particular verb may not yet have occurred in a particular construction is that the combined meaning may be implausible or not generally relevant. For example, *sneeze* may not be commonly witnessed in the "caused motion" construction simply because it is unusual for a sneeze to cause something to move. But it is not impossible, as the attested examples in (9) from Google illustrate:

(9a) I just sneezed the cat into the trash can;
(9) b. an Italian man was shot in the head and sneezed the bullet out

In fact, one of the most remarkable aspects of language is that speakers do produce sentences that they have never heard before. In English, even fairly frequent verbs are sometimes used in new ways, as illustrated in the following examples (which are similar to the examples with *sneeze* above):

(10) I grabbed the guy [and] popped him a punch that sent him sprawling onto the poster spread (COCA corpus, Davies, 2009).
(11) Some part of me wanted to yell my father into the world (Google)
(12) Women have dribbled and rowed their way into sporting history (COCA corpus Davies, 2009).

Therefore, simply not having heard a particular sentence should not lead learners to conclude that the formulation is necessarily unacceptable.

To date, most of the work on statistical preemption has been done using native speakers, as described above (see also Ambridge, Pine, & Rowland, 2012; Ambridge, Pine, Rowland, & Chang, 2012; Ambridge, Pine, Rowland, Freudenthal, & Chang, 2012). A recent exception is Ambridge and Brandt (2013), who investigated acceptability judgments from L1 German/L2 English speakers on English locatives, such as the familiar *She filled the cup with water* and the novel and disfavored ??*She filled water into the cup*. Ambridge and Brandt used mixed effects modeling to estimate the relative contribution of different predictors to the acceptability ratings for nonnative speakers. They found that positive evidence of a particular verb use served to boost the acceptability of that same use, but they found no evidence that verb frequency had a negative impact on other novel uses of the verb. This result suggests that L2 learners may not implicitly compare the novel use of a verb with a competing use of that verb—as is needed for statistical preemption. At the same time, the relatively small sample size (30 nonnative speakers) leaves open the possibility that the lack of effect was due to a lack of power. Furthermore, recent findings indicate that there might be reason to expect nonnative speakers to treat novel utterances differently than native speakers in a systematic way. In the following section, we detail how a critical difference between native and nonnative language processing may influence how the different populations perceive novel sentences.

Language Processing in Nonnative Speakers

L2 learners' persistent errors in the use of argument structure patterns may have its roots, in part, in an intriguing finding from recent L2 literature. Various researchers have reported that L2 learners are less likely than native speakers to predict upcoming forms during online comprehension, even when they demonstrate knowledge of the constraints used to predict the forms in production in off-line tasks. Grüter, Hurtado, Marchman, and Fernald (2014) refer to this as the "RAGE hypothesis," where RAGE stands for Reduced Ability to Generate Expectations during language processing. The RAGE hypothesis may provide some insights into nonnative difficulties with argument structure restrictions. If nonnative speakers are less able to predict upcoming utterances, then they lose the opportunity to learn from the prediction error that would normally occur when the observed utterance does not match the predicted utterance.

In key studies that helped inspire the RAGE hypothesis, Lew-Williams and Fernald (2007, 2010) found that native speakers actively use gender information to predict the upcoming referent in a spoken utterance. Using eye-tracking to record predictive fixations, Lew-Williams and Fernald (2007) demonstrated that child and adult native speakers of Spanish were able to predict the correct referent based on the gender information provided in the preceding determiner. However, native English speakers who learned L2 Spanish did not show evidence of using gender information; instead, they waited until the noun label was provided before looking to the correct referent, even though they demonstrated knowledge of the objects' grammatical gender in off-line tasks. Lew-Williams and Fernald (2010) found that this difference between L1 and L2 speakers held even when they used newly learned novel nouns to control for experience with the lexical items (Lew-Williams & Fernald, 2010; see also Grüter, Lew-Williams, & Fernald, 2012).

Martin et al. (2013) similarly found evidence that L2 learners of English failed to predict upcoming words to the same extent as native speakers when reading sentences that strongly biased a particular final word. The authors recorded event-related potentials (ERPs) from native English speakers and L2 learners while they read sentences in which the final noun started either with a vowel or a consonant, and which was either expected or unexpected (but always plausible):

- Since it is raining, it is better to go out with an umbrella (expected: vowel).
- Since it is raining, it is better to go out with a raincoat (unexpected: consonant).

- She has a nice voice and always wanted to be a singer (expected: consonant).
- She has a nice voice and always wanted to be an artist (unexpected: vowel).

As predicted by earlier work (DeLong, Urbach, & Kutas, 2005), native speakers showed an increase in the N400—a general index of semantic unexpectedness—at the article (*an* or *a*) when the following noun was unexpected. L2 learners, however, did not, even though they used the two articles correctly and shared the same off-line judgments of which nouns were expected and unexpected in the given sentential contexts.

The use of articles in these experiments to determine whether speakers are anticipating upcoming words is not accidental. It is clear that all fluent speakers, whether native or nonnative, actively integrate semantic information as soon as it is encountered. While the articles themselves do not provide relevant semantic information in the contexts of these experiments, they allow comprehenders to create predictions (Lew-Williams & Fernald, 2010) or verify predictions (Martin et al., 2013) about which words will be coming next. The experiments just described indicate that L2 learners may not actively predict upcoming linguistic material to the same degree that L1 learners do.

We need not assume that a reduced ability to generate expectations results in a qualitative or permanent difference between L1 and L2 speakers. Individual differences, degree of proficiency, the relationship between L1 and L2, and task demands undoubtedly play a role in the extent to which predictions about upcoming forms are made by L2 comprehenders (e.g., Borovsky, Elman, & Fernald, 2012; Kaan, 2014). Lexical access is typically delayed in the L2, so it is possible that L2 learners simply do not have time to make predictions about upcoming words because they are still attempting to process their current input. Alternatively, to the extent that cognitive resources are required to inhibit an L2 comprehender's L1, resources will be correspondingly less available for non-essential on-line predictive processes (Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009; Linck, Kroll, & Sunderman, 2009; Martin et al., 2013; Rossi, Kroll, & Dussias, 2014). At very high proficiency levels, Hopp (2013) has found that English-German L2 speakers do appear to exhibit predictive processing of grammatical forms. Predictive abilities also appear to be enhanced when parallel constructions exist across the L1 and L2, as demonstrated by Foucart, Martin, Moreno, and Costa (2014), who found that Spanish-French bilinguals (who have grammatical gender in both their L1 and L2) also show anticipation at the gender-marked article in their L2 French. It is also important to note that reduced anticipatory processing is not likely to be exclusive to L2 learners. Evidence of a decline in such processing has been found in older native speakers as well (DeLong, Groppe, Urbach, & Kutas, 2012).

Nonetheless, to the extent that L2 learners' ability to predict upcoming grammatical forms is reduced, the effectiveness of statistical preemption should be correspondingly reduced. That is, if nonnative speakers do not anticipate upcoming utterances to the same extent as native speakers, they will have less opportunity to learn from predictions that are subsequently corrected. Concretely, nonnative speakers should show less ability to distinguish novel sentences with a competing conventional alternative and novel uses without a readily available alternative. Another anticipated difference between native and nonnative speakers concerns the role of verb frequency on perceptions of acceptability. Statistical preemption predicts that verb frequency should inversely influence the acceptability for novel sentences that have a competing alternative, but not those that do not. If nonnative speakers do not compare the novel use of a verb with a conventional use that competes with the novel use, they should not show the same effects of verb frequency in novel sentences as native speakers do. The following study sets out to test these predictions.

The Present Study

Relevant to the present study, statistical preemption predicts that learners should treat novel sentences that have a competing alternative differently than those that do. Using a set of novel sentences that either had or did not have an agreed-upon competing alternative phrasing, Robenalt and Goldberg (2015) collected acceptability judgments for two versions of each sentence that involved a low frequency verb and a high frequency verb. Example stimuli are provided in Table 1.

Table 1 Example stimuli from Robenalt and Goldberg (2015) that varied whether novel sentence types had a readily available alternative and whether the main verb was high or low frequency

Description	High frequency verb	Low frequency verb
Has competing alterative formulation (in parentheses)	(a) Amber explained Zach the answer (cf. Amber explained the answer to Zach).	(b) Amber recited Zach the answer (cf. Amber recited the answer to Zach).

No readily available competing alternative formulation	(c) Megan smiled her boyfriend out the front door.	(d) Megan grinned her boyfriend out the front door.

Statistical preemption predicts that verb frequency should only matter when a sentence has a readily available competing formulation, and this was found to be the case. For sentences with a competitor, native speakers disfavored high frequency verbs more strongly than low frequency verbs, a result that is consistent with the previous studies described above (e.g., Ambridge et al., 2008; Brooks & Tomasello, 1999; Theakston, 2004). At the same time, for sentences with no competitor, the high and low frequency versions of the sentences were judged to be equally acceptable. That is, while speakers judged sentences like (a) in Table 1 to be less acceptable than those like (b), they judged sentences like (c) and (d) to be equivalent. Furthermore, native English speakers perceived novel sentences that had a competing alternative (e.g., a and b) to be less acceptable than those that did not (c and d). These results indicate that native speakers are highly sensitive to the presence of competing alternative phrasing. When speakers already have a conventional way to express a particular meaning, they judge novel re-formulations of that meaning to be less acceptable. However, when no competitor exists, speakers display a willingness to extend a verb to a construction in which it does not normally appear. In the present study, we aim to replicate Robenalt and Goldberg's findings (2015) and, crucially, extend the study to L2 learners of English.

Method

Participants

Native speakers of English (*n* = 157) and adult learners of English (*n* = 276) were recruited from Amazon's Mechanical Turk (https://www.mturk.com). Nonnative speakers completed the same task as the native speakers, provided additional biographical data on their language background, and self-rated their proficiency in speaking, reading, and writing in English. The mean age of first exposure to English was 12.9 years (*SD* = 3.9 years, 5–56). Of the 276 nonnative speakers of English, 63 participants' L1 was Tamil, 55 participants' L1 was Spanish, and 33 participants' L1 was Hindi. The remaining balance included 1–17 speakers of 36 other languages. Computer IP addresses were restricted to those within the United States. Each participant was modestly reimbursed for their participation, which lasted approximately 10 minutes.

Materials

Fourteen verb classes were selected, and one high and low frequency alternative were identified for each class (14 verb pairs; 28 total verbs). High frequency verbs were, on average, 15.74 times more frequent than their low frequency counterparts according to verb frequencies from the Corpus of Contemporary American English (COCA, Davies, 2009). Two sentence frames were developed for each verb pair, such that one sentence frame exemplified the typical use of the verb while the other exemplified an atypical use of the verb. For example, one pair of verbs involved *disappear* (high frequency) and *vanish* (low frequency). Since each of these verbs is typically used intransitively, the first sentence frame was intransitive (e.g., *Ashley disappeared/vanished*). These familiar sentence frames served as a baseline measure of acceptability for each verb and are referred to as baseline sentences (see Appendix S1 in Supporting Information online for a full list of the sentence pairs used in this study).

The novel sentence frame was constructed such that the formulation was semantically sensible, but extremely atypical; the argument structure construction was one that the verbs rarely, if ever, appear in. For example, the atypical sentence frame for *disappear* and *vanish* was transitive (e.g., *The magician disappeared/vanished the rabbit*). Atypical uses were confirmed to be novel by searches in the full 450-million-word COCA corpus. As the corpus has been tagged for parts of speech, but not parsed for syntactic structure, all search results were checked by hand. Novel uses never accounted for more than .12% of the total instances of the verb in question and were completely unattested for eight verbs (for details, see Appendix S2 in Supporting Information online). In a separate norming study reported in Robenalt and Goldberg (2015), all novel sentences were presented to a separate group of 20 native speakers, who were asked whether they could think of a better way to express the intended meaning of the sentence. If they chose to paraphrase the sentence, they were instructed to enter their response in a free form text box. If they felt the sentence did not need to be improved, they were asked to copy and paste the sentence exactly as written.[1] Responses were analyzed by two independent raters and coded as exact repetitions, paraphrases, or other. "Other" sentences included misinterpretations of the sentence meaning or ungrammatical formulations and accounted for 6% of the norming data. Coders agreed on 92% of the responses; disagreements were arbitrated by a third coder.

The paraphrase responses for each sentence were examined for consistency and used to bin the sentences into two categories. For half of the sentences, 50% or more of the paraphrases were identical, suggesting that independent raters shared an intuition about a better way to express the meaning of the novel utterance. Such sentences were said to have a competing alternative (henceforth, novel/hasCA). The remainder of the sentences showed little consensus within the paraphrases, suggesting that there was no standard alternative way to express the meaning of the novel utterance. These sentences were said to have no competing alternative (novel/noCA). Figure 1 plots the proportion of total responses that corresponds to the most commonly shared paraphrase for each sentence. Taller bars indicate higher paraphrase consistency across the independent raters. The color coding of the bars indicates the result of a median split that was used to divide the sentences into the two categories (hasCA vs. noCA).

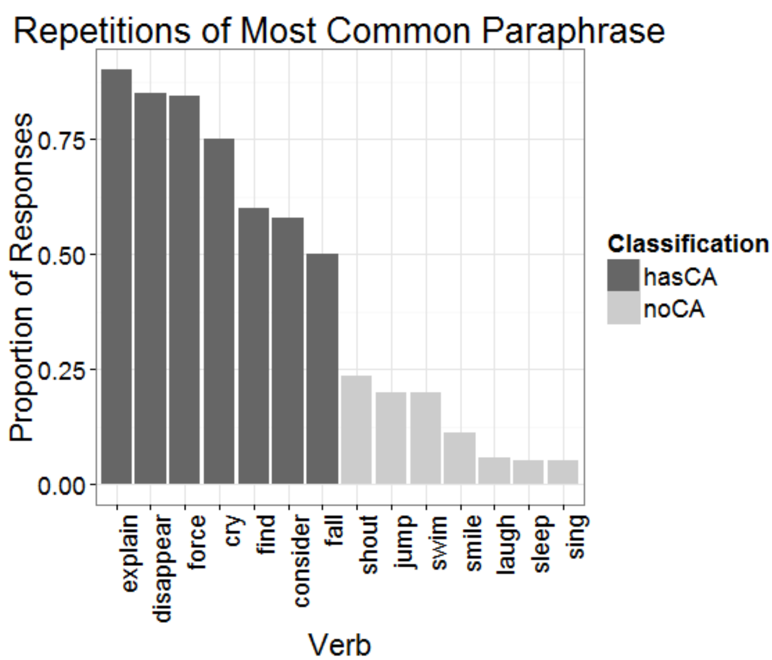

Figure 1 Most common paraphrases for novel sentences. Individual bars represent the proportion of responses that corresponded to the most common paraphrase for that sentence, with verbs used to index the sentence stimuli (see Appendix S1 in Supporting Information online for a complete list of stimuli). Only high frequency verbs were used to create the hasCA/noCA bins; the low frequency counterpart was assigned to the same category based on the median split for the high frequency items. Figure copied by permission from Robenalt and Goldberg (2015).

We did not attempt to gather numbers for exactly how often each verb occurs in its competing alternative, because whether or not the alternative formulation truly competes depends on contextual factors that would require prohibitively time-consuming hand-coding of tens of thousands of examples. For example, a sentence like *She made him cry* competes with a simple causative *??She cried him*, only if the causation is construed to be direct in the particular context. If the causation is construed as indirect, then the periphrastic causative is to be preferred for independent, semantic reasons (see Boyd & Goldberg, 2011, Experiment 3 for discussion).

A second norming study reported in Robenalt and Goldberg (2015) was conducted to ensure that all novel sentences described situations that were semantically plausible. Native English speakers (n = 20) were recruited via Mechanical Turk to rate sentences for plausibility. Responses were made on a 5-point Likert scale. These ratings were averaged for each sentence and included as a covariate in statistical analyses to ensure that we were estimating changes in grammatical acceptability, independent of situational plausibility. We also included sentence length as a covariate because our sentences varied considerably in length (2–10 words).

Procedure

Participants rated sentences for grammatical acceptability using a 5-point Likert scale, in which a rating of 5 indicated that the sentence was completely acceptable and a rating of 1 indicated that the sentence was completely unacceptable. Before rating any of the experimental items, participants completed three practice trials that included one sentence that was unambiguously acceptable and one sentence that was unambiguously unacceptable. Participants who failed to rate these sentences appropriately were excluded from the analyses described below. Practice materials and instructions were adapted from Ambridge et al. (2008).

Results

Data were analyzed using R, a free software environment for statistical computing (R Development Core Team, 2008). Mixed effects models were run using the lme4 package (Bates, Maechler, Bolker, & Walker, 2013) and graphics were generated using the ggplot2 package (Wickham, 2009). But first, Figure 2 provides a visualization of the differences in acceptability ratings for native versus nonnative speakers for all sentence types including the baseline sentences, which involved familiar uses of verbs. Native ratings are shown in the left panel and nonnative ratings are shown in the right panel. The numerical means and standard deviations for each of the bars in Figure 2 are provided in Appendix S3 in Supporting Information online.

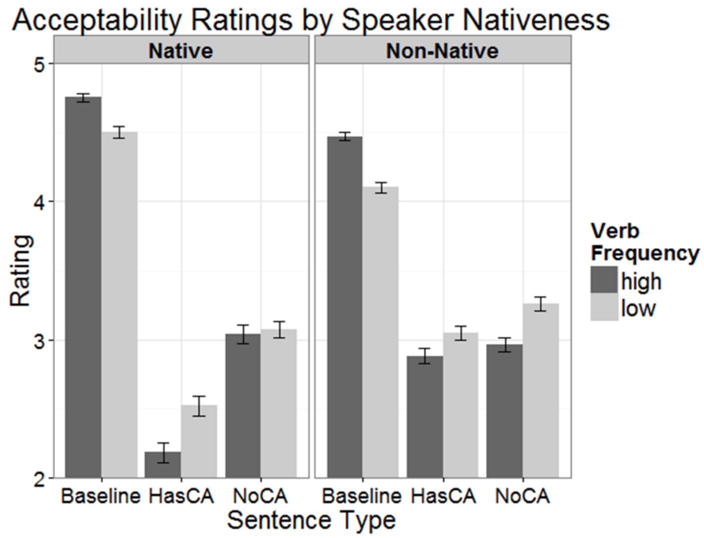

Figure 2 Sentence acceptability ratings for native versus nonnative speakers. The data are split by verb frequency (high vs. low) and sentence type (baseline, novel/hasCA, novel/noCA). Error bars represent standard error.

We made a few relevant observations with reference to Figure 2 before confirming these observations statistically using linear mixed models. Overall, nonnative speakers were less likely than native speakers to use the ends of the scale, which may be a result of nonnative speakers' lower confidence in their L2 ability. As predicted by usage-based frameworks, the frequency of positive exemplars clearly played a role for both native and nonnative speakers. In particular, both groups judged familiar (baseline) sentences to be more acceptable than either type of novel sentence; recall that baseline sentences involved verbs used in their most frequent construction (e.g., *Ashley disappeared*). Moreover, the baseline sentences with higher frequency verbs (e.g., *Ashley disappeared*) were more acceptable than comparable sentences with lower frequency verbs (e.g., *Ashley vanished*), presumably because the sentence types with higher frequency verbs are more familiar.

Turning our attention to judgments on novel sentences, in the case of native speakers (Figure 2, left panel), the sentences that have a competing alternative were treated differently than the sentences that did not. Specifically, we replicated the findings of Robenalt and Goldberg (2015), which are consistent with indirect negative evidence via statistical preemption: Native speakers find novel sentences that have no competing alternative to be somewhat more acceptable than novel sentences that do, and verb frequency only makes a difference for sentences that have a competing alternative. We suggest this is because native speakers have benefited from error-driven learning. Many potential but unwitnessed combinations of verbs and constructions have been statistically preempted by the attested forms with which they compete. The opportunities to learn from statistical preemption increase with the frequency of the competing form, which is indexed here by the frequency of the main verb. If there is no competing form for the intended message, verb frequency does not matter. In the case of nonnative speakers (Figure 2, right panel), the essential finding was that novel sentences with and without a competing alternative appear to be treated alike. We return to explore this in detail in a model that directly compares participants' performance on novel sentences in a later section.

Native Versus Nonnative Speakers: Initial Model

The observations described with respect to Figure 2 were confirmed with mixed linear models[2] that were used to estimate effects of speaker nativeness on perceived acceptability of novel and familiar (baseline) sentences. All independent variables were treatment coded; following Barr, Levy, Scheepers, and Tily (2013), we used the maximal random effects structure for which the model would converge. This resulted in the inclusion of random intercepts for participants and random slopes for the high versus low frequency version of each verb pair. The first model run included speaker group as a third independent variable. The full output of this model is provided in Appendix S4 in Supporting Information online. Due to a significant three-way interaction between speaker group, verb frequency, and noCA sentences, we split our data by speaker group and ran two separate models for native versus nonnative speakers. We report the estimates for each speaker group in Table 2, with the relevant factors highlighted. The reference level, which provides the intercept, is for baseline sentences with high frequency verbs. Fixed factors included high versus low verb frequency, sentence type (baseline/familiar vs. novel/hasCA vs. novel/noCA), and their interactions. The rightmost column of the table indicates where speaker nativeness was found to interact with other factors, as determined by the combined model reported in Appendix S4. All estimates in the following analyses should be interpreted in terms of the 5-point Likert scale used to rate sentence acceptability.

Table 2 Mixed linear model for acceptability rating

Random effects	Group	*SD*			
Subject (intercept)	Native	0.30			
	Nonnative	0.36			
Verb pair (intercept)	Native	0.40			
	Nonnative	0.37			
VF (slope)	Native	0.09			
	Nonnative	0.30			
Fixed effects	**Group**	**Estimate**	***SE***	***t***	**Native vs. nonnative**
(Intercept)	Native	6.34	0.25	25.69	$t = -5.41$, $p < 0.001$
	Nonnative	5.83	0.21	27.40	
High → Low VF	Native	–0.25	0.05	–5.21***	$t = -2.17$, $p = 0.03$
	Nonnative	–0.37	0.09	–4.30***	
Baseline → HasCA	Native	–2.60	0.06	–46.58***	$t = -14.09$, $p < 0.001$
	Nonnative	–1.63	0.05	–34.06***	
Baseline → NoCA	Native	–1.13	0.08	–14.35***	$t = 2.80$, $p = 0.01$
	Nonnative	–1.04	0.07	–15.63***	
Plausibility	Native	–0.12	0.04	–2.85**	—
	Nonnative	–0.10	0.03	–2.99**	

Sentence Length	Native	–0.19	0.01	–12.95***	—
	Nonnative	–0.16	0.01	–13.02***	
Low VF × HasCA	Native	0.59	0.08	7.85***	$t = –0.31$, $p = 0.75$
	Nonnative	0.44	0.07	6.53***	
Low VF × NoCA	Native	0.23	0.08	3.05**	$t = 4.24$, $p < 0.001$
	Nonnative	0.76	0.07	11.11***	

Note. *** $p < .001$, ** $p < .01$. VF = verb frequency; HasCA = has competing alternative; NoCA = does not have competing alternative. Because plausibility and sentence length were entered as covariates rather than independent factors, no direct comparison between native and nonnative speakers is available for these rows.

Familiarity was found to play an important role in perceptions of acceptability for both native and nonnative speakers. All speakers judged baseline sentences to be more acceptable than novel sentences, regardless of whether competing alternatives existed for the novel sentences. That is, both native and nonnative speakers showed a significant difference between baseline and hasCA sentences (native difference = –2.60, $t = –46.58$, $p < .001$; nonnative difference = –1.63, $t = –34.06$, $p < .001$) and also between baseline and noCA sentences (native difference = –1.13, $t = –14.35$, $p < .001$; nonnative difference = –1.04, $t = –15.63$, $p < .001$). In addition, both native and nonnative speakers displayed a preference for sentences containing higher frequency verbs in the baseline condition (native difference between high and low frequency = –0.25, $t = –5.21$, $p < .001$; nonnative difference = –0.37, $t = –4.30$, $p < 0.001$). Both of these effects are expected by usage-based models, insofar as greater familiarity results in greater acceptability.

Although both groups showed sensitivity to the frequency of the main verb, this interacted with sentence type in different ways. Native speakers' results replicated what had been found for native speakers in Robenalt and Goldberg (2015). Specifically, native speakers show a significant preference for low frequency verbs relative to high frequency verbs in novel/hasCA sentences by .59 points, $t = 7.85$, $p < .001$. At the same time, no preference was found for high versus low frequency verbs in novel/noCA sentences (the adjustment from High → Low Frequency is –.25 points, which is negated by the interaction between LF and NoCA that adds +.23 points back onto the predicted acceptability rating). These findings reinforce the evidence for statistical preemption in the case of native speakers, suggesting that native speakers do not simply tally all occurrences of verbs, but rather attend to verb frequency in context. We propose that context allows native speakers to generate implicit predictions about the types of sentences that are likely to occur in the upcoming discourse, and use prediction errors to make inferences about the conventional way to express a particular meaning. If a novel use is witnessed when a conventional alternative would be appropriate, a high frequency verb tends to sound less acceptable than a low frequency verb. When there is no such phrasal alternative, frequency is not predictive of acceptability.

In the case of nonnative speakers, both types of novel sentences were less acceptable than baseline (familiar) sentence types (estimates = –1.63 and –1.04 for hasCA and noCA, respectively; *t*s = –34.06 and 15.03, respectively; both *p*s < .001). Also, in the case of both types of novel sentence, the significant preference for high frequency verbs that is evident for the baseline sentences is eliminated. That is, significant interactions for low frequency verbs were found for both hasCA sentences and noCA sentences (estimates = +.44 and + .76, respectively; *t*s = 6.53 and 11.11, respectively; both *p*s < .001). In other words, nonnative speakers implicitly keep track of the sentence tokens they have heard, such that more familiar tokens become more acceptable. Deviations from familiar types of sentences are clearly less acceptable; however, nonnative speakers show no evidence of generating phrasal expectations in the way that native speakers do. They are less aware that sentences with a competing alternative are perceived to be less acceptable by native speakers; while native speakers downgrade hasCA sentences by –2.60 points relative to baseline sentences, nonnative speakers only penalize these sentences by –1.63. The difference between these two adjustments is significant according to the combined model, *t* = –14.09, *p* < .001 (see Table 2). The inability of nonnative speakers to distinguish when a novel sentence has a conventional competing alternative suggests that nonnative speakers are not generating online discourse expectations in the same way as native speakers.

Native Versus Nonnative Speakers: Follow-up Model

The inclusion of baseline sentence ratings in the model in Table 2 obscures whether judgments on novel sentences with low frequency verbs are lower in absolute terms than those with high frequency verbs, or whether they are only less different than the baseline sentences. In the latter case, it could be that higher frequency verbs simply get more of a boost by being in their familiar sentence frames than do lower frequency verbs. That is, when difference scores are used—as they are in Table 2 and in other work (e.g., Ambridge et al., 2008)—it is unclear whether low frequency verbs are actually more acceptable in novel sentences than high frequency verbs, or whether, instead, they are only less different from their baseline versions than are high frequency verbs (Ambridge et al., 2012; Ambridge & Brandt, 2013).

In order to investigate this question, we redid the analysis in Table 2 without the baseline sentences, and the resulting model is summarized in Table 3. Once again, a significant three-way interaction between speaker group, sentence type, and verb frequency prompted us to split the data according to speaker group and run two separate models for each group. The output of the combined model is provided in Appendix S5 in Supporting Information online, while the results shown in Table 3 show the separate estimates for native versus nonnative speakers. The rightmost column indicates where speaker group was found to interact with the factors of sentence type (hasCA vs. noCA) and verb frequency (high vs. low), as evidenced by the output from the combined model. The reference level for these models represents the estimated acceptability for hasCA sentences using a high frequency verb.

Table 3 Mixed linear models for novel sentences only

Random effects	Group	*SD*
Subject (intercept)	Native	0.59
	Nonnative	0.56
Verb pair (intercept)	Native	0.70
	Nonnative	0.62
VF (slope)	Native	0.26
	Nonnative	0.47

Fixed effects	Group	Estimate	*SE*	*t*	Native vs. nonnative
(Intercept)	Native	3.57	1.16	3.07	$t = -1.29, p = 0.20$
	Nonnative	4.30	1.11	3.86	
High → Low VF	Native	0.35	0.12	2.80**	$t = -1.74, p = 0.08$
	Nonnative	0.16	0.19	0.85	
HasCA → NoCA	Native	1.22	0.45	2.69**	$t = -10.43, p < 0.001$
	Nonnative	0.11	0.38	0.28	
Plausibility	Native	–0.05	0.15	–0.36	—
	Nonnative	–0.21	0.16	–1.28	
Sentence Length	Native	–0.19	0.12	–1.55	—
	Nonnative	–0.08	0.09	–0.86	
Low VF × NoCA	Native	–0.34	0.17	–2.02*	$t = 4.04, p < 0.001$
	Nonnative	0.08	0.26	0.29	

Note. ** $p < .01$, * $p < .05$. VF = verb frequency; HasCA = has competing alternative; NoCA = does not have competing alternative. Plausibility and sentence length are included as covariates rather than independent factors and thus no direct comparison is available for the native versus nonnative contrast.

Native speakers indeed display the pattern already discussed, in line with the predictions of statistical preemption, in a replication of Robenalt and Goldberg (2015). In particular, they show an overall preference for novel/noCA sentences over novel/hasCA sentences (+1.22 points, $t = 2.69$, $p = .005$), and an advantage for low over high frequency verbs only in the case of sentences that have a competing alternative (novel/hasCA). This can be seen in Table 3: The significant Low VF × NoCA interaction for native speakers (estimate = –.34 points, $t = 2.02$, $p = .04$) is negated by the adjustment for High → Low VF. That is, native speakers do not penalize novel uses of high frequency verbs relative to low frequency verbs when there is no readily available alternative to the novel use. Nonnative speakers, on the other hand, do not significantly distinguish novel/noCA from novel/hasCA, either in terms of any overall preference for the former (+.11 points, $t = .28$, $p = .78$) or in terms of any effect of verb frequency on novel sentences. Thus, nonnative speakers do not distinguish between novel sentences with and without a competing alternative in the ways that native speakers do. We suggest that this is because nonnative speakers are less able to anticipate upcoming utterances, so they have less opportunity to learn from incorrect predictions and therefore less ability to discriminate between sentences that have a ready competitor and those that do not.

Although there is a numerical preference for low frequency verbs over high frequency verbs, this difference does not approach significance (+.16, $t = .85$, $p = .40$). There is also no interaction between verb frequency and novel sentence type (estimated change = .08 points, $t = .29$, $p = .77$), indicating that verb frequency plays no more of a role in either the noCA or the hasCA condition. Thus for nonnative speakers, novel sentences are recognized to be novel (as indicated by their lower-than-baseline ratings), but they do not discriminate against sentences for which there exists a more conventional alternative formulation, nor against verbs that have been witnessed in other conventional constructions with high frequency. We can be confident that the lack of differences is not due to a lack of power, since there were 276 nonnative speakers in our sample. Moreover, we see the same pattern of results in subsets of the data that are divided by speakers' L1, particularly in Tamil (*n* = 63) and in Hindi (*n* = 33). The subgroup of Spanish native speakers (*n* = 55) look numerically more like native speakers, which may well be due to the fact that they had a higher self-reported speaking proficiency in English (85 out of 100) compared to the Tamil or Hindi speakers (72 and 74 out of 100, respectively).

We interpret the present findings regarding nonnative speakers as follows. L2 learners track what they have heard, and thus can clearly distinguish between familiar (baseline) sentences and unfamiliar sentences. Further, like native speakers, they prefer familiar sentences containing higher frequency verbs to those containing lower frequency verbs. Thus, L2 learners take good advantage of positive statistical evidence. At the same time, nonnative speakers make less use of comparisons between what is said and *what might have been said*. Indeed, if L2 learners have a reduced ability to generate expectations during online comprehension, then they should have less opportunity for the error-driven learning required for statistical preemption to work. Of course not all nonnative speakers are alike. In the following section, we investigate individual differences within the L2 learners.

Individual Differences in Proficiency

L2 learners are a heterogeneous group, as their native languages and the conditions under which they acquired their L2 vary extensively. Nonnative speakers may vary in age of first exposure to their L2; they may use their L2 in isolated and infrequent contexts, or they may be fully immersed in their L2; and they may use their L2 primarily in speaking or primarily in reading. We were interested in seeing whether self-reported measures of language experience or proficiency would be significant predictors of near-native judgments by L2 learners. We operationalized nativelike ratings by calculating the average rating for each of our experimental items from the native speaker group. We then subtracted each nonnative speaker's raw acceptability score from the average native rating to yield a difference score. If the speaker produced ratings that were similar to the average native rating, then these difference scores would be relatively small (close to 0). If the speaker produced ratings that were different from the average native rating, then these difference scores would be large. Negative difference scores indicate that the nonnative speaker underestimated acceptability relative to a native speaker, while positive scores indicate the reverse.

Five biographical predictors of difference scores were collected from the nonnative speakers: age of first exposure to English, total years of exposure to English, and self-rated estimates of English proficiency (speaking, reading, and writing). To collect the proficiency measures, each participant was presented with a sliding bar bounded between 0–100 and asked to rate their speaking, reading, and writing proficiency on a scale between 0 (*I can't do this at all*) and 100 (*I'm as good as a native speaker*). Model comparisons were used to determine whether any of the biographical variables could explain significantly more variance in the difference scores than the independent variables of sentence type (baseline, novel/hasCA, novel/noCA) and verb frequency (high, low). Speaking proficiency was by far the best predictor of nativelike ratings ($\Delta R^2 = 62.20$, $F = 18.67$, $p < .001$). The full model output for the linear regression including speaking proficiency is listed in Table 4.

Table 4 Results of the linear regression relating nonnative speaking proficiency to the difference from nativelike sentence ratings

Variable	Estimate	*SE*	*t*
(Intercept)	–1.10	0.16	–7.00
High → Low VF	–0.12	0.06	–1.85
Baseline → HasCA	3.08	0.22	13.89***
Baseline → NoCA	1.34	0.22	6.02***
Speaking proficiency	0.01	0.00	5.42***
Low VF × HasCA	–0.05	0.09	–0.58
Low VF × NoCA	0.40	0.09	4.47***
HasCA × Speaking proficiency	–0.03	0.00	–9.93***
NoCA × Speaking proficiency	–0.01	0.00	–5.37***

Note. *** $p < .001$. VF = verb frequency; HasCA = has competing alternative; NoCA = does not have competing alternative.

Once again, the reference level, represented by the intercept, is for baseline sentences with high frequency verbs. Therefore, the negative value for the intercept indicates that relative to native speakers, nonnative speakers tended to underestimate the acceptability of baseline items. The positive estimates for Baseline → HasCA and Baseline → NoCA indicate the reverse, namely that nonnative speakers overestimated the acceptability for both types of novel items. The positive value for speaking proficiency as a main effect indicates that the higher proficiency speakers underestimate the baseline scores to a lesser extent than lower proficiency speakers. In mathematical terms, the .01 estimate for speaking proficiency is multiplied by the proficiency rating for a given speaker, such that higher speaker proficiency results in the predicted difference score becomes closer to 0 (i.e., more nativelike). Similarly, negative estimates for the interactions between speaking proficiency and both novel sentence types demonstrate that higher proficiency speakers judged these sentences to be less acceptable. These estimates are once again multiplied by the proficiency rating for a given speaker. Since the estimate for NoCA × Speaking proficiency is equivalent to the estimate for Speaking proficiency alone, these values cancel each other out. The HasCA × Speaking proficiency estimate of .03 indicates that nonnative speakers found hasCA sentences to be progressively less acceptable as their speaking proficiency increased. Thus, overall lower speaking proficiency was associated with lower acceptability ratings for baseline items and greater acceptability ratings for novel sentences, especially for hasCA sentences. This relationship is illustrated in Figure 3, which displays the average difference from nativelike ratings at four tiers of speaking proficiency, as detailed below.

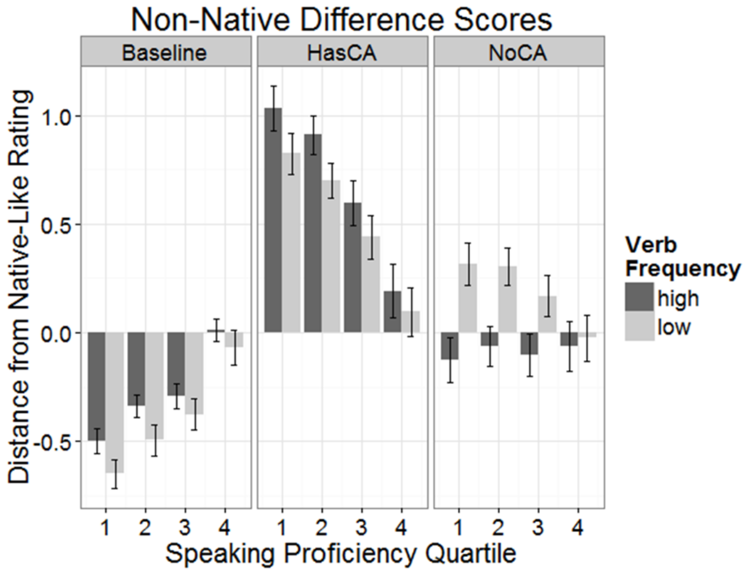

Figure 3 Average difference from nativelike ratings at four tiers of speaking proficiency**.** **N**onnative speakers were divided into four bins for speaking proficiency (low, mid-low, mid-high, high; numbered 1–4, respectively). Negative difference scores indicate that the nonnative speakers thought the sentences were less acceptable than native speakers did; positive difference scores indicate the reverse. Smaller absolute values are closer to nativelike ratings.

The highest proficiency tier was constructed by titrating the nonnative speakers based on speaking proficiency until they were statistically indistinguishable from native speakers. The cutoff point was determined by running a combined mixed model on native speakers and all nonnative speakers above a particular speaking proficiency rating until speaker nativeness ceased to be a predictive factor in the model. This cutoff point was 92 or greater, or the top 22% of all speakers. The other three tiers were constructed by dividing all remaining speakers as evenly as possible into three groups (n = 63, 73, 78, respectively). As Figure 3 demonstrates, L2 learners with low to mid-high proficiency judged baseline sentences to be less acceptable than native speakers while they judged novel sentences with a competitor to be more acceptable than native speakers. All nonnative speakers were relatively close to nativelike ratings for novel/no-CA sentences when a high frequency verb was used, but only high proficiency speakers correctly estimated the acceptability of these items when a low frequency verb was used. Interestingly, the area in which nonnative speakers showed the greatest deviation from nativelike scores was the hasCA sentences. This is also evident in the mixed model represented in Table 4, where the estimate for the interaction between Speaking proficiency and HasCA is three times what it is for noCA. Given that the .03 estimate is multiplied by the speaking proficiency rating for a speaker in order to estimate his or her distance from nativelike ratings, this suggests that an important component of speaking like a native speaker is the ability to recognize novel uses that already have a conventional alternative.

Discussion

Native speakers and nonnative speakers both show an advantage for baseline sentences compared to novel sentences, and for higher frequency verbs in baseline sentences relative to lower frequency verbs. These are expected effects of greater experience with baseline sentences, particularly those with higher frequency verbs. When it comes to novel sentences, the present results show precisely the pattern predicted by statistical preemption for native speakers: They are willing to use verbs in new ways to the extent that there is no readily available competing alternative phrasing for the meaning that they are trying to express. The results also replicate previous findings suggesting that when a novel sentence has no competing alternative, the frequency of the main verb makes no difference to native speakers (Robenalt & Goldberg, 2015). When there exists a competing formulation, the competing alternative with a high frequency verb is more readily available than the same sentence with a lower frequency verb. This results in an increase in acceptability for novel sentences for lower frequency verbs when the sentence has a competing alternative (Ambridge et al., 2008; Brooks & Tomasello, 1999; Theakston, 2004; Robenalt & Goldberg, 2015).

Nonnative speakers, on the other hand, show no evidence of taking alternative formulations into account when judging novel sentences. Unlike native speakers, they show no preference for novel sentences that do not have a readily available conventional alternative over novel sentences that do. They are also not significantly more likely to accept a novel sentence like *She vanished the rabbit* than *She disappeared the rabbit*, even though *disappear* has presumably been witnessed more frequently in a periphrastic causative that competes with the transitive use (e.g., *She made the rabbit disappear*). A likely interpretation of these findings is that nonnative speakers are less influenced by comparisons between how a verb is being used and how it might have been used.[3]

The present results are not entirely unexpected, given previous work indicating that L2 learners have a reduced ability to generate expectations on the basis of grammatical forms during comprehension (Grüter et al., 2012, 2014; Lew-Williams & Fernald, 2010). Our work extends this hypothesis to the anticipation and recognition of alternative grammatical forms. That is, nonnative speakers may be less able to make the necessary prediction of an upcoming grammatical choice that is needed for competition-driven learning to work. Statistical preemption requires that the learner initially expect a particular form to appear in order for that prediction to be violated, so that the error can provide a chance to learn. The inability to generate competing forms to the same extent as native speakers may well reduce the role of statistical preemption in the acquisition of L2 argument structure.

Consistent with the present findings are results from Trahey and White (1993) in a study of 11-year-old French speakers learning English as a L2. While French allows adverbs to occur directly after the verb, English does not (??*John kisses often Mary*). Intensive exposure to the English pattern of using adverbs before the verb led participants to increase their production of the correct order, but participants showed no evidence of avoiding the French-like order. That is, nonnative English speakers took advantage of the positive exemplars they witnessed and produced more utterances of that type, but they showed no evidence of learning that the unwitnessed form was unacceptable. Relatedly, Long, Inagaki, and Ortega (1998) found that explicit recasts were markedly more effective for L2 learners than statistical preemption, quite possibly because recasts do not require the same anticipatory processes. Instead of comparing what is said with what might have been said (statistical preemption), recasts only require the comparison of what was said by one person with what is said immediately after by another person. Also consistent are the findings of Ambridge and Brandt (2013), who collected judgments on novel sentences that had competing alternatives. They found that verb frequency correlated inversely with acceptability for native speakers of English, but not German learners of English.

Of course, nonnative speakers are not all alike. Posthoc analyses revealed that self-reported speaking proficiency was a strong predictor of more nativelike judgments. Those nonnative speakers who rated their speaking proficiency to be 92 or greater on a scale of 0–100 (the top 22% of L2 learners in our pool) showed judgments that were statistically indistinguishable from those of native speakers. Thus, L2 learners do appear to take alternatives into account at very high proficiency levels (see also Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Grüter et al., 2012; Hopp, 2013). That is, with sufficient proficiency, L2 learners can make use of alternative formulations and can achieve nativelike intuitions in this domain.

Statistical preemption requires that learners take advantage of competition dynamics: When one form is witnessed in a given context, other forms that might have been expected to occur become slightly inhibited. When this happens repeatedly, in the absence of positive evidence, the unwitnessed form is suppressed in such a way that native speakers come to recognize it as unacceptable. Native speakers essentially know that there is a "better" way to express the intended message. Nonnative speakers, on the other hand, appear less able to take competing alternatives into account. It is for this reason, we suggest, that they are more likely to produce the type of examples in (1)–(4). While upon reflection they are likely to realize that the forms are relatively novel, they have not learned to suppress those forms in favor of an alternative formulation. Again, at very high levels of L2 proficiency, as judged by self-reported speaking proficiency, nonnative speakers do attain nativelike competence in this area.

Statistical preemption is widely acknowledged to be at work in the learning of irregular morphology, by native and nonnative speakers alike. All speakers who learn to avoid producing *goed* and *foots* presumably learn by repeatedly witnessing *went* and *feet* in the contexts that otherwise would have been appropriate for the regular forms. But while *went* and the hypothetical *goed* are true synonyms, phrasal paraphrases are never fully synonymous, and verbs routinely appear in more than one phrasal construction. Thus, learning that, for example, *explain something to me* preempts *explain me something* requires learners to realize that the former has been witnessed in contexts in which the latter might have been expected, and this is clearly a more difficult task.

Limitations

The nonnative speakers in the present study came from a wide range of language backgrounds, which made it impossible to systematically investigate possible effects of transfer from speakers' L1. It could be that the existence or non-existence of cognate verbs or constructions in participants' L1 played a role in their acceptability judgments. It is also possible that some L2 participants were unfamiliar with some of the lower frequency verbs. However, if this were the case, we would expect them to find the novel sentences that contained those verbs to be judged less acceptable than sentences with higher frequency verbs, but we did not find this. In fact, the lower frequency verbs were numerically, but not significantly, more acceptable than the higher frequency verbs in novel sentences.

We relied on self-reports of speaking proficiency, which is clearly an imperfect measure; however, the fairly large sample size of 276 and the correlation of self-reported speaking proficiency with more nativelike intuitions was reassuring. While effects of sentence length, plausibility, and relative verb frequency were taken into account in the present study, one factor that was not systematically controlled for was the target construction involved. In fact, most of our noCA stimuli happened to involve the "caused-motion" construction (Goldberg, 1995), whereas the hasCA stimuli made use of a wider range of constructions, including the double-object construction, the simple transitive, the verb phrase complement constructions, and the caused-motion construction. Crucially, the between-groups differences between native and nonnative speakers cannot be attributed to a difference in the constructions used, since the stimuli were held constant across the two groups. A second study reported in Robenalt and Goldberg (2015) on native speakers controlled for target construction amongst all novel stimuli. That study focused on novel instances of the caused-motion construction that had a competing alternative (*The landscaper surrounded rocks around the garden*) and those that did not (*The sound rattled the bats out of their hiding place*). With semantic factors and plausibility taken into account, the study replicated the key results reported here for native speakers. That is, novel sentences with lower frequency verbs were judged more acceptable than those with higher frequency verbs, but only for sentences that had a competing alternative; verb frequency played no role in novel sentences without a competing alternative. It will be important to replicate that study with nonnative speakers in future work.

It will also be worth exploring whether native and nonnative speakers are equally aware of the general distribution of various constructions. For example, if a learner does not realize that the English double-object construction is favored for transfer events that are described with a pronominal recipient and a lexical theme argument, he or she would be less able to use statistical preemption because she or he would not necessarily expect a formulation like *explain me the story*. Finally, another potentially fruitful direction for future work will be to compare native English-speaking children to adults learning L2 English. Other work has found that native speakers as young as three use indirect negative evidence involving conservatism via entrenchment for the causative construction, while statistical preemption was not in evidence until age six or seven (Brooks & Tomasello, 1999). It remains to be seen whether nonnative speakers acquiring high levels of L2 ability show evidence of conservatism via entrenchment before statistical preemption.

Conclusion

Two of the longest standing puzzles in language acquisition are (a) how native speakers learn to avoid using certain formulations that are semantically sensible and syntactically licensed and (b) what factors distinguish L2 learners from native speakers. The present results support the idea that novel formulations are particularly avoided by native speakers when there exists a readily available alternative way to express the same message. We find evidence that L2 learners, on the other hand, are less aware that such novel formulations are actually disfavored. In fact, L2 learners of English show little evidence of taking competing formulations into account when judging novel sentences for acceptability. Recent findings that L2 speakers are not as likely to anticipate linguistic choices on the basis of grammatical forms converge with the present results, insofar as weak or absent predictions hinder L2 learners' ability to anticipate competition between possible upcoming forms. We argue that competing alternatives are acquired through a process of anticipating a particular form and repeatedly hearing a competitor in its place. Over time, the incorrect prediction is weakened and eventually replaced by the competitor. A reduced ability to predict upcoming forms thus has major ramifications for learning from competition dynamics between multiple candidate utterances in the way that native speakers use such prediction. This is not to say that nonnative speakers never achieve a nativelike appreciation of the competition between utterances. Self-rated speaking proficiency was found to correlate with more nativelike judgments, suggesting that learners may be able to cultivate their ability to generate expectations through extensive use of their L2. Indeed, the top 22% of our nonnative sample produced judgments that were statistically indistinguishable from those of native speakers. Future work will explore the connection between language use and online predictions in language processing, and the psychological realization of the reduced ability to generate expectations among L2 learners in particular.

Final revised version accepted 6 February 2015

Notes

1 Participants occasionally added or removed an adverb from the novel sentences (e.g., *The chief will laugh you back to your desk job* may have been changed to *The chief will laugh you right back to your desk job*). These changes were not counted as a paraphrase unless the argument structure of the critical verb was altered.

2 We make the assumption that the 5-point Likert scale was interpreted in a linear fashion (for a histogram of the acceptability responses by condition and group, see Figure A1 as part of Appendix S3 in Supporting Information online).

3 We considered the possibility that nonnative speakers use a less context-dependent form of indirect negative evidence—*conservatism via entrenchment* (Ambridge et al., 2012; Ambridge, Pine, Rowland, et al., 2012; Ambridge, Pine, Rowland, Freudenthal, et al., 2012; Braine & Brooks, 1995). Conservatism via entrenchment is the idea that as a verb becomes more entrenched in any pattern, it becomes less available for all novel uses. This does not require that other uses necessarily compete with the novel use. Conservatism via entrenchment, then, predicts that judgments on novel sentences should vary inversely with frequency, not only for sentences that have a competing alternative, but for those that do not as well. In this study, however, while lower frequency verbs in novel sentences were judged numerically more acceptable than higher frequency verbs, this difference did not approach significance.

Acknowledgments

We wish to thank Pavel Trofimovich for extremely helpful editorial advice, two anonymous reviewers and Lourdes Ortega for insightful comments on an earlier draft, and Bodo Winter for very useful statistical advice.

References

Ambridge, B. & Brandt, S. (2013). Lisa filled water into the cup: the roles of entrenchment, pre-emption and verb semantics in German speakers' acquisition of English locatives. *Zeitschrift für Anglistik und Amerikanistik*, *61*, 245–263. doi:10.1515/zaa-2013-0304

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument-structure overgeneralization errors. *Cognition*, *106*, 87–129. doi:10.1016/j.cognition.2006.12.015

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, *123*, 260–79. doi:10.1016/j.cognition.2012.01.002

Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (2012). The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language*, *88*, 45–81. doi:10.2307/41348883

Ambridge, B., Pine, J. M., Rowland, C. F., Freudenthal, D., & Chang, F. (2012). Avoiding dative overgeneralisation errors: Semantics, statistics or both? *Language and Cognitive Processes*, *iFirst*, 1–37. doi:10.1080/01690965.2012.738300

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, *76*, 28–55. doi:10.2307/417392

Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, *10*, 533–581.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package (Version 1.0-4) [statistical software package]. Retrieved from http://CRAN.R-project.org/package=lme4

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). New York, NY: Routledge.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational effects in second-language acquisition. *Journal of Memory and Language*, *44*, 235–249. doi:10.1006/jmla.2000.2750

Bley-Vroman, R., & Joo, H. R. (2001). The acquisition and interpretation of English locative constructions by native speakers of Korean. *Studies in Second Language Acquisition*, *23*, 207–219. doi:10.1017/S0272263101002042

Bley-Vroman, R., & Yoshinaga, N. (1992). Broad and narrow constraints on the English dative alternation: Some fundamental differences between native speakers and foreign language learners. *University of Hawai'i Working Papers in ESL*, *11*, 157–199.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology*, *63*, 545–545. doi:10.1037/0022-3514.63.4.545

Borovsky, A., Elman, J., Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with the timecourse of incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*, 417–436. doi:10.1016/j.jecp.2012.01.005

Bowerman, M. (1988). The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford: Blackwell.

Boyd, J. K. & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, *87*, 55–83.

Braine, M. D. (1971). On two types of models of the internalization of grammars. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 153–186). New York, NY: Academic Press.

Braine, M. D. S., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 353–376). Hillsdale, NJ: Lawrence Erlbaum.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
Brooks, P., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, *75*, 720–738. doi:10.2307/417731
Brooks, P. J., & Zizak, O. (2002). Does preemption help children learn verb transitivity? *Journal of Child Language*, *29*, 759–781. doi:10.1017/S0305000902005287
Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development*, *70*, 1325–1337. doi:10.1111/1467-8624.00097
Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J.R. Hayes (Ed.), *Cognition and the development of language* (pp. 11–54). New York, NY: Wiley.
Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). Amsterdam: John Benjamins.
Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455. doi:10.1080/01690969508407111
Bybee, J. L. (2010). *Language, usage and cognition* (Vol. 98). Cambridge, MA: Cambridge University Press.
Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum.
Collins, P. (1995). The indirect object construction in English: An informational approach. *Linguistics*, *33*, 35–49. doi:10.1515/ling.1995.33.1.35
Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*, 159–190. doi:10.1075/ijcl.14.2.02dav
DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. doi:10.1038/nn1504
DeLong, K. A, Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*, 226–239. doi:10.1016/j.bandl.2012.02.006
DiSciullo, A. M., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.
Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*, 353–387. doi:10.1017/S0272263112000915
Ellis, N. C. (2002). Frequency effects in language processing. *Studies in second language acquisition*, *24*, 143–188. doi:10.1017.S0272263102002024
Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics—Introduction to the special issue. *Applied Linguistics*, *27*, 558–589. doi:10.1093/applin/aml028
Erteschik-Shir, N. (1979). Discourse constraints on dative movement. In T. Givón (Ed.), *Syntax and semantics 12: Discourse and syntax* (pp. 441–467). New York, NY: Academic Press.
Eskildsen, S. W. (2009). Constructing another language: Usage-based linguistics in second language acquisition. *Applied Linguistics*, *30*, 335–357. doi:10.1093/applin/amn037

Flege, E. F., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, *41*, 78–104. doi:10.1006/jmla.1999.2638

Flett, S., Branigan, H. P. & Pickering, M. J. (2013). Are non-native structural preferences affected by native language preferences? *Bilingualism: Language and Cognition*, *16*, 751–760. doi:10.1017/S1366728912000594

Foraker, S., Regier, T., Khetarpal, N., Perfors, A., & Tenenbaum, J. B. (2007). Indirect evidence and the poverty of the stimulus: The case of anaphoric one. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the twenty-ninth annual Cognitive Science Society* (pp. 275–280). New York, NY: Lawrence Erlbaum.

Foucart, A., Martin, C. D., Moreno, E., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*, 1461–1469. doi:10.1037/a0036756.

Gennari, S. P., & Macdonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*, 161–187. doi:10.1016/j.jml.2007.07.004

Goldberg, A. E. (1993). Another look at some learnability paradoxes. In E. V. Clark (Ed.), *The proceedings of the twenty-fifth annual Child Language Research Forum* (pp. 60–75). Stanford, CA: Center for the Study of Language and Information.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: Chicago University Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*, 131–153. doi:10.1515/cogl.2011.006

Green, G. M. (1974). *Semantics and syntactic regularity*. Bloomington, IN: Indiana University Press.

Grüter, T., Hurtado, N., Marchman, V. A., & Fernald, A. (2014). Language exposure and online processing efficiency in bilingual development. In T. Grüter & J. Parasis (Eds.), *Input and experience in bilingual development* (pp. 15–36). Amsterdam: John Benjamins.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*, 191–215. doi:10.1177/0267658312437990

Havik, E., Roberts, L., van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*, 73–112. doi:10.1111/j.1467-9922.2009.00501.x

Herbst, T. (2014). The valency approach to argument structure constructions. In T. Herbst, H. J. Schmid, & S. Faulhaber (Eds.), *Constructions, collocations, patterns* (pp. 159–207). Berlin: Mouton de Gruyter.

Hirsh-Pasek, K., Treiman, R., & Schneiderman, M. (1984). Brown and Hanlon revisted: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, *2*, 81–88. doi:10.1017/S0305000900005596

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*, 33–56. doi:10.1177/0267658312461803

Hovav, M. R. & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of Linguistics*, *44*, 129–167. doi:10.1017/S0022226707004975

Hubbard, P. L., & Hix, D. (1988). Where vocabulary meets grammar: Verb subcategorization errors in ESL writers. *CATESOL Journal*, *1*, 89–100.

Inagaki, S. I. (1997). Japanese and Chinese learners' acquisition of the narrow-range rules for the dative alternation in English. *Language Learning*, *47*, 637–669. doi:10.1111/0023-8333.00024

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*, 257–282. doi:10.1075/lab.4.2.05kaa

Langacker, R. W. (1988). A usage-based model. *Topics in Cognitive Linguistics*, *50*, 127–163. doi:10.1075/cilt.50.06lan

Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *33*, 193–198. doi:10.1111/j.1467-9280.2007.01871.x

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*, 447–464. doi:10.1016/j.jml.2010.07.003

Linck, J. A., Kroll, J. F., Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*, 1507–1515. doi:10.1111/j.1467-9280.2009.02480.x

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, *82*, 357–371. doi:10.2307/329961

Marcotte, J. (2005). *Causative alternation errors in child language acquisition* (Doctoral dissertation). Stanford University, Stanford, CA.

Marcus, G.F. (1993). Negative evidence in language acquisition. *Cognition*, *46*, 53–85. doi:10.1016/0010-0277(93)90022-N

Martin, C. D., Thierry, G., Kuipers, J., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*, 574–588. doi:10.1016/j.jml.2013.08.001

Oehrle, R. T. (1976). *The grammatical status of the English dative alternation* (Doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Oh, E. (2010). Recovery from first-language transfer: The second language acquisition of English double objects by Korean speakers. *Second Language Research*, *26*, 407–439. doi:10.1177/0267658310365786

Pinker, S. (1981). Comments. In C. L. Baker & J. J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 53–63). Cambridge, MA: MIT Press.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

R Development Core Team (2008). R: A language and environment for statistical computing (ISBN 3-900051-07-0) [statistical computing software]. Vienna, Austria: R Foundation for Statistical Computing.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

Robenalt, C., & Goldberg, A. E. (2015). Judgment evidence for statistical preemption: It is relatively better to *vanish* than to *disappear* a rabbit, but a lifeguard can equally well *backstroke* or *swim* children to shore. *Cognitive Linguistics 26-4: 1-20.*

Rossi, E., Kroll, J. F., & Dussias, P. E. (2014). Clitic pronouns reveal the time course of processing gender and number in a second language. *Neuropsychologia*, *62*, 11-25.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*, 533–536. doi:10.1038/323533a0

Rutherford, W. E. (1989). Preemption and the learning of L2 grammars. *Studies in Second Language Acquisition*, *11*, 441–457. doi:10.1017/S0272263100008433

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development*, *19*, 15–34. doi:10.1016/j.cogdev.2003.08.001

Thompson, S. A. (1995). The iconicity of "dative shift" in English: Considerations from information flow in discourse. In M. E. Landsberg (Ed.), *Syntactic iconicity and linguistic freezes: The human dimension* (pp. 155–175). Berlin: Mouton de Gruyter.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Trahey, M., & White , L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition*, *15*, 181–204. doi:10.1017/S0272263100011955

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly*, *46*, 176–186. doi:10.1002/tesq.9

Tyler, A. (2008). Cognitive linguistics and second language instruction. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 456–488). New York, NY: Routledge.

Wasow, T. (2002). *Postverbal behavior*. Stanford, CA: Center for Study of Language and Information.

Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied linguistics*, *16*, 180–205. doi:10.1093/applin/16.2.180

Wickham, H. (2009). *Gplot2: elegant graphics for data analysis*. New York, NY: Springer.

Supporting Information

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

Appendix S1. Full List of Stimuli and Average Acceptability Ratings.
Appendix S2. Results from COCA Searches on Novel Uses for Novel Sentence Stimuli.
Appendix S3. Average Acceptability Ratings.
Appendix S4. Full Mixed Linear Model for Acceptability Rating.
Appendix S5. Mixed Linear Model for Novel Sentences.