LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Reduced Competition Effects and Noisier Representations in a Second Language

Karina Tachihara and Adele E. Goldberg
Princeton University

Native speakers strongly disprefer novel formulations when a conventional alternative expresses the same intended message, presumably because the more conventional form competes with the novel form. In five studies, second language (L2) speakers were less influenced by competing alternatives than native speakers. L2 speakers accepted novel interpretable sentences more readily than native speakers, and were somewhat less likely to offer competing alternatives as paraphrases or to prefer competing alternatives in forced-choice tasks. They were unaffected by exposure to competing alternatives immediately before judgments. Reduced sensitivity to competing alternatives was confirmed by L2 speakers' greater divergence from native speakers on judgments for novel formulations compared to familiar ones. Reduced sensitivity to competing alternatives also predicts noisier linguistic representations; consistent with this, L2 speakers performed worse on a verbatim recognition task, with performance correlating with more nativelike judgments. Proficiency was a modest predictor of judgments, but transfer effects were not.

Keywords second language; language acquisition; constructions; competition; memory; representations

We thank Editors Kara Morgan-Short and Pavel Trofimovich and as well as three anonymous referees for very helpful comments. We are also grateful to Libby Barak, Sammy Floyd, and Ting Qian for very useful discussions.

This article has been awarded a Preregistered Research Design badge. Preregistration for this study's research design and analyses is publicly accessible through AsPredicted.org. Direct links to the five preregistered experiments are available in the Supporting Information file. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to Karina Tachihara, Princeton University, Psychology, Peretsman-Scully Hall NJ, Princeton, NJ 08540, USA. E-mail: tachihara@princeton.edu.

DOI: 10.1111/lang.12375

Introduction

Adults who learn a new language typically fail to reach the same level of proficiency that first language (L1) speakers do (Hartshorne, Tenenbaum, & Pinker, 2018; although see Dąbrowska, 2018). Even learners who live in countries that predominantly use the second language (L2) sometimes produce errors that L1 speakers avoid. For example, Hubbard and Hix (1988) observed that intermediate and even advanced adult learners of English continue to produce verbs in "constructions they do not belong in" (p. 89). That is, except at the very highest levels of proficiency, L2 learners of English are prone to produce certain errors like those in Example 1 that were found online (Bley-Vroman & Joo, 2001; Bley-Vroman & Yoshinaga, 1992; Hubbard & Hix, 1988; Inagaki, 1997; Martinez-Garcia & Wulff, 2012; Oh, 2010).

Example 1
a. ? . . . the Secretary is absolutely forced that he has to give the information out . . .
b. ? Please explain me the meaning.
c. ? she considered to go back to her parents' house.

The errors in Example 1 involve acceptable syntactic patterns of English: The same patterns are fully acceptable with different main verbs (e.g., *convince*, *tell*, *want*, respectively). Moreover, none of the utterances in Example 1 is likely to result in a communication failure, because the intended interpretations are clear. Nonetheless, native English speakers strongly disprefer these types of sentences in favor of more conventional alternative ways to express the same intended messages. Specifically, L1 speakers prefer the conventional formulations in Example 2.

Example 2
a. . . . the Secretary is absolutely forced to give the information out . . .
b. Please explain something to me.
c. she considered going back to her parents' house.

Evidence from several studies has suggested that familiarity with a competing alternative that would express the same intended message influences L1 speakers' judgments about novel sentences. In particular, other things being equal, the more frequent, or entrenched, a conventional formulation is, the less acceptable L1 speakers judge a novel formulation intended to express the same message. For example, the expressions in Example 3 are both unconventional combinations of verb and construction, but L1 speakers tend to judge *drip* in 3a

to be somewhat more acceptable than the novel use of *spill* in 3b (Theakston, 2004).[1]

Example 3
a. ? I dripped the carpet with juice.
b. ? I spilled the carpet with juice.

Correspondingly, as illustrated in Example 4, *drip* is less frequent than *spill* in the more conventional formulation used to express the same intended messages as those in Example 3.

Example 4
a. I dripped juice on the carpet. (27 instances of DRIP <liquid> in Corpus of Contemporary American English; Davies, 2008)
b. I spilled juice on the carpet. (381 instances in SPILL <liquid> in Corpus of Contemporary American English; Davies, 2008)

Similarly, Ambridge, Pine, Rowland, and Young (2008) elicited acceptability judgments from L1 speakers on novel transitive sentences involving three pairs of verbs that are conventionally used in periphrastic causatives: *fall/tumble*; *disappear/vanish*; *laugh/giggle.* They reported that speakers judged novel sentences with lower frequency verbs to be more acceptable than the nearly synonymous sentences with high frequency verbs; sentences with novel verbs (which have a prior frequency of zero) were judged the most acceptable of all. Thus, the more entrenched a conventional alternative is, the less acceptable a novel formulation intended to express the same message is judged to be (for a production task with related results, see Brooks & Tomasello, 1999). In this way, L1 speakers generally judge novel formulations to be less acceptable to the extent that a better way to express the same intended message is well-entrenched (Goldberg, 2019).

Further evidence for this point comes from Robenalt and Goldberg's (2015) finding that L1 speakers judged novel sentences for which there is no well-entrenched conventional competing alternative as more acceptable than novel sentences for which a clear conventional competing alternative exists. In order to determine if a sentence has a competing alternative, Robenalt and Goldberg asked participants to paraphrase sentences. For example, for the novel expression in Example 5a, speakers tended to suggest a wide range of paraphrases (e.g., Examples 5b to 5d), without converging on a readily available formulation. They therefore considered Example 5a to lack a well-entrenched competing alternative.

Example 5
a. The teacher frowned a warning to the back of the class.
b. The teacher frowned at the kids at the back of the class.
c. The teacher warned the back of the class with a frown.
d. The teacher frowned as a warning to the back of the class.

Novel combinations of verbs and constructions (e.g., Example 3a) without a clear competing alternative were judged as more acceptable to L1 speakers than corresponding sentences for which a readily available alternative existed (i.e., speakers converged on the same paraphrase). Thus, L1 speakers tolerated novel combinations of verb and construction better when there was no obvious better way to express the intended message.[2]

The focus of the current work is on the extent to which competing alternatives affect the acceptability judgments of moderately proficient L2 speakers. Thus, we tested novel interpretable sentences, each of which has a competing alternative that L1 speakers strongly prefer. The fact that sentences exist that do not appear to violate any systemwide constraints but that do not involve nativelike selection is not new (Pawley & Syder, 1983). In this study, we concentrated on cases in which competition makes a difference for L1 speakers in order to determine whether competition is less effective for L2 speakers.

Robenalt and Goldberg (2016) followed up on their earlier work by comparing the judgments of a large group of L2 speakers from a variety of language backgrounds with a new group of L1 speakers on judgments of sentences with and without well-entrenched competing alternatives. The L1 pattern of judgments reported by Robenalt and Goldberg (2015) was replicated, and like L1 speakers, L2 speakers judged familiar or baseline combinations of constructions and verbs as markedly more acceptable than novel combinations. Critically, however, except at the highest proficiency levels, the L2 speakers did not appear to distinguish novel sentences that had a competing alternative from those that did not. In particular, L2 speakers judged novel combinations of verb and constructions that had a competing alternative to be just as acceptable as novel sentences that did not. Kang (2017) replicated this finding with a group of Korean learners of English. Likewise, in a study of acceptability judgments made by Chinese learners of English on denominal verbs (e.g., *She sweatered the child*), Zhang and Mai (2018) found that only at the highest proficiency levels did competing alternatives appear to be taken into account.[3]

We can also interpret certain older studies as consistent with the idea that L2 speakers are less likely to consider competing alternatives when judging novel formulations. For instance, English does not allow the verb and its direct object

to be separated by an adverb (??*John kisses often Mary*), although French does. Trahey and White (1993) reported that when French learners of English received intensive exposure to acceptable English order, they increased their production of that order, but they did not appear to learn that the unencountered order was unacceptable. Similarly, Kellerman (1979) had observed that, although advanced learners of English were better able to recognize familiar idioms in English than less advanced learners, they were no better at recognizing incorrect idioms as unconventional than less advanced learners.

The Present Study

We hypothesize that the fact that L2 speakers do not take competing alternatives into account may be related to their having a noisier or less specific linguistic representation compared to L1 speakers.[4] Noisier representations predict that memory for linguistic formulations are less strong, and distinctions between different linguistic formulations are not as clearly defined. Although previous studies have demonstrated that L2 speakers appear to take competing alternatives less fully into account than L1 speakers do, the nature of this difference and the possible mechanisms behind it have not been systematically investigated. In a series of experiments, we investigated the following key finding: L2 speakers judge novel (unconventional) sentences with competing alternatives more generously than L1 speakers do. After replicating the effect in Experiment 1, we tested whether it was due to a reduced awareness of the competing alternatives. In particular, we tested whether L2 speakers supply competing alternatives when asked for paraphrases in Experiment 2, whether they recognize competing alternatives to be more acceptable in Experiment 3, and whether exposure to the competing alternatives before the judgment task encourages L2 judgments to align more closely with those of L1 speakers in Experiments 4 and 5. If L2 speakers' linguistic representations are noisier, we expected L2 speakers to perform more poorly on tests of verbatim memory in L2, a possibility that we tested in Experiment 5. Because participants performed the same judgment task in four of the five experiments with remarkably stable results, we combined the judgment data and explored the effects of proficiency and transfer from L1 (focusing on Spanish-to-English transfer) in a sufficiently large sample of participants.

General Aspects of the Stimuli and Participants Across Experiments

In Table 1, the left-hand column lists the six target novel (unconventional) sentences that we used in our studies. These sentences included three instances each of two English constructions: the double-object construction and

Table 1 Target unconventional combinations of verb construction judged on 0–100 scale of acceptability in Experiments 1, 3–5 (left panel), and conventional competing alternatives (paraphrases) of each sentence to the left used for coding (Experiment 2) and used in nonjudgment tasks of Experiments 3 and 4 (right panel)

Target novel (unconventional) sentences	Conventional competing alternatives
Daniel forced that Helen compete.	Daniel forced Helen to compete.
Ken convinced that Laura clean her room.	Ken convinced Laura to clean her room.
Nick encouraged that Michelle finish school.	Nick encouraged Michelle to finish school.
Amber explained Zach the answer.	Amber explained the answer to Zach.
Lucas described Mike the apartment.	Lucas described the apartment to Mike.
Gary returned the museum his paintings.	Gary returned his paintings to the museum.

a clausal complement construction. We chose these constructions to provide different target constructions from those used in previous work (Kang, 2017; Robenalt & Goldberg, 2016; Zhang & Mai, 2018). All verbs had relatively high frequency but were unattested in combination with the argument structure used in the experimental stimuli in the 500+ million word Corpus of Contemporary American English (Davies, 2008). We kept the number of stimuli small so that we could keep the length of the entire experiment to under 20 minutes. This was necessary in order to test 70 unique L1 and L2 speakers in each condition in five experiments: a total of 980 participants (490 L1 speakers).

We asked participants to rate the acceptability of the unconventional sentences (Experiments 1, 3, 4, and 5) or to paraphrase them (Experiment 2). In the right-hand column of Table 1 are the conventional, competing alternatives that we used for coding the paraphrases (Experiment 2) and in the nonjudgment tasks (Experiments 3 and 4). We also included a set of 22 fillers in the judgment tasks. We expected 14 of the fillers to be relatively acceptable, and a subset of eight of these acceptable fillers (baseline sentences) were instances of the double-object and clausal complement construction with verbs that routinely occur in those constructions. We expected another eight fillers to be relatively unacceptable in that they included agreement errors, article omissions, and pronoun case errors (see Appendix S1 in the Supporting Information online).

Preregistration and Data Collection

We preregistered each of the five experiments on AsPredicted.org (for preregistration links, see Appendix S2 in the Supporting Information online), specifying our hypotheses, the dependent measures, the data collection process including restrictions on participants and sample sizes, and the main statistical test prior to data collection. We preregistered all of the analyses that are reported, unless specifically noted as exploratory, in order to increase the validity of our results.[5]

We collected the data using the crowdsourcing platform Amazon Mechanical Turk (https://www.mturk.com). Our selection criteria restricted participation to those who had to an IP address within the United States. We also used IP addresses and account information to prevent participants from engaging in any experiment more than once. We collected data until we had reached the desired number of usable participants for each condition in each experiment ($n = 70$), following our preregistered collection procedure and sample size.

We used the same preregistered criterion for classifying L1 and L2 participants for each of the five experiments. The task listed on Amazon Mechanical Turk included a description of the language requirement for recruiting L2 and L1 speakers separately. To qualify for inclusion, all participants initially answered questions about their language skills. In particular, they reported their speaking proficiency in English on a 0–100 scale (100 = "I'm as good as a native speaker"), their L1s, the age at which they had started learning English, and their current age. We only included L2 speakers who rated their proficiency in English to be 85 or lower; 85 was used as a criterion because, in the pilot test, three quarters of respondents reported having a proficiency of 85 or lower and because prior work had found that judgments by speakers at the highest levels in self-rated speaking proficiency align with those of L1 speakers (Robenalt & Goldberg, 2016). Across the five experiments, we collected data from a total of 490 L2 speakers. We report the number of participants excluded on the basis of their self-reported proficiency being between 86 and 99 in English for each experiment below. We checked whether the inclusion of participants with proficiency within this range would have changed the quantitative results in any of the experiments, and it did not. Combined results for all L2 participants with self-rated proficiency in English between 0 and 99 are provided in the Combined Judgment Data section. We also collected other measures of proficiency—age of acquisition and years speaking English (see Appendix S3 in the Supporting Information online), but they had no discernable effect.

We did not prescreen for specific L1s, and participants came from 73 different language backgrounds (Figure 1). Spanish was by far the most common language ($n = 150$), and we analyzed this subgroup separately. Other native

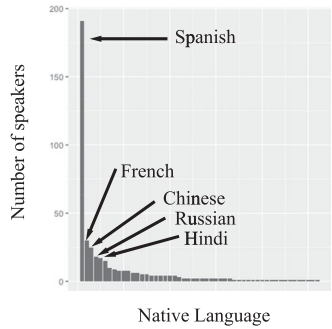

Figure 1 Distribution of the 72 native languages of participants by the number of speakers for each language.

languages included French ($n = 24$), Chinese ($n = 21$), Russian ($n = 15$), and Hindi ($n = 13$). All other languages had fewer than 10 speakers. The relative distribution was similar in each experiment (with Spanish accounting for a plurality and with a variety of other languages represented). We return to consider effects of specific languages in the Combined Judgment Data section. L1 English speakers met the selection criteria when they reported that English was their L1 and that their speaking proficiency in English was 100/100. We recruited most of the L1 speakers (77.5%) from a separate request for L1 English speakers.

In each experiment, L1 and L2 participants completed the same tasks, each requiring 3 to 20 minutes to complete. Participants received moderate financial compensation, appropriate for the estimated time for them to complete the tasks. Within each task in each experiment, we randomly ordered items for each participant. Coding ensured that participants could not review or change any responses that they had already completed.

Experiment 1: Judgment Task

In Experiment 1, we investigated whether L2 participants rated target unconventional sentences, for which a conventional, competing alternative existed, to be more acceptable than L1 participants did. We did this in order to confirm, with the current stimuli, the key finding that L2 speakers judge novel interpretable sentences to be more acceptable than L1 speakers do (Kang, 2017; Robenalt & Goldberg, 2016).

Method

Participants

Seventy adult native English speakers and 70 adult L2 learners of English were recruited on Mechanical Turk according to the preregistered criteria outlined above. During data collection, another 16 L2 participants whose self-reported proficiency fell between 86 and 99 took the survey, but they were excluded because they did not meet the preregistered selection criteria.

Materials and Procedure

The stimuli included the target unconventional sentences (see Table 1) and 22 filler sentences (see Appendix S1). We instructed participants to rate how acceptable or natural-sounding each of the 28 sentences was on a scale of 0 ("absolutely not acceptable and does not sound natural in English") to 100 ("completely acceptable and natural sounding"). To familiarize participants with the task, we provided two examples initially. Specifically, "I like the show" received a score of 95, and "To me pleases the show" received a score of 5. Sentences appeared one at a time on the screen with the 0–100 scale underneath. Participants clicked the "Next" button to move on to the next sentence.

Results

The preregistered multilevel model included group (L1 vs. L2) as the fixed effect and fit the maximal random terms that convergence would allow (Barr, Levy, Scheepers, & Tily, 2013; Kuznetsova, Brockhoff, & Christensen, 2017), using the lmerTest library (R Development Core Team, 2008; see Appendix S4 in the Supporting Information online for relevant R codes). In this case, we included random intercepts for participants and random intercepts and slopes for items. The model confirmed that L2 participants rated the target unconventional sentences as more acceptable than the L1 participants did, $\beta = 14.04$, $t = 3.70$, $p < .001$, as we had predicted. Figure 2 is a graphic representation of the mean judgment scores for the two groups. The results of Experiment 1 provided a benchmark for acceptability judgments on a new set of target unconventional sentences for L1 and L2 participants. In Experiments 2 and 3, we investigated whether the discrepancy in judgments between groups was due to the fact that the L2 participants were unaware of the preferred way of expressing the intended messages, that is, whether L2 participants were unaware of the competing alternatives (i.e., listed on the right-hand side of Table 1).

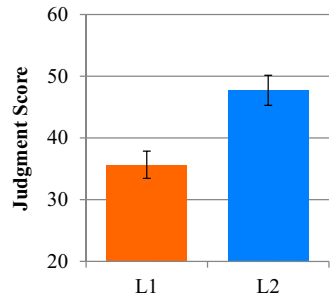

Figure 2 Experiment 1 judgment task: Mean judgment scores for unconventional sentences on a scale of 0–100 (*fully acceptable*) by group. Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

Experiment 2: Paraphrase Task

In order to assess whether L2 speakers were aware of the competing alternatives, we asked new groups of L1 and L2 participants to paraphrase each target unconventional sentence. Of interest was whether L2 participants would be as likely as L1 participants to provide the intended competing alternative for each target unconventional sentence.

Method

Participants

A new set of 140 participants was recruited for Experiment 2 (70 L1 speakers) following the same criteria as in Experiment 1. Another 16 participants whose self-reported proficiency fell between 86 and 99 took the survey during data collection.

Materials and Procedure

The same target unconventional sentences from Experiment 1 were used. We asked participants to generate paraphrases by rewriting each sentence in a different way while keeping the same meaning and using the same verb. Participants saw each target unconventional sentence on a separate screen and typed their responses into a sentence-length box.

Coding

Two independent coders who were blind as to whether participants were L1 or L2 speakers classified responses. We instructed the coders to ignore minor spelling, capitalization, article choice, and tense differences between

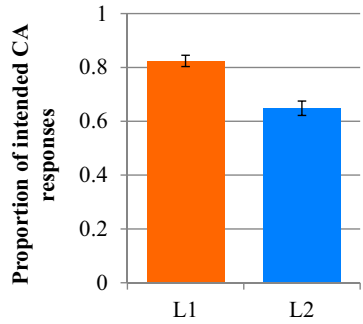

Figure 3 Experiment 2 paraphrase task: Mean proportion of participants' responses that provided the intended paraphrase (the competing alternative). Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

the paraphrases and the intended target competing alternatives. They coded passive forms of the target competing alternatives as correct target paraphrases. The intercoder reliability was 94%. A third coder adjudicated remaining disagreements.

Results

132 out of 840 responses were excluded from the analysis because participants failed to follow the instructions for those items. For example, the target unconventional sentence was copied exactly as it appeared, the response did not contain the target verb, or the response did not convey the original meaning. The intended competing alternatives were provided as paraphrases 81.59% of the time by L1 participants compared to 63.95% of the time by L2 participants. We again used the preregistered multilevel model with group (L1 vs. L2) as the fixed effect with maximal converging random structure. We included random intercepts for participants and random intercepts and slopes for items. This model revealed that L2 participants were less likely than L1 participants to supply the intended competing alternatives as paraphrases (see Figure 3), $\beta = -0.18$, $t = -2.73$, $p = .009$.

Experiment 3: Two Alternative Forced-Choice and Judgment Tasks

Although results of Experiment 2 indicated that L2 speakers are less likely to spontaneously suggest competing alternatives as paraphrases, L2 speakers may still be able to recognize the competing alternatives as preferable alternatives to the target unconventional sentences. Experiment 3 asked participants to choose

which of two sentences they preferred: the target unconventional sentence or its intended competing alternative. We repeated this two alternative forced-choice task for each target pair. After this task, we asked participants to rate the acceptability of the full set of six target unconventional sentences and 22 fillers as we had done in Experiment 1.

Method

Participants

A new set of 140 participants for Experiment 3 was recruited (70 L1 speakers), following the same criteria as for Experiments 1 and 2. Another 15 L2 participants whose self-reported proficiency fell between 86 and 99 took the survey during data collection, but we excluded them because they did not meet our preregistered selection criteria.

Materials and Procedure

In the forced-choice task, each participant saw six pairs of sentences, each including an unconventional sentence and its corresponding competing alternative. We told participants that one sentence would be more acceptable than the other and gave them these instructions, "Please choose the sentence that is more acceptable in English. If you are not sure, simply choose the sentence that sounds better, or more natural to you." A new pair of sentences appeared on each page. We randomized the order of presentation of each pair for each participant as well as the order of the conventional and the unconventional sentences on each trial. After the forced-choice task, all participants performed the same judgment task that was described in Experiment 1.

Results

Forced-Choice Task

When asked to choose between an unconventional combination of verb and construction and a familiar, competing alternative, the L1 participants nearly always chose the competing alternatives, as we had expected. We again used the preregistered multilevel model with group (L1 vs. L2) as the fixed effect with maximal converging random structure. We included random intercepts for participants and random intercepts and slopes for items. This model revealed that the L2 participants were less likely than the L1 participants to choose the competing alternatives as the more acceptable sentence, $\beta = 0.10$, $t = 3.27$, $p = .002$ (see Figure 4). At the same time, preregistered analyses revealed that L1 and L2 participants both demonstrated a clear preference for the competing

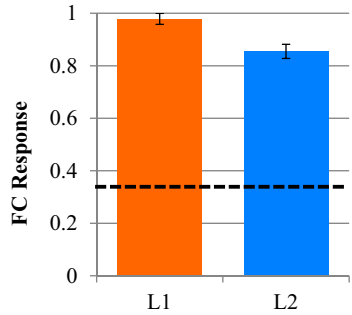

Figure 4 Experiment 3 forced-choice task: Mean proportion of two alternative forced-choice task responses in which the conventional competing alternative was chosen. The dotted line represents chance performance. Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

alternatives when we compared their performance to chance, as t tests confirmed, $M_{L1} = .95$, $t = 70.57$, $p < .001$; $M_{L2} = .86$, $t = 36.59$, $p < .001$.

Judgment Task

The judgment results replicated those of Experiments 1 and 2: The L2 participants rated the unconventional sentences as more acceptable than the L1 participants did. We again used the preregistered multilevel model with group (L1 vs. L2) as the fixed effect with maximal converging random structure. We included random intercepts for participants and random intercepts and slopes for items. This model revealed that the L2 participants rated the unconventional sentences to be more acceptable than the L1 participants did, $\beta = 16.62$, $t = 4.35$, $p = .001$ (see Figure 5a).

Next, we analyzed only judgments targeting the stimuli for which participants selected the competing alternative as preferable. This included 401 items for the L1 participants and 360 items for the L2 participants. Even for these items, the L2 participants remained significantly more generous in their ratings than the L1 participants did in this exploratory analysis, $\beta = 15.25$, $t = 3.04$, $p < .001$ (see Figure 5b). That is, even when the L2 participants correctly chose the competing alternative as more acceptable than the unconventional sentence, they continued to rate the unconventional sentence as more acceptable than the L1 participants did.

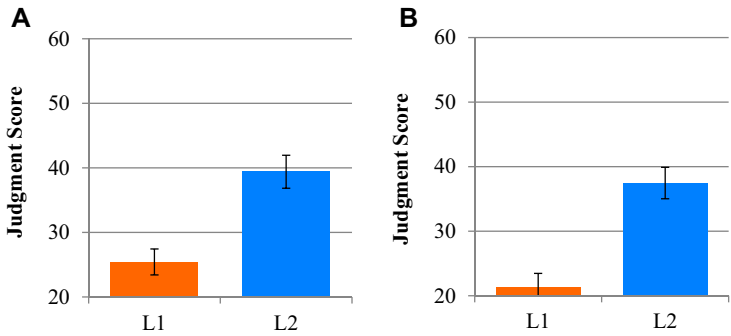

Figure 5 Experiment 3 judgment task: Mean judgment scores for L1 and L2 participants for all target items (A) and for only those items for which the participants preferred the competing alternative during the two alternative forced-choice task (B). Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

Experiment 4: Positive Exposure to Competing Alternatives Followed by Judgment Task

In Experiment 4, we determined whether exposure to the competing alternatives in naturalistic contexts (where they can be assumed to be acceptable) would serve to remind participants of, or familiarize them with, the competing alternatives. We hypothesized that this might result in L2 judgments aligning more closely with those of L1 speakers. Therefore, we randomly assigned participants to one of two conditions. In the competing alternative condition, they read competing alternatives, each of which was embedded at the end of a short story. A control condition similarly involved reading short stories, but the stories did not include any competing alternatives. After reading all the stories, both groups judged the same set of unconventional sentences (and fillers) used in Experiments 1 and 3. We predicted that the L2 participants in the competing alternative condition would judge target unconventional sentences to be less acceptable, thus aligning their judgments more closely with those of the L1 participants when compared to the control condition. We also compared L1 participants' judgments in the two conditions although we did not expect to see a difference because they had already rated unconventional sentences as quite unacceptable without any manipulation.

Table 2 Sample short stories and follow-up attention questions for the competing alternative and the control conditions

Competing alternative condition	Control condition
Daniel and Helen were in a competitive swimming team. Helen did not want to compete because she felt ill. Daniel forced Helen to compete.	Daniel and Helen were in a competitive swimming team. Helen did not want to compete because she felt ill. Daniel competed in her place.
Who was forced to compete?	Who competed?
1. Coach	1. Helen
2. Daniel	2. Daniel
3. Helen	3. No one

Method

Participants

A new set of 280 participants (70 × 2 groups × 2 conditions) were recruited for the experiment, using the same selection criteria from Experiments 1 to 3. Another 29 participants whose self-reported proficiency fell between 86 and 99 took the survey during data collection, but did not meet the criteria for inclusion.

Materials and Procedure

Participants read 12 stories, including six distractor filler stories that were identical in both conditions. All stories consisted of three sentences: The first two sentences provided a context for the third and final sentence. Only the last sentence differed depending on the condition. In the competing alternative condition, the last sentence in each target story was a competing alternative sentence. In the control condition, it was a sentence that did not share the main verb or meaning with a target (unconventional) sentence. Table 2 provides example stories.

We randomly assigned participants to the competing alternative condition or to the control condition. Participants read 12 stories, six stories that depended on the condition and six filler stories. Each story appeared on a separate page and was followed by an attention check question on the next page. These questions were simple multiple-choice questions and were all based on the meaning of the final sentence of the immediately preceding story (see Table 2 and Appendix S5 in the Supporting Information online). Participants clicked on the answer to the question and then continued to the next page. If they answered incorrectly, they returned to the story page having received the message "let's try again." They then answered the same multiple-choice question a second

time. At that point, regardless of their answer, the survey continued to the next story. We added the repeated attention check question to encourage participants to read and understand the story before continuing. We randomized the order of the stories for each participant. Immediately after the exposure task, all participants took the same judgment task used in Experiments 1 and 3.

Results

We considered responses to the attention check questions to be correct only when participants answered them correctly the first time that they encountered the question. Our preregistered analysis included only participants who were accurate on 11 out of 12 attention check questions, but this criterion turned out to be too strict because 31.15% of participants made at least two errors. Therefore, we used an accuracy threshold of 10 out of 12 questions, which allowed us to retain 81.64% of the participants. Results did not differ qualitatively when we used only the smaller sample based on the preregistered criterion of 11 out of 12 questions; for consistency, we used the same criteria for Experiment 5.

We again used the multilevel model with group (L1 vs. L2) as the fixed effect with maximal converging random structure. We included random intercepts for participants and random intercepts and slopes for items. We replicated the results from Experiments 1 and 3, as L2 participants once again rated the target unconventional sentences as more acceptable than the L1 participants did, $\beta = 16.99$, $t = 4.92$, $p < .001$ (see Figure 6a). To compare results in the two conditions, we used a preregistered model with proficiency, condition (competing alternative vs. control), and their interaction as fixed effects with maximal converging random structure. In this case, we included random intercepts for participants and random intercepts and slopes for items. Because we expected the manipulation to make a difference only for L2 participants, we looked at the difference by condition for each group separately. For L2 participants, the model revealed no difference between the competing alternative condition and the control condition, $\beta = 3.80$, $t = 0.52$, $p = .602$ (see Figure 6b). In other words, reading the competing alternative immediately before judging the corresponding unconventional sentence did not lead to the L2 participants' judgments on the sentences aligning more closely with those of the L1 participants.

There also was no significant effect of proficiency for the L2 participants, $\beta = 0.32$, $t = 1.50$, $p = .14$, although there was an interaction of condition and proficiency, $\beta = –0.54$, $t = –2.07$, $p = .042$, such that in the control condition only, the proficiency of the participants and their judgments were negatively correlated (i.e., the higher the proficiency, the lower the judgment score). This

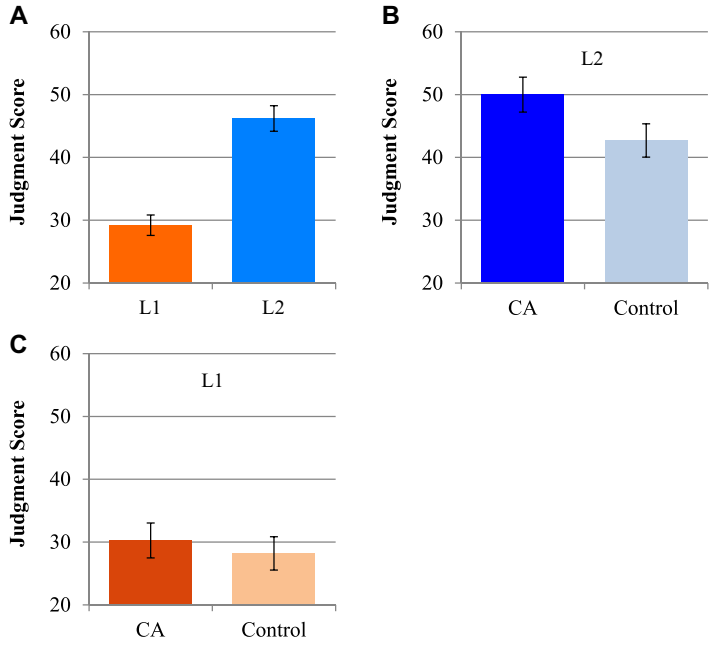

Figure 6 Experiment 4 exposure task: Mean judgment scores for L1 and L2 participants collapsed across conditions (A), by condition for L2 participants (B), and by condition for L1 participants (C). Error bars represent standard errors. CA = competing alternative. [Color figure can be viewed at wileyonlinelibrary.com]

correlation was in the expected direction although we did not anticipate that it would exist in only the control condition. We ran a similar model for the L1 participants, without proficiency as a variable (all L1 participants had a self-rated proficiency of 100). The maximal converging model included condition (competing alternative vs. control) as a fixed effect, random intercepts for participants, and random intercepts and slopes for items. As expected, we found no difference between conditions, $\beta = 2.06$, $t = 0.51$, $p = .62$ (see Figure 6c).

Contrary to our expectations, encountering the conventional competing alternatives immediately before the judgment task did not affect the L2 participants' judgments on the unconventional formulations. This raised the question of whether the L2 participants were able to remember the conventional competing alternatives after they had encountered them at the end of short stories. In order to determine this, in Experiment 5, we again exposed L2 participants to the short stories followed by a judgment task, and then we had the same participants perform a verbatim recognition memory task.

Experiment 5: Exposure, Judgments, and Verbatim Recognition Tasks

The lack of an effect of exposure to competing alternatives on sentence judgments in Experiment 4 may have been due to a failure by L2 participants to notice, encode, and/or remember the difference in constructions between the competing alternatives provided during the story exposure and the target unconventional sentences that we later asked them to judge. To test this possibility, we added a recognition memory test after the judgment task. We again included a competing alternative condition and a control condition to compare performance with and without exposure to the competing alternatives during exposure.

Method

Participants

We recruited a new set of 280 participants (70 × 2 groups × 2 conditions) for this experiment, using the same criteria as for Experiments 1 to 4. Another 24 participants whose self-reported proficiency fell between 86 and 99 took the survey during data collection, but were excluded based on our preregistered selection criteria.

Materials and Procedure

We again exposed each participant to 12 stories including six distractor filler stories that were identical across conditions. As in Experiment 4, all stories consisted of three sentences: Sentences 1 and 2 provided a context for the third and final sentence. We modified the six target stories used in Experiment 4 by using different arguments but keeping the construction and the verb combinations the same. For example, instead of "Daniel forced Helen to compete," participants read "Melissa forced Tom to dance." We also changed the two preceding sentences in each story in order to provide an appropriate context for the final sentence. We modified the stories for the control condition in an analogous manner, such that, as in Experiment 4, only the last sentence of each target story differed between the conditions: The competing alternative condition used competing alternative constructions in the last sentence, and the control condition used unrelated sentences that did not share verbs or meanings with the target unconventional sentences. The attention check questions were similar to those used in Experiment 4. The exposure task was the same as in Experiment 4, except for the six target stories which were edited as just described.

The same judgment task used in Experiments 1, 3, and 4 immediately followed the exposure. After the judgment task, we included a new memory

task. The memory task consisted of 12 items: six sentences that participants had read in the preceding stories (old items), drawn equally from the target stories and filler stories, and six new sentences. For example, "Melissa forced Tom to dance" appeared in the exposure story condition; subsequently, participants judged "Daniel forced that Helen compete"; and finally, in the memory task, participants had to decide if, for example, they had seen "Melissa forced that Tom dance" (in this case, the correct answer was "no"). Including different arguments in stories and the judgment task allowed us to test recognition memory for story sentences without interference from the intervening judgment task. We counterbalanced new and old items across participants. We instructed participants to "indicate if the sentence is an old sentence (i.e., you have seen this exact sentence before) or a new sentence (i.e., you have never seen this exact sentence before)." On each page, they saw a sentence and two choices, "This is a new sentence" and "This is an old sentence." They could continue to the new page only after they had made a choice between those two options.

Results

As in Experiment 4, we had preregistered a criterion of accurate responses of 11 out of 12 attention check questions, but this was again too strict as it would have required us to omit 59.64% of the participants from the analyses. Therefore, we used the threshold of 10 out of 12 accuracy, allowing us to use a consistent cutoff in both Experiments 4 and 5. This allowed us to include 78.21% of the participants.

Judgment Task

As in Experiment 1, 3, and 4, we found that L2 participants rated the target unconventional sentences as more acceptable than the L1 participants did, $\beta = 16.80$, $t = 5.26$, $p < .001$ (see Figure 7a), based on a multilevel model with group (L1 vs. L2) as a fixed effect with maximal converging random structure (random intercepts for participants and random intercepts and slopes for items). To determine whether the experimental manipulation had an effect on the L2 participants, we ran a preregistered model with proficiency, condition (competing alternative vs. control), and their interaction as fixed effects with maximal converging random structure (again, random intercepts for participants and random intercepts and slopes for items). As in Experiment 4, the model revealed no difference between the competing alternative condition and the control condition, $\beta = –4.18$, $t = –0.64$, $p = .52$ (see Figure 7b). There was also no significant effect of proficiency, $\beta = –0.03$, $t = –0.15$, $p = .89$, and no interaction, $\beta = –0.10$, $t = –0.37$, $p = .72$. For completeness, we also we used a

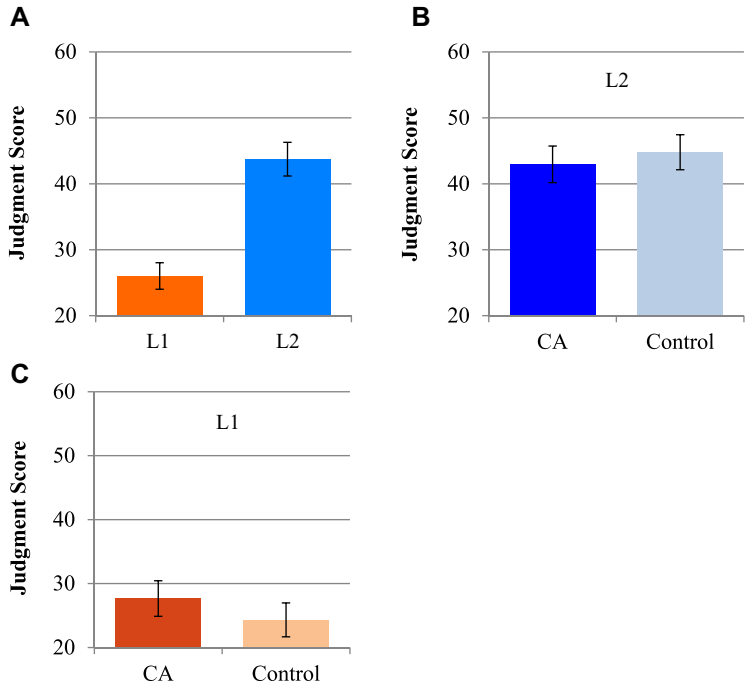

Figure 7 Experiment 5 judgment task: Mean judgment scores by L1 and L2 participants on a scale of 0–100 (*fully acceptable*) collapsed across conditions (A), for L2 participants by condition (B), and for L1 participants by condition (C). Error bars represent standard errors. CA = competing alternative. [Color figure can be viewed at wileyonlinelibrary.com]

maximal converging model that included condition (competing alternative vs. control) as a fixed effect for the L1 participants (again with random intercepts for participants and random intercepts and slopes for items). As expected, we found no difference between conditions, $\beta = -3.34$, $t = -1.01$, $p = .32$ (see Figure 7c), replicating Experiment 4.

Memory Task

Using signal detection theory, we calculated the d' score on the memory recognition results for each participant. A t test on d' scores (preregistered) revealed that both L1 and L2 participants performed above chance, $d'_{L1} = 0.59$, $t = 8.72$, $p < .001$; $d'_{L2} = 0.31$, $t = 4.58$, $p < .001$. However, compared to the L1

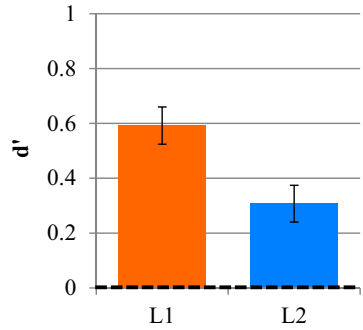

Figure 8 Experiment 5 memory task: Mean d' by group. The dotted line represents chance performance. Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

participants, the L2 participants were less accurate on the recognition memory task, $t = 2.98$, $p = .003$ (see Figure 8).

We further conducted an exploratory analysis of the data from the recognition memory task to better characterize the difference between L1 and L2 participants' performance. The L1 participants in the competing alternative condition were more accurate ($d' = 0.87$) than those in the control condition ($d' = 0.32$), $t = –4.37$, $p < .001$ (see Figure 9), which may have been due to the following aspect of the design. Our aim was to determine whether L2 speakers are able to remember the distinction between unconventional sentences and their competing alternatives. This led us to test the recognition of unconventional sentences in the competing alternative condition. These sentences were never encountered during the stories (all stories contained fully acceptable sentences). For instance, after encountering "Melissa forced Tom to dance" in a story, participants in the competing alternative condition judged "Daniel forced that Helen compete." Then during the memory task, we asked them whether they had seen the following (unconventional) sentence in a story: "Melissa forced that Tom dance." In total, three out of six of the new items in the competing alternative condition's memory task were unconventional. In the control condition, all stimuli in the memory task were conventional (and acceptable); the new sentences were simply slightly different from those encountered in the stories (e.g., "Kelly stared at the trees" vs. "Kelly looked at the trees"). The L1 participants in the competing alternative condition may have recognized the new sentences as new because all of the sentences in the stories had been acceptable, and some of the new sentences used in the memory task

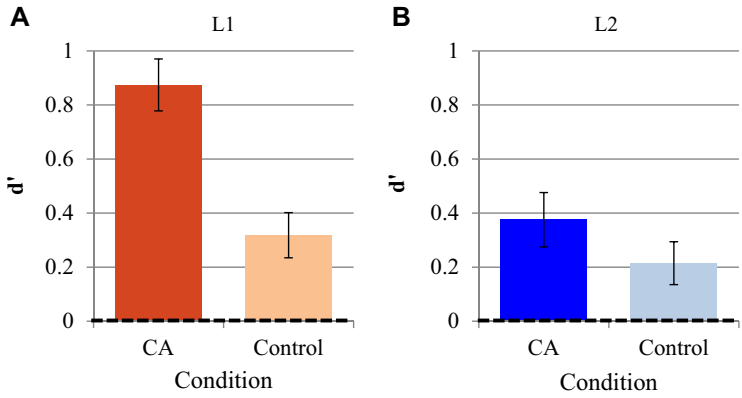

Figure 9 Experiment 5 memory task: Mean d' by group and condition. Dotted line represents chance performance. Error bars represent standard errors. CA = competing alternative. [Color figure can be viewed at wileyonlinelibrary.com]

were unacceptable. That is, the L1 participants may have realized, for example, that "Melissa forced that Tom dance" was new because it was unacceptable and that all of the sentences in the stories had been acceptable.

On the other hand, and most relevant in the current context, recognition accuracy for the L2 participants did not differ by condition: L2 participants in the competing alternative condition ($d' = 0.38$) were not significantly more accurate than those in the control condition ($d' = 0.21$), $t = 1.26$, $p = .57$ (see Figure 9). Insofar as the L2 participants did not use unacceptability as a cue to newness, it may suggest, once again, that they were not as sensitive to the unacceptability of unconventional sentences as the L1 participants were.

The above analysis of recognition memory used the d' measure from signal detection theory. This is the appropriate measure because it reports accuracy while taking into account any potential bias toward answering yes or no. But we were able to perform exploratory analyses on responses to old and new items separately by considering the data in Figure 10. Figure 10a displays responses to old items; it shows the mean number of hits (old items identified correctly as old) versus misses (old item erroneously identified as new). Figure 10b displays responses to new items; it shows the mean number of false alarms (new items erroneously identified as old) versus correct rejections (new items correctly identified as new). L1 participants and L2 participants tended to accurately recognize old sentences as old (hits) and rarely forgot or erroneously categorized old sentences as new (misses). That is, using a multilevel model for just hits

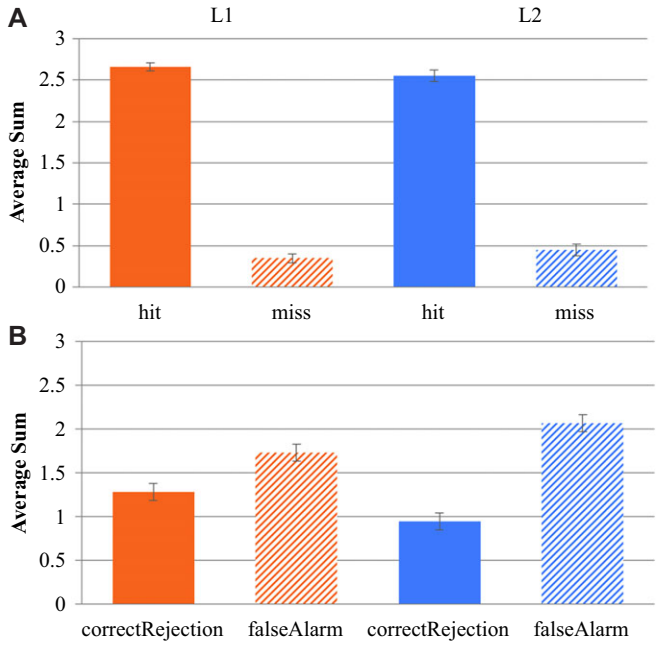

Figure 10 Experiment 5 memory task: Mean sum of each response type by group. Old items: hits or misses (A). New items: correct rejections or false alarms (B). Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

versus misses, group (L1 vs. L2), and their interaction as fixed effects with maximal converging random structure (random intercept per participant), there was no main effect of group, $\beta = 0.72$, $t = 1.68$, $p = .094$, and no interaction, $\beta = –0.21$, $t = –1.70$, $p = .09$. There was a main effect of hits versus misses, $\beta = –2{,}11$, $t = –22.82$, $p < .001$, which meant that there were more hits than misses for both L1 and L2 participants.

Figure 10b tells a different story. The L1 participants were less accurate with new items (correct rejections vs. false alarms), and the L2 participants were even worse. That is, the L2 participants were fairly likely to say that the new sentences were old. Using the same model, we compared correct rejections versus false alarms in the two groups and found a main effect of group, $\beta = 1.01$, $t = 3.24$, $p = .001$, and a main effect of response, $\beta = 1.13$, $t = 7.51$, $p < .001$. Importantly, there was a significant interaction of response and group, such that the L2 participants were more prone to false alarms than the L1 participants were, $\beta = –0.68$, $t = 3.41$, $p < .001$. To summarize, the

results demonstrated that the L2 participants were as good as the L1 participants at recognizing old items. But recognizing that new items were new proved more difficult, and the L2 participants were particularly likely to erroneously believe new sentences were old (false alarms compared to correct rejections). Thus, the L2 participants were less adept at discriminating new sentences from ones that they had encountered; that is, they demonstrated a bias toward believing that they had seen items even when they had not.

This is consistent with the hypothesis that L2 speakers' representations of encountered sentences may be less specific—noisier—than those of L1 speakers (see also Futrell & Gibson, 2017). That is, if L2 speakers' linguistic representations of encountered exemplar types exist but are noisier, they should find it challenging to recognize that new items are distinct from similar representations. This predicts that L2 speakers with more accurate recognition memory may have richer and more nuanced memories of which constructions they have encountered with which verbs. We therefore hypothesized that as verbatim memory (d') increases, judgments should align more closely with L1 speakers' for unconventional sentences. A Pearson correlation test confirmed, in fact, a significant correlation between participants' d' scores and their average judgments given to target unconventional sentences, $r = –.27$, $p = .008$. That is, better verbatim memory predicted judgments that were more aligned with those of the L1 participants.

Perhaps verbatim memory is simply a proxy for greater proficiency in English. To test this possibility, we ran a logistic regression with d' and self-reported proficiency as interacting predictors. While d' showed a significant relationship with judgments, $\beta = –10.52$, $t = –2.28$, $p = .03$, proficiency was not significantly correlated, $\beta = –0.02$, $t = –0.19$, $p = .85$. There was also no interaction between proficiency and d' scores, $\beta = –0.14$, $t = –0.75$, $p = .45$. This suggested that the L2 participants' memory was a better predictor of their judgments than self-rated proficiency.

Nonetheless, it would be quite strange if proficiency did not correlate with the degree to which L2 participants' judgments correlated with those of the L1 participants, particularly given that self-rated proficiency was previously found to be a significant predictor in a similar judgment task (Robenalt & Goldberg, 2016). In fact, we had purposely narrowed the range of proficiency with the exclusion criteria and we expect self-rated measures to be noisy; therefore, the lack of effect of proficiency on judgments in Experiment 5 may have been due to a power issue. Because none of the manipulations in Experiments 4 and 5 affected participants judgments (recall that the verbatim recognition task

followed the judgment task), we combined the data from all four experiments that included the judgment task in order to increase power.

Combined Judgment Data

Experiments 1, 3, 4, and 5, performed by unique groups of participants, all included the same judgment task. Despite the experimental manipulations preceding the judgment task in Experiments 4 and 5, all four experiments revealed that the L2 participants were significantly more generous in their judgments of the target unconventional sentences than the L1 participants were. In this section, we explore the full dataset of judgments in more detail, attempting to answer several questions. The first question is whether self-rated proficiency predicts L2 participants' judgments in the larger dataset. A second question is whether the L2 participants simply regressed toward the mean on judgments of unacceptable and acceptable sentences. If so, this could mean that L2 speakers are less secure overall in their judgments than L1 speakers are, a plausible possibility. To address this question, we compared L1 and L2 participants' judgments on target unconventional sentences with their judgments on a comparable subset of filler sentences—the set of eight baseline conventional sentences. Finally, the combined data allow us to consider possible transfer effects of the L2 participants' L1s; in particular, we consider the performance of the subpopulation of Spanish-speaking learners of English in some detail because this subgroup of 150 participants was of comparable size to those used in each of the individual experiments. Since we had preregistered each of the five experiments separately, the analyses on the combined data must be considered exploratory.

Effect of Proficiency on Judgment Scores

With the power of the combined, full dataset, we predicted that participants who reported higher proficiency in English would judge target unconventional sentences more like L1 speakers did (i.e., as less acceptable than less proficient speakers). We used a multilevel model with proficiency as a fixed effect and with a maximal converging random structure (random intercepts for participants, items, and experiments). When both L1 and L2 participants were included, the model demonstrated a significant negative relationship between proficiency and judgment scores for target unconventional sentences, $\beta = -0.30$, $t = -8.51$, $p < .001$. However, this relationship is not surprising because the L1 participants had proficiency scores of 100 and had low judgment scores for target unconventional sentences. More relevantly, we ran the same model using only the L2 participants who had met our preregistered criterion of having

self-rated proficiency scores of 85 or lower. Surprisingly, we did *not* find a significant effect of proficiency on judgments in this group of 334 L2 participants, $\beta = –0.07$, $t = –1.14$, $p = .25$. Self-rated proficiency is a noisy measure, but we had expected that it would nonetheless show some correlation with the judgment data. In order to increase the variation in proficiency as much as possible, we performed an additional analysis in which we included all L2 participants with a proficiency level of 99 or lower. With this full set of L2 participants ($n = 445$), we did find the expected relationship between proficiency and judgments, $\beta = –0.12$, $t = –2.48$, $p = .014$. Thus, although the L1 participants and the L2 participants differed in their judgments of unconventional sentences, L2 participants' judgments aligned more closely with L1 participants' judgments as their proficiency increased.

We used the same dataset of participants with a proficiency level of 99 or lower and the same model to test if any other variables, such as age of acquisition, current age, or number of years of use predicted the L2 participants' judgments. None was a significant predictor of judgment scores: age of acquisition, $\beta = –0.09$, $t = –0.86$, $p = .39$; current age, $\beta = –0.01$, $t = –0.18$, $p = .86$; number of years of English use, $\beta = –0.03$, $t = –0.43$, $p = .67$.

Comparing Performance for Novel Versus Baseline Sentences in L1 and L2

The judgments included 22 filler sentences, eight of which we predicted to be quite unacceptable and 14 of which we predicted to be acceptable (see Appendix S1). Of the acceptable sentences, eight sentences involved the same constructions that we had used in the target unconventional sentences (double-object and clausal complement constructions), with verbs that readily appeared in those constructions (e.g., "The realtor showed the happy couple a bungalow"). This subset of fillers involving conventional combinations of verbs and argument structure constructions provided a suitable baseline because it controlled for length, complexity, and type of construction in sentences that we expected to be acceptable for a comparison with the target unconventional sentences that we expected to be unacceptable. That is, comparison against these baseline sentences was a useful way of accounting for potential differences in how L1 and L2 participants used the scale, perhaps due to different degrees of confidence, certainty (R. Ellis, 1991), or metalinguistic awareness (R. Ellis, 2004).

In an analysis of the combined data, reassuringly, L2 participants rated the baseline (conventional) sentences as more acceptable than the target (unconventional) sentences, using a multilevel model with sentence type (unconventional vs. conventional) as a fixed effect with maximal converging

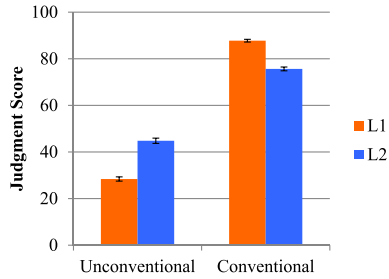

Figure 11 Combined judgment data: Mean judgment scores for unconventional and conventional sentences by group. Results show a significant sentence type × group interaction, $\beta = -1.96$, $t = -2.30$, $p = .03$. L2 judgments for novel sentences differed more (in absolute value) from L1 judgments than judgments for baseline sentences. Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

random structure and random intercepts for participants and items and random intercept and slope for experiment, $\beta = 32.05$, $t = 6.15$, $p < .001$. Next, we compared the difference in judgment scores between L1 and L2 participants for unconventional and conventional sentences. We found that L2 participants not only rated unconventional sentences as more acceptable, they also rated baseline conventional sentences as less acceptable. This was confirmed by a multilevel model with group (L1 vs. L2) as a fixed effect with maximal converging random structure (random intercepts for participants and random intercepts and slopes for items and experiments), $\beta = -12.34$, $t = 7.90$, $p < .001$ (see Figure 11). In order to compare the size of the differences, we used the scaled scores created by RStudio over the entire dataset. Because acceptable sentences had positive values and unacceptable sentences had negative values, we used the absolute values of the scaled scores to determine distance from the mean. A multilevel model with sentence type (unconventional vs. conventional) and group (L1 vs. L2) as interacting fixed effects with maximal converging random structure was fit with the maximal random terms that convergence would allow. In this case, we included random intercepts for participants and random intercepts and slopes for items and experiments.

Results showed a significant interaction of sentence type and group, $\beta = -1.96$, $t = -2.30$, $p = .03$. In other words, the difference between L1 and L2 participants for target (unconventional) sentences was larger than the difference between L1 and L2 participants for baseline (conventional) sentences. To summarize, uncertainty may play a role in the discrepancy between L1 and L2 judgments; in fact, our data confirmed that L2 speakers do tend to

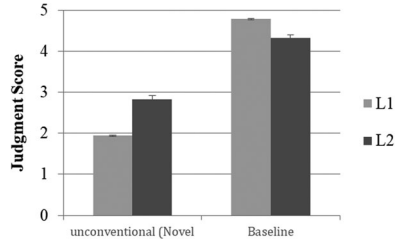

Figure 12 Reanalyzed data from Robenalt and Goldberg (2016) showed an interaction between groups and the difference between baseline (acceptable) sentences and novel (unacceptable) sentences, $\beta = 1.33$, $t = 11.41$, $p < .001$. L2 judgments for novel sentences differed more from L1 judgments (in absolute values) than judgments for baseline sentences. Error bars represent standard errors.

stay closer to the middle of the scale. However, the discrepancy between L1 and L2 judgments was especially large for unconventional sentences, which was the phenomenon that we had set out to investigate.

In order to confirm that novel interpretable sentences are of particular interest and that L2 speakers do not simply provide more middle of the road judgments overall, we performed a new, parallel analysis using the data reported by Robenalt and Goldberg (2016) that we present graphically in Figure 12. As in the analysis of the current data, we took the absolute values of the scaled score. We used a multilevel model with sentence type (target unconventional vs. baseline conventional) and group (L1 vs. L2) as interacting fixed effects with maximal converging random structure and fit the maximal random terms that convergence would allow. In this case, we included random intercepts for participants and random intercepts and slopes for items. As in the current data, the data collected by Robenalt and Goldberg revealed a significant interaction of sentence type and group, $\beta = –0.46$, $t = –12.89$, $p < .001$. That is, once again, the L2 participants were particularly different from the L1 participants on judgments of unconventional sentences.

Effect of Native Language and Transfer

One variable that may play a role in judgments involves transfer effects from L2 participants' already well-entrenched L1. That is, adults are highly practiced in the linguistic skills needed for their L1, and these skills constitute ingrained linguistic habits that can influence the learning and use of a newer L2 (Ambridge & Brandt, 2013; Austin, Pongpairoj, & Trenkic, 2015; Bates & MacWhinney, 1987; Bley-Vroman & Joo, 2001; N. Ellis, 2002; Finn & Hudson Kam,

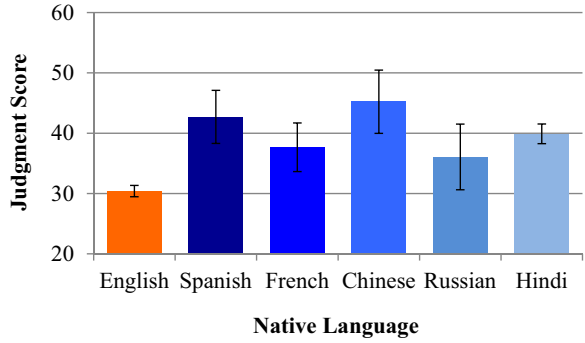

Figure 13 Mean judgment scores on the target unconventional sentences for each of the six most common participants' L1s in the combined data from Experiments 1, 3, 4, and 5. Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

2015; Foucart & Frenck-Mestre, 2011; Hernandez, Li, & MacWhinney, 2005; Kellerman, 1995; MacWhinney, 2006; Rutherford, 1989; Sabourin, Stowe, & de Haan, 2006; Selinker & Lakshmanan, 1992). We expected that participating language learners would face a particular challenge when a distinction required in L2 was irrelevant in their L1, especially if the distinction was an obscure or arbitrary one. Thus, we were interested in seeing whether L2 participants' specific L1 affected their judgments.

The combined dataset allowed us to consider L2 participants' L1 as a variable to investigate whether the key findings held when we treated L1s as a random effect. We again used the multilevel model with group (L1 vs. L2) as the fixed effect with maximal converging random structure. We included random intercepts for languages and participants and random intercepts and slopes for items. As in Experiments 1, 3, 4, and 5, we found that L2 participants rated the target unconventional sentences as more acceptable than the L1 speakers did, $\beta = 18.03, t = 6.51, p < .001$.

We also compared judgment scores across groups of L2 participants who shared a L1. We used a multilevel mixed model to compare Spanish, the most dominant language, to the four next most frequent L1s and fit the maximal random terms that convergence would allow. In this case, language was a fixed effect with random intercepts for participants and experiment and with random intercepts and slopes for items. There were no significant differences between Spanish and any of the four other most dominant languages (see Figure 13):

French, $\beta = –4.67$, $t = –0.65$, $p = .52$; Chinese, $\beta = 0.76$, $t = 0.07$, $p = .94$; Russian, $\beta = 2.68$, $t = 0.31$, $p = .76$; Hindi, $\beta = –6.03$, $t = –0.72$, $p = .49$.

To examine possible effects of transfer, we analyzed data from speakers of the one language for which we had a large enough number of L1 participants, Spanish ($n = 150$). Spanish is a particularly interesting test case because of the particular stimuli used in our studies. Both Spanish and English allow the same general range of verb meanings to appear with clausal complements while the two languages differ in how verb translations are used in their dative constructions. If transfer from Spanish was responsible for the fact that the L2 participants were more generous on judgments of unconventional sentences, then they should have behaved more like L1 English participants on the clausal complement target stimuli than on the double-object stimuli.

Specifically, English and Spanish both use clausal complements for verbs meaning *think*, *believe*, and *want* (e.g., Example 6a, Example 7a) and neither language allows clausal complements for verbs that mean *force*, *encourage*, or *convince* (e.g., Example 6b, Example 7b). For the latter verbs, both Spanish and English prefer a direct object plus infinitival complement (e.g., Example 6c, Example 7c).

Example 6
a. Melissa thought that Helen played.
b. ? Melissa forced that Helen play.
c. Melissa forced Helen to play.

Example 7
a. Melissa pensó que Helen había jugado.
b. ? Melissa obligó que Helen juegue.
c. Melissa le obligó a Helen a jugar.

The other type of unconventional sentences used in our stimuli were instances of the English double-object construction with Latinate sounding verbs. In particular, although English verbs that sound Germanic and mean *transfer* typically can appear in the double-object construction (e.g., Example 8a), each of the Latinate verbs strongly prefers a different, *to*-dative construction (e.g., Examples 8b and 8c). In Spanish, however, there are no constructions that distinguish Latinate-sounding verbs from Germanic-sounding verbs. With the exception of rare borrowings, all Spanish verbs are of Latin origin, and all verbs that express literal or metaphorical transfer are expressed in the same construction, which includes the animate dative marker *a* on the recipient argument and allows a different word order (Examples 9a and 9b).

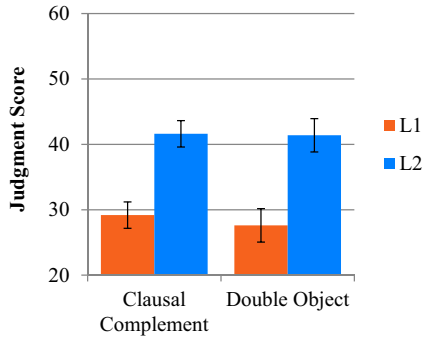

Figure 14 Mean judgment scores for native English participants (L1) and Spanish-speaking learners of L2 English by construction. Error bars represent standard errors. [Color figure can be viewed at wileyonlinelibrary.com]

Example 8
a. Amber told Zach the answer.
b. ? Amber explained Zach the answer.
c. Amber explained the answer to Zach.

Example 9
a. Amber le dijo a Zach la respuesta.
b. Amber le explicó a Zach la respuesta.

If the participating Spanish-speaking learners of English were simply transferring their knowledge of verb and construction from their L1 (Spanish) to their L2 (English), they should have recognized that the clausal complement constructions with *force*, *convince*, and *encourage* were unacceptable because they are unacceptable in Spanish as well. We might have expected, then, to see the Spanish-speaking participants have more trouble with the double-object construction because Spanish treats the verbal translational equivalents of *explain* and *tell* (*explicar* and *decir*) alike. Specifically, Spanish speakers might have difficulty recognizing that verbs like *explain* do not allow the double-object construction because verbs like *tell* do.

However, L1 Spanish participants had just as much difficulty with the clausal complement constructions as with the double-object construction, with their judgments on the two types of unconventional combinations patterning the same way (see Figure 14). That is, we analyzed the double-object construction and the clausal complement constructions separately, and, to a comparable degree in both, the L1 Spanish participants rated unconventional sentences

as more acceptable than the L1 English participants. In particular, we used a multilevel model with group (L1 English speakers vs. L2 speakers) as the fixed effect and maximal converging random structure (random intercepts for participants and random intercepts and slopes for items): double-object, $\beta = -13.78$, $t = -4.32$, $p = .004$; clausal complement, $\beta = 12.43$, $t = -5.15$, $p < .001$. Although transfer effects were certainly likely to partially account for L2 judgment scores in certain cases, data from L1 Spanish participants suggested that it was not the only reason L2 participants rated unconventional sentences as more acceptable than L1 participants did.

Discussion

Summary of Findings

We conducted five experiments with L1 English participants and with L2 participants from a wide variety of L1 backgrounds. The L2 participants all lived in the United States and were able to follow instructions in English, indicating that they had at least moderate proficiency in their L2, although they did not consider themselves to possess nativelike proficiency (higher than 85 on scale of 100). In the four experiments that included a judgment task (Experiments 1, 3, 4, and 5), we found that these moderately proficient L2 participants were more generous than L1 participants in accepting sentences that involved novel combinations of verb + construction, which each had a conventional alternative.

We found evidence to support the idea that the L2 participants were less aware of the competing alternatives than the L1 participants. Specifically, the L2 participants were less likely to suggest the intended competing alternatives as paraphrases of the unconventional sentences (Experiment 2); additionally, when we asked the L1 and L2 participants to choose whether the unconventional sentence or the competing alternative was preferable, the L2 participants were less likely to accurately choose the competing alternative sentences (Experiment 3). At the same time, if simple recognition that a competing alternative was preferable were sufficient to result in alignment of L1 and L2 judgments, we would have expected L2 judgments for those items for which the L2 participants did recognize that the competing alternative was preferable to pattern like those of the L1 participants. But a secondary analysis revealed that the L2 participants judged even those items more leniently (Experiment 3). Thus, it seems that L2 speakers do not take competing alternatives into account when judging novel sentences even when they recognize the competing alternatives to be more conventional. In Experiments 4 and 5, we investigated whether exposure to competing alternatives just before the judgment task would reduce

the discrepancy in judgments between L1 and L2 speakers. There was no discernable effect. That is, despite having just encountered acceptable competing alternatives, the L2 participants judged the unconventional sentences to be more acceptable than the L1 participants did; in fact, the L2 participants' judgments in the competing alternative condition were not different from L2 judgments in the control condition in which they had not encountered competing alternatives.

In Experiment 5, we tested whether participants were able to remember the competing alternative sentences that they had encountered with a verbatim memory recognition task that followed the collection of judgments. Here we found that the L2 participants' memory was less accurate than that of the L1 participants, and importantly, there was a significant correlation between accuracy on the verbatim memory task and the degree to which judgment scores aligned with those of the L1 participants. Moreover, when we looked at the verbatim memory results more closely, we found that the L2 participants were as accurate as the L1 participants at identifying old sentences, but differed specifically in being more likely than the L1 participants to incorrectly identify new sentences as having been encountered within the experimental context. The memory results are consistent with the idea that L2 speakers have somewhat noisier representations than L1 speakers do.

Noisy Representations for L2 Constructions

Noisier representations predict that L2 speakers should differ more from L1 speakers for novel combinations of verb and construction compared to familiar combinations. That is, familiar sentence types should be consistent with memory traces even if those traces are noisier, while less precise representations should present a particular issue for new sentence types because noisier or vaguer representations may erroneously be perceived as matches to novel sentences. We illustrate this idea in Figure 15. The colored clouds are intended to capture the representations of a familiar verb + argument structure combination for L1 speakers (Figure 15a) and L2 speakers (Figure 15b). We expect that a new instance of the conventional formulation will fall within the colored cloud for both L1 and L2 speakers, but an unconventional formulation would be more difficult for L2 speakers to distinguish because the boundaries of their representations are less clearly delimited.

We performed an analysis on the combined data in order to compare judgments on familiar and novel sentence types. There is evidence that L2 speakers are less confident (and so less extreme) in their judgments than are L1 speakers in that L2 participants' ratings were both higher for unconventional sentences *and* lower for conventional sentences. At the same time, as predicted by the

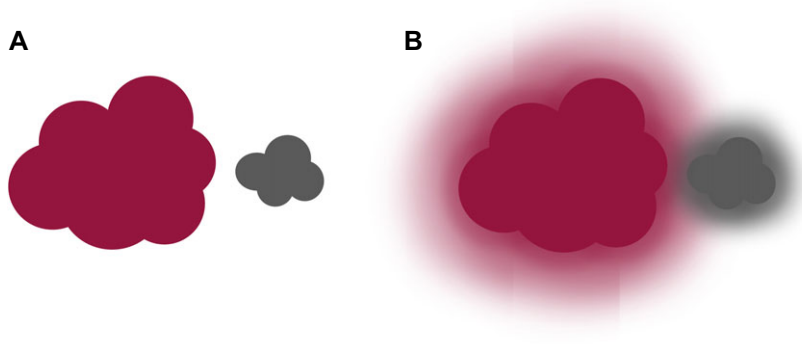

Figure 15 L1 speakers' more clearly delimited representations (A) and L2 speakers' noisier representations (B). [Color figure can be viewed at wileyonlinelibrary.com]

idea that L2 speakers' representations are noisier than those of L1 speakers, we found a greater discrepancy between groups in judgments of unconventional sentences than of the conventional sentences: The L2 participants differed more from the L1 participants on novel combinations of verb and construction than on familiar combinations. Indeed, a new analysis of the Robenalt and Goldberg (2016) data found the same interaction. Moreover, other researchers have previously observed a similar asymmetry between judgments of acceptable and unacceptable sentences for L2 speakers, where L2 speakers had more difficulty with implicit learning of unacceptable sentences (R. Ellis, 2004; Shirai, 2019). This suggests that L2 speakers are not equally challenged by all aspects of L2 but have particular difficulty with sentences that fall outside of their experience—the unconventional sentences.

L2 speakers may have noisier representations of English combinations of verb + construction because their input is itself noisier. In particular, many L2 speakers interact regularly with other L2 speakers, and this may lead them to repeatedly hear unconventional sentences produced by other L2 speakers (Ellis & Larsen-Freeman, 2006). Our sample was restricted to participants in the United States, but we did not have more precise information about the statistics of the English that they had encountered. We therefore acknowledge that the types of sentences that we considered unconventional may in fact be more conventionally used by L2 speakers. At the same time, this leaves us with a version of our original question: Why are L2 speakers more likely to produce unconventional sentences in the first place?

The current experiments confirmed the phenomenon that moderately proficient L2 speakers judge unconventional combinations of verb and construction as more acceptable than L1 speakers do. We have suggested that this effect is due to learners' having noisier linguistic representations and being less aware of more conventional competing alternatives. We further speculate that these two variables may be related to one another. In particular, L2 speakers may take less advantage of *statistical preemption*, that is, the competition between alternative ways of formulating a particular message in context. Researchers have argued that statistical preemption finetunes L1 speakers' knowledge of language (Boyd & Goldberg, 2011; Goldberg, 1995, 2011, 2019; Perek & Goldberg, 2017). As we reviewed in the Introduction, this idea predicts that L1 speakers should more readily accept novel combinations when no standard conventional alternative exists, and they appear to do so (e.g., Robenalt & Goldberg, 2015).

On a representational level, we believe that the competition between conventional ways of expressing particular messages sculpts the boundary conditions of how constructions are used. This would lead to clearly delimited representations of forms that are preferable compared to those that are not. Evidence that L2 speakers take less advantage of statistical preemption comes from the findings reviewed in the Introduction and in the current judgment data. L2 speakers do not appear to take competing alternatives into account in their judgments of unconventional sentences, except at the highest levels of proficiency when their judgments align with those of L1 speakers (cf. Figure 2 in Robenalt & Goldberg, 2016; see also Ambridge & Brandt, 2013; Zhang & Mai, 2018).

Why should statistical preemption have less of an effect on L2 speakers? There is growing evidence that L1 speakers anticipate or predict upcoming grammatical choices as they experience utterances unfolding (Dahan, Swingley, Tanenhaus, & Magnuson, 2000; DeLong, Groppe, Urbach, & Kutas, 2012; Kutas, DeLong, & Smith, 2011; Lew-Williams & Fernald, 2007; Wicha, Moreno, & Kutas, 2004). When L2 speakers encounter an unanticipated word or morpheme rather than an anticipated one, one can expect an error signal that will subsequently serve to slightly inhibit the anticipated form in favor of the encountered form when the same message in context is required next. However, L2 speakers appear to be less likely than L1 speakers to predict upcoming forms, even when they demonstrate knowledge of the forms during production and in offline tasks (Grüter, Hurtado, Marchman, & Fernald, 2014; Ito, Martin, & Nieuwland, 2017; Kaan, 2014; Kaan, Dallas, & Wijnen, 2010; Kaan, Kirkham, & Wijnen, 2016; Lew-Williams & Fernald, 2010; Martin et al., 2013), perhaps due to greater cognitive load from increased self-monitoring (Levelt, 1983) or the need to inhibit their L1 (Green, 1998). To the extent that

L2 speakers are less likely to predict upcoming grammatical choices, they will have less opportunity to learn from predictions that are subsequently falsified by the speech that they encounter.

Moreover, we speculate that L2 speakers' experience with between-language competition may affect their sensitivity to competition as a cue. We know that L2 speakers experience competition from their L1 to some extent, even in monolingual settings (Costa, 2004; de Groot, 1993; Marian & Spivey, 2003). This competition between words and constructions from the L1 needs to be ignored, and this may inadvertently reduce L2 speakers' sensitivity to competition effects when they use their L2. That is, perhaps L2 speakers are less affected by competition from the competing alternatives in their judgments because they have essentially needed to reduce the impact of competition when using their L2. Consistent with this idea is the fact that monolingual speakers are more likely to resist assigning a second label to a concept, presumably because of competition from the first label, while bilingual speakers are less resistant to accepting a second word label for a concept (Byers-Heinlein & Werker, 2009; Davidson & Tell, 2005). Further investigation into the possible effect of between-language competition on within-language competition is needed to clarify this possibility.

To summarize, we propose that for L1 speakers, competition leads to more clearly delineated representations of conventional forms, which allows speakers to clearly distinguish novel formulations that might be used to express the same message. L2 speakers, we suggest, show a reduced sensitivity to competition, which results in noisier representations.

Transfer and Proficiency Effects

An analysis of the combined data allowed us to consider possible transfer effects for the L2 participants with various L1s, with the most common L1 being Spanish. Although transfer effects have been well-documented in other studies, we did not find evidence for transfer as an explanation here for the L2 participants' tendency to accept unconventional sentences more readily than L1 participants. First, we found no difference among L2 speakers from a variety of L1 backgrounds. Second, English and Spanish share the same restrictions on one of the constructions we tested (clausal complement) but not the other (double-object), and yet participants whose L1 was Spanish showed the same overly generous judgments on novel instances of both constructions.

In line with previous findings, self-rated proficiency revealed a significant effect when L2 participants from the full self-rated proficiency range were included (Robenalt & Goldberg, 2016). That is, judgments of the L2 participants

at the very highest proficiency level aligned more closely with those of the L1 participants. But surprisingly, when we included only the participants who conformed to our preregistered criterion of self-rated proficiency (<85), the effect of proficiency was only marginal. No effect of age of acquisition or years speaking English was evident. Thus, it appears that that the type of novel combinations of verb + construction that we investigated remains challenging for L2 speakers even at moderate proficiency levels.

At the same time, it is important not to overstate the differences in ratings between L1 and L2 speakers. The L2 participants systematically and appropriately judged the unconventional sentences as less acceptable than the conventional baseline sentences. Also, although L2 participants were less likely than L1 participants to produce the intended competing alternative as a paraphrase of the unconventional sentences (Experiment 2), they did do so 65% of the time. The L2 participants also selected the competing alternative over the unconventional alternative (Experiment 3) 85% of the time. Finally, although the L2 participants' verbatim memory was not as accurate as that of the L1 participants, it was well above chance (Experiment 5). We expected the L2 participants' judgments to align more closely with the L1 participants' judgments at higher levels of proficiency, and we found support for this in the combined analysis. That is, self-rated proficiency correlated (inversely) with the L2 participants' judgments on the unconventional sentences, which is consistent with evidence that L2 speakers can become fully proficient in a L2 (Dussias, Marful, Gerfen, & Bajo, 2010; Foucart, Martin, Moreno, & Costa, 2014; Havik, Roberts, van Hout, Schreuder, & Haverkort, 2009; Hopp, 2013; Rossi, Diaz, Kroll, & Dussias, 2017). It remains quite possible that sufficient exposure to the relevant competing alternatives would induce more nativelike judgments; the designs in Experiments 4 and 5 provided only single instances of the conventional alternatives.

Limitations and Future Work

Although Experiments 4 and 5 showed that single exposures to conventional alternatives had no effect on subsequent judgments, we remain optimistic that greater exposure to the conventional alternatives would ultimately lead to a stronger preference for the conventional over the unconventional means of expression. If nativelike judgments of unacceptable sentences depend on implicit learning, then it is likely that continued repeated exposure over time is required (R. Ellis, 2004; McClelland, McNaughton, & O'Reilly, 1995; Shirai, 2019). We hope to explore whether more extended and repeated exposure to conventional alternatives will hasten the learning process in the future.

The current work failed to reveal strong effects of self-rated proficiency except when we included the entire set of L2 participants (with self-rated proficiency between 1 and 99). This may be in part due to the noisy nature of self-rated proficiency scores. Because we recruited our participants online (within the United States) to achieve a large sample, we were unable to determine their proficiency levels more precisely. Future work also needs to address whether a more accurate or more sensitive measure of proficiency or other aspects of language background would predict judgments more strongly.

The analysis of the subgroup of 150 native Spanish speakers revealed that simple transfer effects were unlikely to be the root cause of L2 speakers' more generous ratings on unconventional sentences given that Spanish only fails to make key distinction required for nativelike English judgments in one of the constructions tested, and yet judgments on both constructions were equally generous. Still, more careful consideration of various L1s and how they differ from English requires more investigation. Conversely, if the higher acceptability of unconventional sentences in L2 speakers is based on the effects of competition as suggested here, L2 speakers of languages other than English should display similar effects in those languages. Additionally, whether knowledge of multiple languages impacts the current findings is an open question.

In an effort to test new constructions using a large number of participants ($N = 980$) without making the manipulation too obvious (22 filler sentences were included), our preregistered studies only involved six target unconventional sentences involving two constructions. Future work needs to investigate the effects reported here on a larger scale, including more items as well as different constructions. The motivation for the judgment study came from reports of L2 production errors, but our work relied on explicit judgments rather than on more naturalist language use. Future work could use cued recall, reading times, or neural measures to determine in more implicit tasks the extent to which L2 speakers are more tolerant of unconventional formulations with competing alternatives (R. Ellis, 2004).

Conclusion

The current work demonstrates that moderately proficient L2 speakers are more tolerant of novel combinations of verb + constructions that have conventional competing alternatives than L1 speakers are, a finding consistent with previous research. The present five studies and the combined analysis begin to shed light on why this may be the case. First, L2 participants were less aware of the conventional competing alternatives of unconventional sentences than the L1 participants were, both in a paraphrase tasks (Experiment 2) and

in a forced-choice task (Experiment 3). Even when the L2 participants recognized that an alternative was preferable (Experiment 3) or were exposed to the alternative before the judgment task (Experiments 4 and 5), they remained more generous in their acceptance of unconventional combinations of verb + construction. We found a correlation between accuracy on the memory task and more nativelike judgments (Experiment 5), in that the L2 participants who were better able to recognize which of two formulations they had encountered produced judgments on unconventional sentences that aligned more closely with L1 participants' judgments. This finding is consistent with our proposal that L2 speakers' linguistic representations are noisier than those of L1 speakers.

Using combined judgment data across four experiments, we confirmed that the L2 participants' higher acceptability ratings of novel uses of verbs was not solely due to uncertainty: L2 participants' judgments on unconventional sentences diverged more from those of L1 participants' than did the L2 participants' judgments on conventional sentences. We also confirmed this same effect by reanalyzing data from Robenalt and Goldberg (2016). The combined analysis further suggested that L2 participants' more generous judgments of unconventional sentences were not easily attributable to transfer effects insofar as the effect held across various L1s. Moreover, in an analysis of a large subgroup of L1 Spanish/L2 English speakers, judgments were unaffected by whether the key distinction was made in Spanish or not. We suggest that L2 speakers have a reduced sensitivity to within-language competition, leading to noisier representations with less clearly delimited boundaries. This leads L2 speakers to more readily accept novel interpretable sentences for which conventional competing alternatives exist.

Final revised version accepted 27 June 2019

Notes

1 We confirmed this with a search of the Corpus of Contemporary American English for SPILL <pronoun> versus DRIP <pronoun> (search performed on 30 November 2018).

2 In fact, Robenalt and Goldberg (2015) found that, for L1 speakers, the inverse frequency effect was only evident for sentences that had a conventional competing alternative, that is, L1 speakers judged sentences such as "The teacher frowned a warning to the back of the class" to be as acceptable as "The teacher glowered a warning to the back of the class," even though *frown* is more frequent than *glower*. This is presumably because *frown* (not *glower*) is normally used to express the intended meaning of Example 5a.

3 Zhang and Mai (2018) combined ratings on familiar and novel denominal verbs.

4 In an abstract, Futrell and Gibson (2017) made an independent suggestion that L2 representations are noisier than L1 representations. They hypothesized that noisy representations may lead L2 speakers to be more likely to form shallow parses.
5 Preregistration ensures that planned analyses were actually planned in advance in order to guard against various issues that have been found to reduce replicability in experimental research (see Marsden, Morgan-Short, Thompson, & Abugaber, 2018).

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. https://doi.org/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. https://doi.org/10.1016/j.cognition.2006.12.015

Austin, G., Pongpairoj, N., & Trenkic, D. (2015). Structural competition in second language production: Towards a constraint-satisfaction model. *Language Learning*, *65*, 689–722. https://doi.org/10.1111/lang.12108

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

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

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. https://doi.org/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.

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. https://doi.org/10.1353/lan.2011.0012

Brooks, P., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, *75*, 720–738. https://doi.org/10.2307/417731

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic.

Developmental Science, *12*, 815–823. Retrieved from http://doi.org/10.1111/j.1467-7687.2009.00902.x

Costa, A. (2004). Speech production in bilinguals. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 201–223). Oxford, UK: Blackwell.

Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, *178*, 222–235. https://doi.org/10.1016/j.cognition.2018.05.018

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, *42*, 465–480. https://doi.org/10.1006/jmla.1999.2688

Davidson, D., & Tell, D. (2005). Monolingual and bilingual children's use of mutual exclusivity in the naming of whole objects. *Journal of Experimental Child Psychology*, *92*, 25–45. https://doi.org/10.1016/j.jecp.2005.03.007

Davies, M. (2008). *The Corpus of Contemporary American English: 400+ million words, 1990-present.* Retrieved from http://www.americancorpus.org

de Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 27–52). Amsterdam, Netherlands: John Benjamins.

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.

Dussias, P. E., Marful, A., Gerfen, C., & Bajo, M. T. (2010). Usage frequencies of complement-taking verbs in Spanish and English: Data from Spanish monolinguals and Spanish-English bilinguals. *Behavior Research Methods*, *42*, 1004–1011.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188. https://doi.org/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. Retrieved from http://doi.org/10.1093/applin/aml028

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, *13*, 161–186. https://doi.org/10.1017/S0272263100009931

Ellis, R. (2004). The definition and measurement of explicit knowledge. *Language Learning*, *54*, 227–275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Finn, A. S., & Hudson Kam, C. L. (2015). Why segmentation matters: Experience-driven segmentation errors impair "morpheme" learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1560–1569. Retrieved from http://doi.org/10.1037/xlm0000114

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*, 379–399. Retrieved from http://doi.org/10.1017/S1366728911000137

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. Retrieved from http://doi.org/10.1037/a0036756

Futrell, R., & Gibson, E. (2017). L2 processing as noisy channel language comprehension. *Bilingualism: Language and Cognition*, *20*, 683–684. https://doi.org/10.1017/S1366728916001061

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

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

Goldberg, A. E. (2019). *Explain me this: Creativity, competition and the partial productivity of constructions*. Princeton, NJ: Princeton University Press.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81. https://doi.org/10.1017/S1366728998000133

Grüter, T., Hurtado, N., Marchman, V. A., & Fernald, A. (2014). Language exposure and online processing efficiency in bilingual development. *Input and Experience in Bilingual Development*, *13*, 15–36. https://doi.org/10.1075/tilar.13.02gru

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

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. https://doi.org/10.1111/j.1467-9922.2009.00501.x

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, *9*, 220–225. https://doi.org/10.1016/j.tics.2005.03.003

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

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

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

Ito, A., Martin, A., & Nieuwland, M. (2017). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*, 574–588. Retrieved from http://doi.org/10.1016/j.jml.2013.08.001

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

Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. *Linguistics Today*, *164*, 207–214. https://doi.org/10.1075/la.164.23kaa

Kaan, E., Kirkham, J., & Wijnen, F. (2016). Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cogntion*, *19*, 1–18. https://doi.org/10.1017/S1366728914000844

Kang, J. (2017). Statistical preemption and Korean learners of English. *SNU Working Papers in English Language and Linguistics*, *15*, 40–52. Retrieved from http://hdl.handle.net/10371/139639

Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition*, *2*, 37–57.

Kellerman, E. (1995). Crosslinguistic influence: Transfer to nowhere? *Annual Review of Applied Linguistics*, *15*, 125–150.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York, NY: Oxford University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmer test package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition*, *14*, 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

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. https://doi.org/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. https://doi.org/10.1016/j.jml.2010.07.003

MacWhinney, B. (2006). Emergentism–Use often and with care. *Applied Linguistics*, *27*, 729–740. https://doi.org/10.1093/applin/aml035

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*, 97–115. https://doi.org/10.1017/S1366728903001068

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*, 321–391. https://doi.org/10.1111/lang.12286

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. https://doi.org/10.1016/j.jml.2013.08.001

Martinez-Garcia, M. T., & Wulff, S. (2012). Not wrong, yet not quite right: Spanish ESL students' use of gerundial and infinitival complementation. *International Journal of Applied Linguistics*, *22*, 225–244. https://doi.org/10.1111/j.1473-4192.2012.00310.x

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457. https://doi.org/10.1037/0033-295X.102.3.419

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. https://doi.org/10.1177/0267658310365786

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–226). London, UK: Longman.

Perek, F., & Goldberg, A. E. (2017). Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition*, *168*, 276–293. https://doi.org/10.1016/j.cognition.2017.06.019

R Development Core Team. (2008). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Robenalt, C., & Goldberg, A. E. (2015). Judgment and frequency 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*, 1–44. https://doi.org/10.1515/cog-2015-0004

Robenalt, C., & Goldberg, A. E. (2016). Nonnative speakers do not take competing alternative expressions into account the way native speakers do. *Language Learning*, *66*, 60–93. https://doi.org/10.1111/lang.12149

Rossi, E., Diaz, M., Kroll, J. F., & Dussias, P. E. (2017). Late bilinguals are sensitive to unique aspects of second language processing: Evidence from Clitic pronouns word-order. *Frontiers in Psychology*, *8*, 1–13. https://doi.org/10.3389/fpsyg.2017.00342

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

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*, 1–29. https://doi.org/10.1191/0267658306sr259oa

Selinker, L., & Lakshmanan, U. (1992). Language transfer and fossilization: The "Multiple Effects Principle." In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 197–216). Amsterdam, Netherlands: John Benjamins.

Shirai, Y. (2019). *Connectionism and second language acquisition*. New York, NY: Routledge.
Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development*, *19*, 15–34. https://doi.org/10.1016/j.cogdev.2003.08.001
Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition*, *15*, 181–204. https://doi.org/10.1017/S0272263100011955
Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288. https://doi.org/10.1162/0898929041920487
Zhang, X., & Mai, C. (2018). Effects of entrenchment and preemption in second language learners' acceptance of English denominal verbs. *Applied Psycholinguistics*, *39*, 413–436. https://doi.org/10.1017/S0142716417000406

Supporting Information

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

Appendix S1. Sentences Used in the Judgment Task.
Appendix S2. Preregistration Links.
Appendix S3. Descriptive Statistics for Various Participant Background Variables.
Appendix S4. R Code Used for Mixed-Effects Model Analyses.
Appendix S5. Short Stories and Comprehension Questions Used in Experiment 4.
Appendix S6. Items for the Memory Task in Experiment 5.

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

Nonnative Speakers Are More Tolerant of Unconventional Sentences and Have "Noisier" Knowledge of Acceptable Expressions in a Second Language

What This Research Was About and Why It Is Important

Native speakers are particular about which expressions are conventional and which they consider to be errors. For example, native English speakers strongly prefer "considered going" over "considered to go." The latter is judged to be unconventional, or an error, even though it is easily interpretable. In this study, the

researchers performed five experiments that investigated why nonnative speakers of English often continue to produce unconventional expressions despite being reasonably proficient and residing in an English-speaking environment, often for many years. The researchers found that nonnative speakers consistently judged unconventional formulations to be more acceptable than native speakers did. Nonnative speakers also had more trouble distinguishing conventional formulations from unconventional alternatives, which implied that they have "noisier," less clearly defined understanding of the expressions in their second language.

What the Researchers Did

- The researchers used the online crowdsourcing platform Amazon Mechanical Turk (https://www.mturk.com) to recruit 980 participants (490 native English speakers, 490 nonnative speakers of English), all from the United States.
- The researchers then compared native speakers and nonnative speakers on a number of short tasks, with 70 participants included in each group per task.
- In a judgment task, for example, the participants judged the acceptability of unconventional sentences (e.g., "Ken convinced that Laura clean her room," "Amber explained Zach the answer") and their more conventional alternatives (e.g., "Ken convinced Laura to clean her room," "Amber explained the answer to Zach"), using a 0–100 scale (0 = "absolutely not acceptable," 100 = "completely acceptable").
- To encourage nonnative speakers to recognize unconventional sentences as unconventional, the researchers had them read conventional alternatives like "Amber explained the answer to Zach" just before the judgment task. The idea was that conventional formulation would compete with their unconventional alternatives ("Amber explained Zach the answer") and would therefore lower the acceptability of these unconventional alternatives.
- To check whether the participants' ratings of acceptability reflected their memory of having read particular expressions before the judgment task, the researchers also exposed the participants to short stories containing target expressions and then asked them whether they had just read particular sentences or not.

What the Researchers Found

- The nonnative speakers were more accepting of unconventional sentences than the native speakers were.

- When participants were asked to paraphrase unconventional sentences, the native speakers were more likely than the nonnative speakers to converge on the same conventional alternative.
- Although the nonnative speakers tended to recognize the conventional sentences as such, they still judged the unconventional sentences more generously than the native speakers.
- Reading conventional alternatives before the judgment task did not change the nonnative speakers' judgments.
- The nonnative speakers did not perform as well as the native speakers at recognizing sentences they had just read, but those with better memory tended to judge unconventional sentences more like the native speakers.
- The various native languages of nonnative English speakers did not seem to affect their judgments.

Things to Consider

- Nonnative speakers' understanding of which expressions are possible in a second language appears to be noisier, or less clearly delimited, than native speakers' knowledge.
- This work documents and characterizes the issue which is important for teachers to understand, and also raises further questions about how multiple languages are learned.

How to cite this summary: Tachihara, K., & Goldberg. A. E. (2019). Nonnative speakers are more tolerant of unconventional sentences and have "noisier" knowledge of acceptable expressions in a second language. *OASIS Summary* of Tachihara & Goldberg in *Language Learning*. https://oasis-database.org