Judgment evidence for statistical preemption: It is relatively better to *vanish* than to *disappear* a rabbit, but a lifeguard can equally well *backstroke* or *swim* children to shore.

CLARICE ROBENALT & ADELE E. GOLDBERG
PRINCETON UNIVERSITY

ABSTRACT

How do speakers know when they can use language creatively and when they cannot? Prior research indicates that higher frequency verbs are more resistant to overgeneralization than lower frequency verbs with similar meaning and argument structure constraints. This result has been interpreted as evidence for *conservatism via entrenchment*, which proposes that people prefer to use verbs in ways they have heard before, with the strength of dispreference for novel uses increasing with overall verb frequency. This paper investigates whether verb frequency is actually always relevant in judging the acceptability of novel sentences or whether it only matters when there is a readily available alternative way to express the intended message with the chosen verb, as is predicted by *statistical preemption*. Two experiments are reported in which participants rated novel uses of high and low frequency verbs in argument structure constructions in which those verbs do not normally appear. Separate norming studies were used to divide the sentences into those with and without an agreed-upon preferred alternative phrasing which would compete with the novel use for acceptability. Experiment 2 controls for construction type: all target stimuli are instances of the caused-motion construction. In both experiments, we replicate the stronger dispreference for a novel use with a high frequency verb relative to its lower frequency counterpart, but only for those sentences for which there exists a competing alternative phrasing. When there is no consensus about a preferred way to phrase a sentence, verb frequency is not a predictive factor in sentences' ratings. We interpret this to mean that while speakers prefer familiar formulations to novel ones, they are willing to extend verbs creatively if there is no readily available alternative way to express the intended meaning.

Keywords: productivity in language; conservatism via entrenchment; statistical preemption; verb frequency; grammatical acceptability

1. INTRODUCTION

Speakers creatively extend the language they hear in various ways, and yet they systematically avoid certain other novel formulations. The issue of how speakers know that certain novel utterances are acceptable and others are not has been a fundamental question in linguistics and language acquisition for more than 40 years (e.g., Ambridge et al., 2008; Baker, 1979; Bowerman, 1988; Braine, 1971; Brown & Hanlon, 1970; Gennari & Macdonald 2008; Goldberg, 1995; Lakoff 1970; Pinker, 1989). Many constraints are motivated by general semantic, phonological, and/or syntactic facts, but in certain cases, formulations are semantically sensible, phonologically well-formed, and syntactically licensed in the language--and yet they

are still not acceptable. For instance, the examples in (1)-(2) are perfectly interpretable. They also make use of syntactic constructions that are unequivocally licensed in English, and yet both examples are decidedly odd:

(1) ??The magician vanished the rabbit.[1] (cf. The magician hid the rabbit.)
(2) ??She explained him the story.[2] (cf. She told/guaranteed him the story.)

Examples such as those in (1) and (2) occasionally occur in large corpora, but they are dispreferred compared to the nearly synonymous examples in (3) and (4) (see Goldberg, 2011 for a way to quantify the dispreference):

(3) The magician made the rabbit vanish.
(4) She explained the story to him.

How can people possibly learn that a non-occurring pattern is not fully acceptable? Clearly, the language that people hear does not come overtly marked with question marks or asterisks to indicate unacceptability. Moreover, we know that speakers are not reliably overtly corrected for producing ill-formed utterances. Parents, and caregivers more generally, are much more interested in the content of a child's utterance than in its form (Baker, 1979; Bowerman, 1988; 1996; Braine, 1971; Brown & Hanlon, 1970; Marcus, 1993; Pinker, 1989). Therefore, children cannot count on direct negative evidence to help constrain their subtle speech errors. And yet, by the time they reach adulthood, speakers reliably identify examples such as (1) and (2) as being relatively unacceptable. How do we come to acquire this knowledge?

1.2 Conservatism via Entrenchment

It is tempting to believe that the process of hearing a verb with sufficient token frequency plays a key role in preventing that verb from appearing in argument structure constructions that have not been previously witnessed with that verb (Braine & Brooks, 1995; Brooks & Tomasello, 1999). Proponents of this theory, which we term, *conservatism via entrenchment*, argue that positive evidence is the driving force behind constraints on novel verb uses. That is, verbs become increasingly entrenched in those argument structure constructions that they regularly occur in and it is claimed that they therefore become increasingly more resistant to being used in new constructions. Conservatism via entrenchment makes a clear prediction: the difference in acceptability between attested and unattested uses should increase with verb frequency; i.e., more frequent verbs should be less acceptable than less frequent verbs when used in constructions that are new for those verbs. Indeed, several labs have reported results that are consistent with this prediction.

Brooks et al. (1999) performed a production study with children ages 3, 4/5, and 8. An experimenter demonstrated four familiar actions that could be named by two verbs, a more

[1] We use "??" instead of the traditional "*" to indicate that the judgments of ill-formedness are gradient and dependent on many interacting factors.

[2] We recognize that there is a phonological preference for non-Latinate sounding words in the double-object construction (Gropen et al. 1989; Ambridge et al. 2012b), but as is evident from the acceptability of *guarantee* in the double object construction, this constraint does not fully account for the ill-formedness of (2) (Pinker 1989).

familiar (high frequency) verb and a semantically related less familiar (low frequency) verb. For each action, both verbs were modeled in a familiar argument structure construction (e.g., an intransitive frame for *disappear* and *vanish*), and then utterances were elicited that targeted a novel use of the verbs (i.e., a transitive use of *disappear* and *vanish*). Results demonstrated that across all ages, children were more likely to overextend low frequency verbs than high frequency verbs.

Theakston (2004) performed a judgment task with children (ages 5 and 8) and adults on 12 pairs of sentences involving low and high frequency verbs. The critical items all involved verbs used with argument structure patterns in which they did not normally appear. The data revealed that across ages, sentences containing low frequency verbs tended to be judged as more acceptable than sentences containing high frequency verbs. For example, *I dripped the carpet with juice* was judged to be more acceptable than *I spilled the carpet with juice* (*spill* is more frequent than *drip*). Ambridge et al. (2008) similarly performed a judgment task on three pairs of novel sentences involving intransitive verbs used transitively *(fall/tumble; disappear/vanish; laugh/giggle)*. They also taught participants novel verbs with meanings corresponding to "fall," "disappear," and "laugh." They, too, found an advantage for low frequency over high frequency verbs and an even greater advantage for novel verbs (which have a prior frequency of zero). These results are remarkable since, due to general fluency effects, sentences with higher frequency words might have been expected to be judged *more* favorably than semantically equivalent sentences with lower frequency verbs (e.g., Bornstein & D'Agostino, 1992). For example, *The girl liked flowers* can be expected to be judged as more acceptable than *The maiden adored petunias*. These results, however, indicate that this is not the case for verbs being used in novel ways. That is, *ceteris paribus*, a higher frequency verb appears to be judged *less* acceptable than a lower frequency verb when used in an argument structure construction in which the verb does not normally appear.

1.3 Statistical Preemption

Conservatism via entrenchment would seem to predict that higher frequency verbs should generally be unavailable for creative, novel uses, since they are highly entrenched in those argument structure frames in which they have been witnessed many times. However, one of the most remarkable aspects of language is that speakers do creatively produce sentences that they have not heard before. The following attested examples all involve fairly frequent verbs used in novel ways:

(5)a. "Some part of me wanted to yell my father into the world" (Modern Love column NYT June 13, 2013)
b. "he still didn't trust him as far as he could sneeze him" (Earth, Air, Fire and Custard by Tom Holt)
c. "Blinking the rain from my eyes…" (Donna Tart, *The Goldfinch*).

Thus, it is not clear that simply not having heard a verb in a particular construction leads learners to conclude that the formulation is necessarily unacceptable. One possibility is that the non-occurrence of a particular form *can* serve as evidence that the form is unacceptable if the form could reasonably have been expected to occur. For example, if a learner repeatedly hears a formulation, B (e.g., *make it disappear*), in contexts where she might have expected to hear a semantically and pragmatically related alternative formulation, A (e.g., *disappear it*), she could

come to recognize that B is the appropriate formulation in such a context and that A is less acceptable. This is essentially the notion of *statistical preemption,* which proposes that people can learn argument structure restrictions through indirect negative evidence (Boyd & Goldberg, 2011; Clark, 1987; DiSciullo & Williams 1987; Foraker et al., 2007; Goldberg, 1993; 1995; 2006; 2011; Marcotte, 2005; Pinker, 1981).

Preemption is a familiar process in morphology, where it has also been called "blocking." People learn to say *went* instead of *goed* because every time they might have expected to hear *goed*, they hear *went* instead. In the same way, *feet* preempts *foots,* and *referee* preempts *referer* (Aronoff, 1976; Kiparsky, 1982; Rainer, 1988).

Preemption between two phrasal forms requires some elaboration, since expressions formed from distinct phrasal constructions are virtually never semantically and pragmatically identical the way *feet* and the hypothetical *foots* would be, and therefore it is not clear that an instance of one phrasal pattern could preempt the use of another. For example, the double-object construction is distinct, at least in terms of its information structure, from the prepositional paraphrase (Arnold et al., 2000; Bresnan et al., 2007; Collins, 1995; Erteschik-Shir, 1979; Goldberg, 1995; 2006; Green, 1974; Hovav & Levin, 2008; Oehrle, 1975; Thompson, 1995; Wasow, 2002). Thus, knowledge that the prepositional paraphrase is licensed for *explain* should not in any simple way preempt the use of the double-object construction. And in fact, a large number of verbs do freely appear in both constructions (e.g., *tell*, *show*, *give*). This has led some researchers to dismiss preemption as a possible means by which overgeneralizations of phrasal patterns could be avoided (Bowerman, 1996; Pinker, 1989).

However, preemption can play an important role in learning to avoid certain formulations if a speaker's expectations are taken into account in the following way. Suppose learners witness repeated situations in which the double-object construction might be expected—exactly because the relevant information structure suits the double-object construction at least as well as the prepositional paraphrase. If, in these situations, the prepositional alternative is systematically witnessed instead, the learner can infer that the double-object is not, after all, appropriate (Goldberg, 1995; 2006). As Goldberg (2006) emphasizes, the process is necessarily statistical, because a single use of the alternative formulation could be due to some subtle difference in the functions of the two formulations that actually favors the alternative formulation in that context. It is also possible that a single use could simply be due to an error by the speaker. But if an alternative formulation is consistently heard when another is expected, a process of statistical preemption predicts that speakers will learn to use the provided alternative.

Statistical preemption presupposes that speakers attempt to anticipate others' utterances as they witness them. In fact, there is growing evidence that this is the case. We know that language is understood incrementally, with listeners anticipating upcoming words and constructions (Allopenna, Magnuson, & Tanenhaus, 1998; Arnold, Wasow, Losongco, & DeLong, Troyer & Kutas, 2014; Ginstrom, 2000; Elman 2004; Gibson, Bergen & Piantadosi 2013; Johnson, Turk-Browne and Goldberg 2013; Jaeger & Snider 2013; Kamide, Altmann, and Haywood 2003; Lew-Williams & Fernald 2007; Marslen-Wilson, 1973; Pickering & Garrod, 2013; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Traxler et al. 2000). For example, the less predictable an upcoming word in a sentence is, the stronger the N400 component in electroencephalography — an event-related potential related to semantic processing (King & Kutas, 1995; Kutas & Hillyard, 1984). Prediction is possible because language involves probabilistic regularities that constrain our expectations and provide cues about upcoming words, phrases, and content (Bates & MacWhinney, 1987; Glenberg & Gallese,

2012; Goldberg, Casenhiser, & Sethuraman, 2005; Pickering & Garrod, 2004, 2007). Stephens, Silbert, & Uri Hasson (2010) have found that listeners' brains activity literally aligns to a great extent with the speaker's; strikingly, the extent to which the listeners' activity in the relevant regions anticipates the speakers' has been found to correlate with behavioral measures of comprehension.

Statistical preemption of phrasal forms has only received attention in the experimental literature fairly recently. Two of the earliest studies have shown that hearing novel intransitive verbs in periphrastic causative constructions significantly preempts children's use of those verbs in simple transitives (Brooks & Tomasello, 1999; Brooks & Zizak, 2002). For example, Brooks and Tomasello (1999) found that children aged six and seven were less than half as likely to productively use a novel verb causatively when the verb had been modeled in both the intransitive and in a periphrastic causative, than when it was only modeled in the intransitive form. For example, if the child had heard both *The cow is chamming* and *Ernie's making the cow cham*, they were less likely to respond to "what did Elmo do to the cow?" with *Ernie chammed the cow* (causative), than they were if only the intransitive construction had been witnessed. Hearing the novel verb used in the periphrastic causative construction provided a readily available alternative to the causative construction, statistically preempting the use of the latter (cf. also Tomasello, 2003).

In another study, Boyd and Goldberg (2011) investigated a certain class of "a-adjectives" that begin with a schwa and disprefer appearing prenominally (e.g., ??*the asleep boy*). Adult productions were elicited in a naturalistic situation, resulting in either a relative clause or prenominal use of a targeted real or novel adjective:

(6) Prenominal: The *sleepy/??asleep/?adax* fox moved to the star.
(7) Relative Clause: The fox that's *sleepy*/*asleep/adax* moved to the star.

Results show that witnessing a novel *a*-adjective used in a preemptive context (the relative clause) dramatically decreases prenominal uses. Moreover, speakers did *not* display an increased avoidance of prenominal uses when exposed to contexts like (8), presumably because they rationally attributed *adax*'s appearance in the relative clause to the complex adjective (*proud of itself*, cf. 9), rather than to *adax* alone.

(8) The hamster, adax and proud of itself, moved to the star.
(9) *The proud of itself hamster moved to the star.

This result supports the statistical preemption hypothesis because the prenominal form is only avoided if it could have been expected and a competing form was witnessed instead. Recasts of a child's utterance by an adult can be viewed as a type of statistical preemption when the child implicitly recognizes that the adult's formulation differs from what the child might have used in the same context; and recasts have in fact been found to be quite helpful (Saxton et al. 1998; Chouinard & Clark 2003).

Most of the evidence in support of statistical preemption to date has been based on speakers' productions, not on acceptability or grammaticality judgments. And yet the ultimate aim is to account for the judgments involved in examples like those in (1) and (2). Previous attempts to investigate whether acceptability judgments are predicted by conservatism via entrenchment or statistical preemption have yielded mixed results. In a series of studies,

Ambridge, Rowland, Pine and colleagues (2012a, 2012b, 2012c) investigated certain alternations in which particular verbs only appeared in one variant. Conservatism via entrenchment predicts that those verbs' cumulative frequencies (in all other constructions) should inversely correlate with acceptability of the verb in the non-occurring construction, such that higher frequency should result in lower judgments. Statistical preemption, on the other hand, predicts that only verbs' frequency in a competing construction should be relevant; a verb that occurs with higher frequency in a competing construction should be judged less acceptable, but the frequency of the verb in other constructions should not be relevant.

When verbs occur in a preempting construction – as they did in the Ambridge et al. studies – the statistics that are relevant to conservatism via entrenchment are not independent of those that are relevant to statistical preemption. Instances of the verb in a preempting construction are a subset of the verb in all other constructions; in this case, it might be expected that *both* measures correlate with judgments of ill-formedness. This does in fact appear to be the case. While Ambridge et al. (2012a; 2012b) found that statistical preemption provided no additional effect over and above conservatism via entrenchment, using a larger corpus Ambridge et al. (2012c) found the reverse: that conservatism via entrenchment provided no additional effect over and above statistical preemption.

The present studies thus aim to explore whether and to what extent speakers judge verbs used in novel ways to be relatively acceptable and when they are more likely to reject novel uses. The following two studies are designed to decouple evidence for statistical preemption from evidence for conservatism via entrenchment. To that end, two factors are primarily investigated: the relative frequency of the main verb and whether there exists a readily available alternative way to express the intended message. The critical stimuli involve verbs in constructions with which they do not normally occur, and the stimuli systematically vary whether the novel verb use has a readily available alternative formulation. If overall verb frequency is the primary determinant of acceptability, then whether or not the expression involving a novel use of a verb has a readily available alternative should not matter; across the board, lower frequency verbs should be more acceptable than higher frequency verbs when used in constructions that are novel for those verbs. On the other hand, if competition between formulations plays an important role, then lower frequency verbs should only be more acceptable than higher frequency verbs when there exists a competing formulation.

2. Experiment 1

Experiment 1 asked native speakers to judge three types of sentences on a scale of acceptability. They judged novel sentences that had a readily available alternative formulation, novel sentences that did not have a readily available alternative formulation, and familiar or baseline sentences. All sentences were created in pairs: one version involved a lower frequency main verb, and the other of which involved a higher frequency main verb; each pair of sentences was otherwise identical.

2.1 Methods

Materials

Fourteen verb classes were selected in which one high and low frequency alternative were identified for each class (14 verb pairs; 28 verbs).[3] High frequency verbs were, on average, 15.74 times more frequent than their low frequency counterparts according to verb frequencies collected from COCA, the Corpus of Contemporary American English (made freely available by Mark Davies at http://corpus.byu.edu/coca; Davies 2008). Each pairing of high and low frequency verb was assigned a "baseline" sentence frame, in which the verbs were used in their preferred argument structure. For example, one verb pair was *disappear* and *vanish*. Since these verbs are typically used intransitively, the baseline sentence was intransitive: *Ashley <disappeared/vanished> into the darkness*.

A second sentence frame was also generated for each verb pair, with the aim of creating sentences that were semantically sensible but *novel*: each sentence frame used an argument structure pattern that was highly unusual for the verbs. For *disappear* and *vanish*, the novel sentence frame was transitive: *The magician <disappeared/vanished> the rabbit.* All novel sentences were created by the authors and rated for plausibility as described below. By creating stimuli that included either a high frequency verb or a lower frequency verb, while the rest of the sentence was held constant, we are able to compare like with like, as previous studies investigating the role of verb frequency have done (Brooks et al. 1999; Ambridge et al. 2008; Theakston 2004). Thus we are not investigating various degrees of verb frequency, but whether a lower frequency verb is more acceptable in a novel sentence than a higher frequency verb, when other factors are held constant.

Baseline (familiar) sentences for high and low frequency verb pairs

Will slept/napped on the sofa.

Laurie smiled/grinned.

The boys jumped/hopped on the trampoline.

The students laughed/chuckled.

Alex swam/backstroked to the dock.

Taylor sang/crooned a lullaby to the baby.

The coach shouted/hollered at the players.

The professor explained/recited the assignment.

Christina cried/sobbed when her hamster died.

The spy forced/coerced the criminal to confess.

Brandon fell/tumbled out of the tree.

[3] The original list of materials contained sixteen verb classes and a total of 32 verbs. One verb pair was excluded because the COCA frequency for even the low frequency verb was nearly twice that of the high frequency verb in any other verb pair. The second pair was excluded because COCA searches revealed that the novel use was actually attested with some regularity in the corpus. The results reported below show the same significant patterns whether these two verb pairs are included or excluded.

The family considered/contemplated going to Disneyland.

Emily found/located the book she needed.

Ashley disappeared/vanished into the darkness.

Novel sentences without a clear competing alternative phrasing (novel-noCA)

Jeff slept/napped the afternoon away.

Megan smiled/grinned her boyfriend out the front door.

Terry's horse jumped/hopped her straight out of the saddle.

The chief will laugh/chuckle you back to your desk job.

The lifeguard swam/backstroked the children to shore.

The performer sang/crooned the audience into another dimension.

The shopkeeper shouted/hollered the teenagers out of the building.

Novel sentences with a clear competing alternative phrasing (novel-hasCA)

Amber explained/recited Zach the answer.

Anthony's merciless teasing cried/sobbed his little sister.

Daniel forced/coerced that Helen compete.

Jacob fell/tumbled the lamp.

Kayla's boss considered/contemplated to give her a raise.

Please find/locate a new pen to me.

The magician disappeared/vanished the rabbit.

Table 1: stimuli used in Experiment 1 with novel-hasCA and novel-noCA binned according to paraphrase norming study.

To determine whether the sentences are in fact highly unusual for the verbs used, we performed searches on the full 450 million word COCA corpus, which roughly approximates the number of words that a highly educated college student has encountered in her entire lifetime (Levy et al., 2012)**.** Since the corpus is tagged but not parsed, all sentences returned via searches were hand-checked for the intended structure and meaning. The search strings, raw number of times each sentence type occurs, main verb frequency, and percentage of times the verb occurs in the sentence type of interest (columns C/D) are provided in Online Appendix Table 1. Although some of the sentence types did occur in the full corpus—a point we return to in Experiment 2—they accounted for less than 0.12% of occurrences for each verb.

The novel sentence frames were binned into two categories via results from a norming study. One set of sentences had a competing alternative (novel-hasCA); that is, the sentences had a readily available alternative phrasing that was preferred to the formulation used in the study. The other set of sentences did not have a competing alternative (novel-noCA); i.e., respondents showed little consensus about a better way to rephrase the sentence. The procedure of the

norming study is detailed in the following section; a full list of items and various measures from the paraphrase ratings can be found in Online Appendix Table 2.

In Experiment 1, we did not systematically control for construction, and the sentences binned into the hasCA category included verbs in the transitive, double-object, and verb phrase complement constructions, while sentences binned into the noCA category were instances of the caused-motion construction. We address this confound with a second experiment described in Section 3.

2.1.1 *Paraphrase Norming Study 1: Identifying competing alternatives*

The purpose of the first norming study was to divide the novel sentences into two categories, depending on whether there was a readily available alternative formulation for each sentence. To this end, we assessed whether participants consistently generated an alternative formulation to express the meaning of each novel sentence—and if so, whether they were in agreement as to what that alternative formulation should be.

Participants

41 participants were recruited from Mechanical Turk. Responses from one participant were excluded for failure to follow instructions. Participants were compensated for their participation at a rate that is consistent with other online studies of comparable difficulty and duration. Participation lasted approximately ten minutes.

Procedure

Participants were presented with fourteen novel sentences and asked to decide whether there was a better way to express the meaning of each sentence. If they thought the sentence was acceptable as written, they were instructed to copy the sentence exactly. Otherwise, they were asked to provide a paraphrase. Baseline sentences were not included in this portion of the study because they were designed to be perfectly acceptable; there was no reason to expect participants to generate alternative formulations for those items.

Participants saw novel sentences from one of four lists. Each list contained only one version of each sentence (with either the high or low frequency verb). Although only two lists were necessary to ensure all of the items would appear once, the lists were duplicated and reordered to guard against order effects. Both versions of each list were pseudorandomized.

Before participants were shown any of the critical items, three practice sentences were provided. The first two sentences were filled in for the participant, including one fully grammatical sentence that was reproduced without revision and one sentence with word order violations that was rewritten. The third practice sentence was a free response paraphrase that participants could choose to rewrite or leave as written. All practice sentences were adapted from Ambridge et al. (2008).

The addition or removal of an adverb from a novel sentence was not counted as a paraphrase if the argument structure of the critical verb was unaltered (e.g., changing *The chief will laugh you back to your desk job* to *The chief will laugh you right back to your desk job* did not count as a paraphrase but as an exact repetition). Sentences that were ungrammatical or grossly misinterpreted were labeled as "other" and excluded from analyses (38 sentences; 6% of the norming data). Two coders (the first author and a research assistant naïve to the experimental hypotheses) reviewed the responses and categorized them according to whether each sentence had been accepted as written, or had been paraphrased as grammatical and with

the intended meaning. Coders agreed upon 588/640 sentences (92% of all responses). The disputed cases were arbitrated by a third coder, also naïve to the experimental hypotheses.

In order to determine whether each sentence had a competing alternative, we examined how consistently participants generated an identical paraphrase for a particular sentence. During this process, we considered only the responses from sentences involving high frequency verbs; this was because participants often replaced the low frequency verb with its high frequency counterpart in their paraphrases. Such behavior artificially inflated the diversity of paraphrases for sentences involving low frequency verbs. Low frequency items were given the same classification (hasCA or noCA) as their high frequency counterpart.

The results of the paraphrase norming study are shown in Figure 1 (sample responses with ratings are provided in Online Appendix Table 3). Each sentence was reviewed by 20 independent raters. Ungrammatical or misinterpreted items were removed from the pool of responses for each sentence, and the maximum number of times a single paraphrase was repeated for each sentence was tallied. Figure 1 plots the proportion of total responses that corresponds to the most frequent paraphrase. Taller bars indicate high paraphrase consistency across the independent raters. The color coding of the bars indicates the result of a median split that was used to divide the sentences into the two categories (hasCA or noCA). The sentences represented by darker bars were considered to have a competing alternative; the average number of repeated paraphrases for stimuli in this category was 14.14. The sentences represented by lighter bars were *not* considered to have a competing alternative; the average number of repeated paraphrases for stimuli in this category was only 2.43.

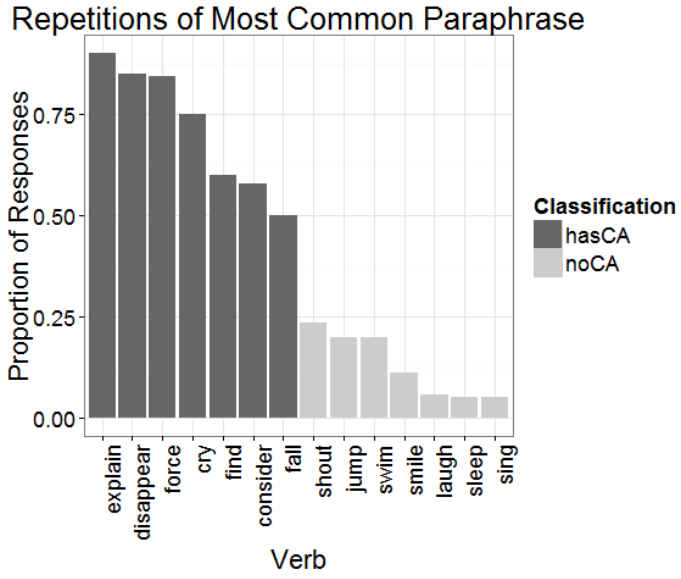

Figure 1: the proportion of responses that corresponded to the most common paraphrase for that sentence is plotted above. The verbs are used to index the sentence stimuli—see Online Appendix Table 2 for a complete list of sentences. Only high frequency verbs were used to create the hasCA/noCA bins; the low frequency counterpart was assigned to the same category after the median split on the high frequency items.

2.1.2 Plausibility Norming Study 1

The purpose of the second norming study was to determine the plausibility of each novel sentence. This norming study was conducted to be certain that we were assessing changes in grammatical acceptability independent of the plausibility of each situation.

Participants

20 participants were recruited from Mechanical Turk. Participants were compensated for their participation at a rate consistent with other online studies of comparable difficulty and duration. Participation lasted under ten minutes.

Procedure

Four lists of sentences were created. Each list contained 14 baseline sentences, 7 novel sentences with no competing alternative, and 7 novel sentences with a competing alternative. Each critical verb appeared exactly once on each list; e.g., if the low frequency verb appeared in a novel sentence, then the high frequency verb appeared in the baseline sentence. All lists were pseudorandomized to prevent clusters of sentences from the same category. A second version of each list was created to guard against order effects for a total of eight lists. Each participant was randomly assigned to one of these lists.

Participants were asked to read each sentence and decide whether the situation it described was plausible. They responded using a five point Likert scale for which a rating of 1 indicated that the situation was completely implausible and a rating of 5 indicated that the situation was completely plausible. Before witnessing any experimental items, participants were given two practice trials to ensure they understood the rating system. The first trial contained two sentences: one completely plausible and one completely implausible sentence. The ratings for these sentences were given and a brief explanation of the answers was provided. The second trial contained three sentences with varying degrees of plausibility. This trial was free response and provided optional textboxes in which participants could justify their rating.

Pilot data indicated that participants struggled to rate sentences for plausibility when they felt the sentences were grammatically unacceptable. To avoid this confound in our ratings, we retrieved the most common response from the first norming study (whether it was a paraphrase or an exact repetition) and substituted that sentence for the critical item. We felt this method best maintained the semantic integrity of each situation while removing effects of questionable grammatical acceptability. Average plausibility ratings are shown in Table 2 in the Results section, and results for each item are provided in Online Appendix Table 4.

2.1.3 Judgment Study 1

Participants

108 native English speakers were recruited from Mechanical Turk. Participants were modestly reimbursed for their participation. The task lasted approximately ten minutes.

Procedure

Item lists were created using the procedure described for the plausibility ratings. Participants rated sentences for grammatical acceptability using a five point Likert scale. A rating of 5 indicated that the sentence was completely acceptable, a rating of 1 indicated that the sentence was completely unacceptable, and a rating of 3 indicated that the sentence was somewhere in between. Instructions and practice sentences were adapted from Ambridge et al. (2008).

2.2 Results and discussion of Experiment 1

Average acceptability scores of baseline, hasCA, and noCA sentences are plotted in Figure 2. Table 2 displays the average acceptability and plausibility ratings for each sentence type at both levels of verb frequency.

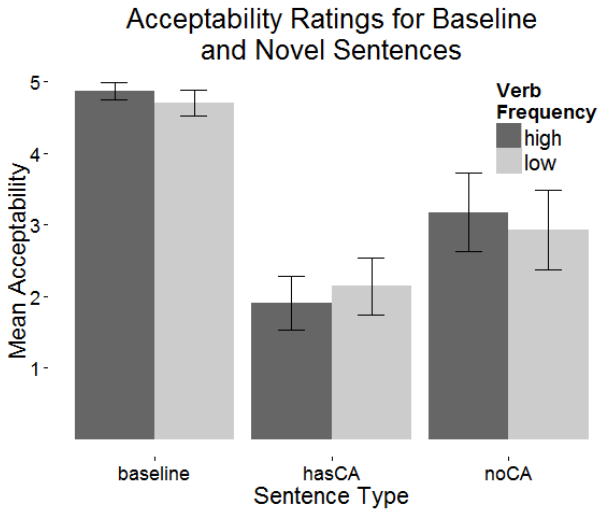

Figure 2: Average acceptability ratings for each sentence type (baseline, novel-hasCA and novel-noCA), for verbs of high and low frequency. Error bars represent standard error.

A mixed linear model was run to predict acceptability judgments from the sentence category (baseline, novel-hasCA, novel-noCA) and verb frequency (high, low). The data were treatment coded and, following Barr et al. (2013), we used the maximal random effects structure for which the model was able to converge. This included by-subject random intercepts and a random slope and intercept for the high versus low frequency version of each verb pair. Model comparisons were used to ensure that the use of random effects resulted in a significantly better model fit. The final model is provided in Table 3.

Sentence Ratings			
Sentence Category	Verb Frequency	Acceptability	Plausibility
Baseline	High	4.87 (0.44)	4.33 (0.89)
	Low	4.70 (0.66)	4.03 (0.82)
Novel-hasCA	High	1.91 (1.07)	4.78 (0.12)
	Low	2.14 (1.12)	4.56 (0.35)
Novel-noCA	High	3.17 (1.34)	3.78 (1.06)
	Low	2.93 (1.36)	3.32 (0.70)

Table 2: Average acceptability and plausibility ratings for baseline, novel-noCA, and novel-hasCA sentences. Standard deviations are in parentheses.

Mixed Linear Model for Acceptability Rating

Random effects:

Groups	**Name**	**Std.Dev.**
Subject	(Intercept)	0.27
VerbPair	(Intercept)	0.45
	FREQ. (slope)	0.83

Fixed effects:

	Estimate	**Std. Error**	***t* value**	***p* value**	**Sig.**
(Intercept)	6.15[4]	0.41	14.84		
Sentence Length	-0.23	0.02	-14.93	<0.001	***
Plausibility	-0.01	0.09	-0.11	0.91	
High → Low freq	-0.18	0.07	-2.60	.009	
Baseline → HasCA	-2.97	0.06	-50.72	<0.001	***
Baseline → NoCA	-1.18	0.08	-15.47	<0.001	***
Low freq * HasCA	0.45	0.08	5.59	<0.001	***
Low freq * NoCA	-0.003	0.08	-0.04	0.97	

Table 3: Mixed effects model of results of Experiment 1. Sentence length, plausibility, verb frequency (high, low), and sentence type (baseline, hasCA, noCA) are treated as fixed effects. Random effects included by-subject random intercepts and a random slope for the high versus low frequency version of each verb pair.

Plausibility and sentence length were considered as possible covariates in order to control for possible general item-level factors. Plausibility ratings (from the second norming study) did not add significantly to the model—in fact, while the noCA stimuli were judged to be more acceptable than the hasCA stimuli, the noCA stimuli were on average numerally *less* plausible than the hasCA stimuli (noCA = 3.54; hasCA = 4.67)**.** We return to this point in the general discussion. Sentence length was included because we did not control for target construction; it was found to be a significant predictor of acceptability, with an estimate of -0.23 points per word (t =-2.60, p = 0.009). The consideration of these covariates ensures that the comparisons that are of theoretical interest have already accounted for differences in sentence length and plausibility. That is, statistically significant comparisons regarding sentence acceptability cannot be explained away by appealing to such item-level differences.

Main effects of both noCA and hasCA sentences indicate that all novel sentences are less acceptable than baseline sentences. Novel-hasCA sentences judged to be -2.97 points less than baseline sentences and novel-noCA sentences judged to be -1.18 points less than baseline sentences (t = -50.72 and -15.47, respectively; $p < 0.001$ for both comparisons). This means that participants preferred verbs used in familiar ways over novel formulations, as is to be expected

[4] The intercept is higher than the highest possible rating (5) because all factors, including the number of words, are negative and therefore subtracted from the intercept. A five word sentence is predicted to have a rating of 6.15 + (5 x -.23) or 5.0 before other factors are taken into account.

on any usage-based account, and as is specifically predicted by conservatism via entrenchment. The significant result of High → Low (verb) frequency indicates that baseline sentences (the reference level of the model) with lower frequency verb were found to be less acceptable than their higher frequency counterparts. In other words, c*eteris paribus*, more familiar sentence frames are more acceptable.

Of central importance is the fact that verb frequency only affects novel sentences that have a competing alternative. That is, higher verb frequency predicts less acceptability only for sentences that have a readily available competing alternative formulation. While sentences are typically more acceptable with high frequency verbs, novel-hasCA sentences show the reverse pattern (with low frequency verbs perceived as 0.45 points more acceptable than high frequency verbs, $t = 5.59$, $p < 0.001$). The use of a high versus low frequency verb has no effect on the predicted acceptability for a novel sentence that does not have a readily available alternative formulation (estimated change = -0.003 points, $t = -0.04$, $p = 0.97$). The suggestion is that this is because what matters is the frequency of the verb in its competing alternative; if there is no competing alternative for a particular sentence, then verb frequency is not predictive. This is predicted by statistical preemption.

We are assuming here that, for sentence types that have a preemptive alternative formulation, the frequency of that formulation will vary depending on the frequency of the verb. That is, the alternative will be more frequent for higher frequency verbs than for otherwise matched lower frequency verbs. We do not attempt to gather numbers for how often each verb occurs in its competing alternative, because whether or not the alternative formulation truly competes depends on contextual factors that would require prohibitively time-consuming hand-coding of tens of thousands of examples. For example, a sentence like *She made him cry* competes with a simple causative *??She cried him*, only if the causation is construed to be direct in the particular context. If the causation is construed as indirect, then the periphrastic causative is to be preferred for independent, semantic reasons (see Boyd & Goldberg 2011, experiment 3 for discussion). The interaction of verb frequency and sentence type (hasCA vs. noCA) is predicted by statistical preemption.

Statistical preemption also predicts that ceteris paribus, novel sentences that do not have a competing alternative should be more acceptable than those that do. In order to compare novel sentences directly, the mixed linear model described above was rerun with novel-hasCA as the reference point for sentence category. This model demonstrated that hasCA sentences were in fact judged to be less acceptable than novel-noCA sentences. Specifically, novel-hasCA sentences were predicted to lose 1.68 points (on a five point acceptability scale) relative to novel-noCA sentences ($t = 17.96$, $p < 0.001$). Some caution must be used in interpreting this comparison since the stimuli were binned from the outset according to how often the sentences were paraphrased the same way. Exact repetitions were allowed for sentences that were deemed acceptable, and these repetitions were not counted as paraphrases (recall Figure 1). Thus, novel sentences that tended to be considered acceptable (and therefore repeated verbatim) were more likely to be classified as not having a competing alternative. Thus, the key evidence in favor of statistical preemption is the interaction between verb frequency and whether a sentence has a competing alternative as described above, which does not suffer from this potential confound.

The inclusion of baseline sentence ratings in the model in Table 3 raises the question of whether, in the case of the hasCA sentences, judgments on novel sentences with low frequency verbs are lower in absolute terms than those with high frequency verbs, or whether the higher frequency verbs are simply judged as more acceptable in their corresponding baseline sentences.

That is, when difference scores are used, as they often are (e.g., Ambridge et al. 2008), it is unclear whether low frequency verbs are actually *more* acceptable than novel sentences with high frequency verbs or whether, instead, they are less different than their baseline versions than are high frequency verbs (Ambridge, Pine and Rowland 2012). In order to investigate this question, we redid the analysis in Table 3 without the baseline sentences (see Table 4).

Mixed Linear Model for Acceptability Rating

Random effects:

Groups	**Name**	**Std.Dev.**
Subject	(Intercept)	0.49
VerbPair	(Intercept)	0.76
	Frequency (Slope)	0.34

Fixed effects:

	Estimate	**Std. Error**	***t* value**	***p* value**	**Sig.**
(Intercept)	3.59	1.30	2.76		
Sentence Length	-0.26	0.14	-1.87	0.06	.
Plausibility	-0.04	0.17	-0.22	0.83	
High → Low freq.	0.28	0.15	1.85	0.06	.
HasCA → NoCA	1.58	0.51	3.11	0.002	**
Low freq. * NoCA	-0.49	0.21	-2.40	0.02	*

Table 4: Mixed effects model of results of Experiment 1 without baseline sentences: hasCA and noCA sentence types are compared directly

When the analysis is conducted with novel sentences only, we find similar patterns of significant results. Sentence length remains a predictor of acceptability, although the finding is only marginally significant (-0.26 points per word, t = -1.87, p = 0.06). Plausibility is not a significant predictor of acceptability. The marginally significant result of High → Low frequency indicates that for hasCA sentences, low frequency items are 0.28 points more acceptable than high frequency items (t = 1.85, p = 0.06). This finding does not hold for noCA sentences, for which low frequency items are -0.49 points less acceptable than high frequency items (t = -2.40, p = 0.002). Again, this interaction is predicted by statistical preemption; conservatism via entrenchment would predict a preference for low frequency items in novel sentences across the board. The strongest finding from this analysis is the difference in acceptability between hasCA sentences and noCA sentences, with the latter estimated to be 1.58 points more acceptable than the former.

While Experiment 1 considered effects of plausibility, sentence length, and relative verb frequency, one factor that was not systematically controlled for was the target construction. In fact, most of our noCA stimuli involved the "caused-motion" construction (Goldberg 1995), whereas the hasCA stimuli made use of the double-object construction, the simple transitive, and a verb phrase complement constructions. This potential confound of construction type does not undermine the key finding related to verb frequency. That is, conservatism via entrenchment

predicts that lower frequency verbs should be more flexible than high frequency verbs when used in any novel way, and so it should predict relatively lower acceptability for higher frequency verbs, regardless of what the novel construction is or whether the novel construction has a competing alternative. Results indicate, however, that verb frequency played a role in the hasCA condition only.

Still, we know that various constructions are productive to varying degrees, and, in fact the caused-motion construction is among the more productive constructions (Goldberg 1995). Thus the overall difference in acceptability between the hasCA and noCA conditions may be attributable to the distinct target constructions involved. Another limitation of Experiment 1 was that several of our "novel" sentence types actually occurred in the 450 million word COCA corpus, albeit they never accounted for more than 0.12% of any verbs' use. The highest "novel" use involved the verb, *tumble,* which occurred in the target caused-motion construction 34 times (out of 4884 instances in total). Finally, we know that verb semantics plays a key role in the acceptability of verbs in argument structure constructions. We attempted to control for this by comparing pairs of verbs with related meanings. However, we could not investigate the role of semantics systematically because of the variety of constructions used, since each construction is associated with its own particular semantics (e.g., Goldberg 1995).

To address these issues, we performed a second experiment in which all of our target stimuli were instances of the same general caused-motion construction. Experiment 2 takes semantic features of the main verbs into account. It also uses a more conservative measure of what is considered a novel sentence type: the general type of sentence can not appear more than 10 times in the 450 million word corpus. In fact, the maximum number of any sentence type was 7, and the majority of the sentence types used did not occur even once.

3. EXPERIMENT 2: Judgment study 2 on novel caused-motion sentences with and without a competing alternative

Experiment 2 uses the same basic design as Experiment 1, with a few exceptions. In Experiment 2, we control for the construction used in all the stimuli: they are all instances of the "caused-motion" construction. The caused-motion construction is associated with a [Verb NP PP_{path}] frame, and, as its name suggests, it is associated with a meaning of caused motion (Pinker 1989; Goldberg 1995; Johnson & Goldberg 2013). Also, in Experiment 2, we collected paraphrase data, acceptability judgments, and plausibility judgments at the same time, as described below.

In a separate norming study, we collected ratings on three semantic features for each verb used. This is because, very generally, we know that the meaning of a verb must be compatible with the meaning of the construction. This has been termed the Semantic Coherence principle (Goldberg 1995: 50): all obligatory participant roles of the verb must be construable as instances of the more general argument roles of the construction. Specific semantic factors such as manner and end-state have been shown to correlate with acceptability in the caused-motion construction when compared with a causative variant (e.g., *She filled the tub with water*) (Gropen et al. 1991; Ambridge et al. 2008). We speculate that force in a direction may be another key semantic factor in the acceptability of novel caused-motion expressions, so ratings on all three features: manner, end-state, and force in a direction were collected for each verb (in isolation) as described in section 3.1.2.

3.1 METHODS

Materials

20 verb classes were selected and one high and low frequency alternative were identified for each class (20 verb pairs; 40 verbs). High frequency verbs were, on average, 6.94 times more frequent than their low frequency counterparts according to verb frequencies in COCA. We again performed searches on the full 450 million word COCA corpus, as in Experiment 1, in order to determine whether the stimuli were in fact novel. The sentence types accounted for no more than 7 tokens, amounting to a proportion of lower than 1 in 1000 for each verb. More than half of the sentence types used as stimuli did not occur even once in COCA. The sentences used in Experiment 2 is provided in Table 5.

Novel sentences without an agreed upon competing alternative phrasing (Novel-noCA)
The sound rattled/ reverberated the bats out of their hiding place.
The lifeguard swam/ paddled a pool toy to the kids.
The teacher frowned/ glowered a warning to the back of the class.
The crowd cheered/ hollered the reluctant candidate to the podium.
The editorial embarrassed/ mortified the poor man out of town.
The lion roared/ snarled the veterinarian out of the enclosure.
The editor smiled/ grinned the new reporter into his office.
The woman screamed/ shrieked the children out of the ice cream store.
The magician fascinated/ enthralled the toddlers into a trance.
Andrew insulted/ derided the potential member out of the club.
Novel sentences with an agreed upon competing alternative phrasing (Novel-hasCA)
The scientist infected/ corrupted bacteria into the sample.
The children soiled/ splotched mud onto the carpet.
The designer decorated/ embellished lace onto the invitation.
The dictator flooded/ inundated propaganda into the city.
The chef coated/ doused ranch dressing over the salad.
The housekeeper soaked/ drenched bleach into the towel.
The nurse bound/ bandaged cotton over the wound.
Natalie smacked/ swatted a newspaper onto the mosquito.
The landscaper surrounded/ bordered rocks around the garden.
The camper blocked/ obstructed a heavy backpack into the entrance.

Table 5: Stimuli used for Experiment 2: caused-motion sentences with high and low frequency verbs. Plausibility ratings and semantic ratings on each verb for manner, end-state, and force in a direction, were included in analyses.

For a complete list of sentences, COCA search strings, and token counts, see Online Appendix Table 5.

As in Experiment 1, the novel sentence frames were binned into two categories as determined by how consistently they were paraphrased. One set of sentences had a competing alternative (novel-hasCA). The other set of sentences did not have a competing alternative (novel-noCA). We consider a gradient factor, the proportion of shared paraphrases (which we refer to as PropPar) as well, as discussed below.

3.1.1. Paraphrase, Plausibility, and Judgment Study 2

Participants

50 native English speakers were recruited from Mechanical Turk. Participant were modestly reimbursed for their participation. The task lasted approximately 35 minutes.

Procedure

Each participant was presented with 10 novel sentences, which were embedded in a list of 26 sentences. The sixteen additional sentences were used to ensure that participants understood how to rate for plausibility versus acceptability: eight were high in both acceptability and plausibility, four were high in acceptability but low in plausibility, and four were high in plausibility but low in acceptability. All participants successfully rated the high acceptability/high plausibility sentences as highly acceptable, and 43 participants successfully differentiated plausibility and acceptability on the remaining sentences (successful discrimination was defined as a mean difference of 2 points or greater between acceptability and plausibility on items that were designed to differentiate the two). Data from the seven remaining participants were excluded from results reported in the following sections.

Participants saw sentences from one of four lists, assigned randomly. Each list contained only one version of each sentence (with either the high or low frequency verb). All lists were pseudo-randomized to ensure that the first four sentences were catch sentences (high in acceptability or plausibility, but not both). For each sentence in the survey, participants were asked to provide an acceptability rating, a plausibility rating, and a paraphrase of the sentence if they felt there was a more natural way to express its meaning. A seven point Likert scale was used for the acceptability and plausibility ratings, with 7 representing maximum acceptability/plausibility and 1 representing the minimum for each measure. If participants could not think of a more natural way to express the meaning of a novel sentence, they were instructed to copy and paste the sentence exactly as written.

Criteria for repetitions were the same as before. Coders agreed on 98% of all responses. The disputed cases were adjudicated through discussion. In order to determine whether each sentence had a competing alternative, we examined how consistently participants generated an identical paraphrase for a particular sentence. During this process, we considered only the responses from sentences involving high frequency verbs, to be consistent with the procedure in Experiment 1. The low frequency counterparts of each high frequency item received the same classification as the high frequency item (hasCA or noCA). The results of the paraphrase norming study are shown in Figure 3.

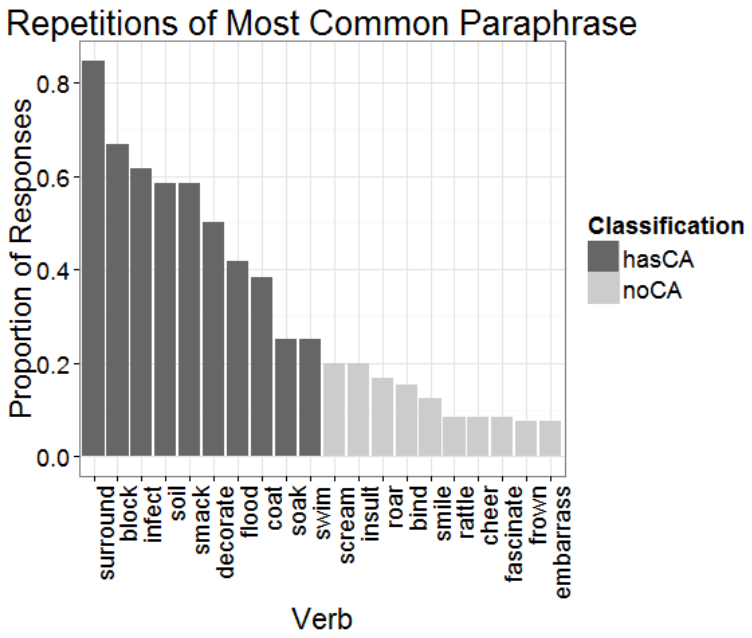

Figure 3: Results of Paraphrase norming study: the height of the bar represents the proportion of usable responses that correspond to the most common paraphrase for that sentence. A median split on this measure divided our sentences into hasCA/noCA bins, which are represented by dark bars and light bars respectively. A gradient measure, the proportion of shared paraphrases (PropPar) was used in a separate analysis. The verbs are used to index the sentence stimuli—see Online Appendix Table 5 for a complete list of sentences. Only high frequency verbs were used to create the hasCA/noCA bins; the low frequency counterpart was assigned to the same category after the median split on the high frequency items.

3.1.2 Semantic features of manner, end-state, and force ratings for Experiment 2

Participants

We collected ratings for three semantic features: manner, end-state, and force in a direction, from a separate group of 13 participants on Mechanical Turk. Participants were modestly compensated for their participation at the same rate as in Experiment 1.

Procedure

Participants were provided with the infinitive form of 23 verbs and asked to rate the extent to which each verb described a manner, an end-state, and force (in a direction) of an action. Ratings were given on a five point Likert scale in which 5 indicated "completely" and 1 indicated "not at all." The first three verbs were test items that were selected to be high in manner ("to skip"), end-state ("to open"), and force in a direction ("to shove"), respectively. Participants were also given detailed explanations for each of these semantic factors, including examples of verbs that should have been rated high, low, or in between. Five of the thirteen participants did not provide correct ratings for these practice items; data from these participants was excluded from further analysis. Ratings from the remaining eight participants were averaged for each verb and used as covariates in the following analyses.

3.2 RESULTS

The complete set of acceptability, plausibility, and semantic feature ratings are provided in Online Appendix Table 6. Average acceptability scores of hasCA, and noCA sentences are plotted in Figure 4.

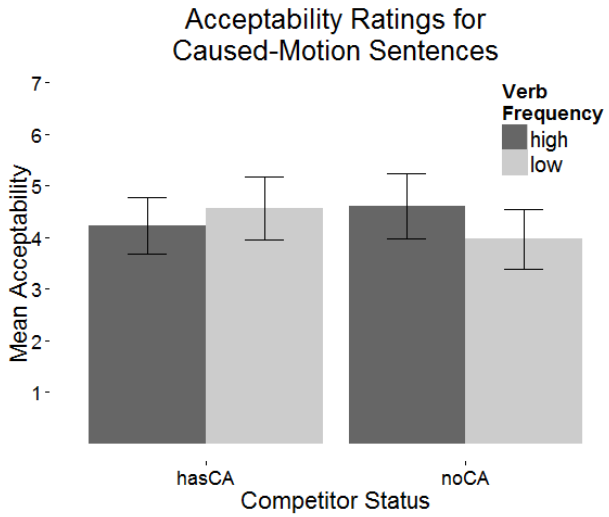

Figure 4. Average acceptability ratings for novel-hasCA and novel-noCA sentences, for verbs of high and low frequency. Error bars represent standard error.

Significant results cannot be gleaned directly from the bar graph because the raw acceptability averages do not factor out the effects of other factors. Therefore, as in Experiment 1, a mixed linear model was run to predict acceptability judgments. Factors included were: sentence category (hasCA, noCA), verb frequency (high, low), plausibility, and the semantic features manner, force in a direction, and endstate. Sentence length was not included as a factor in Experiment 2 because all of the sentences ranged between 6-8 words and had the same major constituents. The data were treatment coded and the random effects structure was the same as that used in Experiment 1 (by-subject random intercepts and random intercepts and slopes for the high versus low frequency version of each verb pair). The model is provided in Table 6.

Mixed Linear Model for Acceptability Rating

Random effects:

Groups	**Name**	**Std.Dev.**
Subject	(Intercept)	0.56
VerbPair	(Intercept)	1.11
	frequency (slope)	0.86

Fixed effects:

	Estimate	**Std. Error**	***t* value**	***p* value**	**Sig.**
(Intercept)	-1.40	1.30	-1.08		
Plausibility	0.44	0.06	7.50	<0.001	***
Manner	0.40	0.20	1.97	0.05	*

Force	0.13	0.20	0.67	0.50	
EndState	0.21	0.18	1.18	0.24	
High → Low freq.	0.75	0.34	2.17	0.03	*
HasCA → NoCA	0.99	0.57	1.72	0.09	.
Low freq. * NoCA	-1.27	0.48	-2.64	0.008	**

Table 6: Mixed effects model of results for Experiment 2. Fixed effects included: Plausibility, semantic factors (manner, force, and endstate), verb frequency (high, low), and sentence type (hasCA, noCA). Random effects included by-subject random intercepts and a random slope for the high versus low frequency version of each verb pair.

Experiment 2 included the semantic features of manner, force in a given direction, and end-state as predictors but only manner was significant (estimate = 0.40 points multiplied by the manner rating for that item, $t = 1.97$, $p = 0.05$). That is, verbs that were judged to encode a manner component were found to be more acceptable in the caused-motion construction, as has been found previously. That force and end-state were not significant is not altogether surprising, since each of the sentences was compared with another sentence that was identical except for the main verb involved. These verb pairs were chosen so as to differ in frequency while being minimally different in semantics. The fact that only one of our semantic features was predictive indicates that the semantics was in fact held reasonably constant between the high and low frequency verbs.

Plausibility was a significant factor in Experiment 2, with speakers judging more plausible sentences to be more acceptable. Recall this was not the case in Experiment 1, which is a puzzle that will require more investigation. Participants in Experiment 2 (but not Experiment 1) were asked to provide both acceptability and plausibility ratings at the same time in an effort to encourage people to distinguish between the two; it could be that simultaneously providing both judgments had the opposite effect. Another explanation is that plausibility exerts a greater influence on acceptability when the construction of interest is held constant, which was true only for Experiment 2.

The key finding of Experiment 1 is replicated in Experiment 2: there is a significant interaction between verb frequency and sentence type (hasCA *vs.* noCA), with higher verb frequency predicting lower acceptability judgments for the hasCA sentences only. The novel noCA sentences with lower frequency verbs are not any more acceptable than noCA sentences with higher frequency verbs. This is the effect predicted by statistical preemption. Recall statistical preemption suggests that, *ceteris paribus,* speakers may use verbs in novel ways as long as there does not exist a readily available alternative way to express the same message. We thus hypothesize that it is not the frequency of the verb overall that is critical, but the frequency of the verb in a competing alternative: the more entrenched the competing alternative is, the less acceptable the target novel formulation is judged to be. We also again see an overall acceptability advantage for noCA sentences as indicated by the positive intercept for hasCA → noCA. This, too, is predicted by statistical preemption. However, this latter finding must be interpreted with caution because of the way the hasCA and noCA stimuli were binned as discussed above.

The model in Table 6 bins sentences categorically into two groups, hasCA and noCA. But as is evident in Figure 3, the split between these two groups is gradual. In fact, Figure 3 suggests a gradient measure of the degree to which a competing alternative is readily available:

the probability with which native speakers suggest the same alternative paraphrase. Specifically, we calculated the proportion of paraphrases in the norming study that were agreed upon across participants. When this measure, proportion of shared paraphrases, or PropPar, is included in the model instead of the binary hasCA/noCA distinction, we find that the model fits the data just as well, with one less degree of freedom. This model is provided in Table 7.

Mixed Linear Model for Acceptability Rating

Random effects:

Groups	**Name**	**Std.Dev.**
Subject	(Intercept)	0.55
VerbPair	(Intercept)	0.92
	frequency (Slope)	0.78

Fixed effects:

	Estimate	**Std. Error**	***t* value**	***p* value**	**Sig.**
(Intercept)	0.62	1.06	0.59		
Plausibility	0.44	0.06	7.67	<0.001	***
Manner	0.19	0.20	0.94	0.35	
Force	0.23	0.18	1.33	0.18	
Endstate	0.22	0.16	1.38	0.17	
High → Low freq.	0.02	0.22	0.10	0.92	
PropPar	-3.39	0.81	-4.18	<0.001	***

Table 7: Mixed linear model for acceptability in Experiment 2, using continuous measure, the proportion of shared paraphrases (PropPar) for strength of competing alternative rather than median split

We reran the data from Experiment 1 using PropPar, however, and this factor was not significant, so the general appropriateness of this measure requires further research.

6. DISCUSSION

Consistent with previous findings, the present two experiments suggest that novel sentences with lower frequency verbs are judged to be (relatively) more acceptable in novel formulations than are novel sentences with higher frequency verbs. Thus, *the magician vanished the rabbit* is less acceptable than *the magician disappeared the rabbit*. The effect held when compared to baseline sentences (i.e., intransitive uses of *disappear* and *vanish*) in Experiment 1 and in absolute terms in both experiments (recall Experiment 2 did not include baseline sentences). But importantly, this frequency effect only held for novel sentences that had a competing alternative.

Whether sentences had a competing alternative was determined by separate norming studies, in which participants paraphrased each novel sentence. When the majority of people agreed upon the same paraphrase, we considered the sentence to have a competing alternative. In the case of *disappear*, for example, the same paraphrase, a periphrastic causative, *The magician made the rabbit disappear*, was suggested by 18 out of 20 participants. (A different, gradient way of determining the *degree to which* a sentence had a periphrastic causative was to use the proportion of shared paraphrases provided by participants. This measure correlated with acceptability judgments in Experiment 2, but not in Experiment 1).

In the case of novel sentences that do not have a readily available alternative, such as *The lifeguard swam/backstroked the children to shore*, the frequency of the verb (here, *swim vs. backstroke*) does not correlate with acceptability, either when compared directly, or when considered in relative terms by considering the difference between the novel and baseline sentences in Experiment 1 (e.g., *swim* and *backstroke* used intransitively). We hypothesize that the frequency effect is due to the frequency of a competing alternative *if there is one,* and not to the overall frequency of the verb in question. The effect of the frequency of competing alternatives is predicted by Goldberg (2011) in terms of the "confidence" of statistical preemption. That is, confidence that one form statistically preempts another was conjectured to vary with the frequency with which the preempting form has been witnessed when the target form may have been expected to occur.

In both experiments, novel sentences for which there was no competing alternative were judged to be somewhat more acceptable overall than sentences for which there was a readily available alternative. This is also predicted by statistical preemption: people are expected to be willing to use verbs creatively as long as the semantics is suitable and as long as there is no already entrenched way to express the intended message. A caveat is required in interpreting the advantage for sentences that had no competing alternative, however. Recall that we asked people to paraphrase the sentence, *if they could think of a better way to formulate it.* We then used the number of shared paraphrases to determine which novel sentences would be considered to have a competing alternative and which not. Sentences that were considered acceptable as written would be paraphrased less often and therefore lowering the opportunities to for shared paraphrases. This potential confound is tricky to avoid if it is in fact the case that sentences are more acceptable when there is no readily available alternative. There seems to us to be a clear sense in which *She vanished the rabbit* is simply "wrong" while *She backstroked the children to shore* is simply novel, but future work is required to better operationalize and test this intuition.

Clearly, the semantic fit between a verb and construction is in general, a strong and important predictor of acceptability (e.g., Ambridge et al. 2008; Goldberg 1995).[5] We found some effect of the degree to which a verb encoded manner in judgments of acceptability in Experiment 2, where the target construction was the caused-motion construction. We did not find significant effects of endstate or force in a direction, most likely because we were comparing pairs of sentences that contained closely related verbs. In this way, semantics was largely controlled for, as we had intended.

It is very clear from the results in Experiment 1 that native speakers find sentences that involve verbs used in familiar ways to be markedly more acceptable than verbs used in novel ways, whether or not the novel sentence has a competing alternative. That is, the baseline sentences were rated as much more acceptable than either group of novel sentences. Speakers find, for example, *She backstroked in the pool all afternoon* to be much more acceptable than

[5] At the same time, as Pinker (1989) noted, it can be difficult to disentangle semantic judgments about how well verbs fit a particular construction, from speaker's knowledge of how those verbs are used. While English speakers would likely rate *fill*, for example, as not having a manner component and "therefore" being unavailable for use in the caused-motion construction, the German cognate verb, *füllen*, does readily occur in the caused-motion construction in German (Ambridge & Brandt 2013), as does a verb that is translated as *fill* in Korean (Bley-Vroman & Joo 2001).

The lifeguard backstroked the children to shore, even when sentence length and plausibility are factored out. This indicates that speakers know which verbs they have witnessed in which constructions, and they prefer familiar uses over novel ones. We also found in Experiment 1, that in the case of baseline (familiar) sentences, participants judged sentences containing high frequency verbs to be more acceptable than sentences with lower frequency verbs. Thus witnessing a verb in a construction more often renders it more acceptable. In this way, speakers are conservative: they prefer familiar formulations, and the more familiar, the better.

4. Conclusion

Are all novel uses of verbs judged to be equally unacceptable? Clearly, the answer is no. In line with previous findings, the present two experiments suggest that, when other factors are held constant, novel sentences with lower frequency verbs are judged to be more acceptable in novel formulations than are higher frequency verbs (Brooks et al. 1999; Theakston 2004, Ambridge et al. 2008). That is, for example, *she vanished a rabbit* is judged to be more acceptable than *she disappeared a rabbit*. But crucially, the present experiments demonstrate that this effect only holds when the novel formulation has a competing alternative. For sentences for which there is no readily available alternative way to express the same message, verb frequency is not predictive. For example, participants judge *The lifeguard swam children to shore* to be just as acceptable as *the lifeguard backstroked children to shore*. The suggestion then is that the frequency effect is due to the frequency of a competing alternative, if there is one, and not to the overall frequency of the verb in question.

Conservatism via entrenchment hypothesizes that speakers prefer to use verbs in familiar ways, and that the more a verb is witnessed, the less available it will be for novel uses. The first part of this idea was resoundingly supported in Experiment 1. Sentences with verbs used in familiar ways are preferred over sentences with verbs used in novel ways, and verbs that have been witnessed frequently in a given construction are judged to be more acceptable than verbs that have been witnessed less frequently in that construction. However, the second aspect of the prediction was not supported: it does not appear to be the case that more frequent verbs are more reluctant to be used in novel ways across the board.

Statistical preemption predicts that novel sentences with high frequency verbs will only be rated as less acceptable than their low frequency counterparts when the novel use has a readily available competing alternative. This prediction was supported by both experiments. That is, in the case of novel sentences without a competing alternative, verb frequency made no difference in judgments of acceptability. This finding supports the idea that speakers take competing alternatives into account; i.e., a readily available alternative formulation statistically preempts a novel form, when such a readily available alternative formulation exists. Participants are equally willing to extend a frequent as an infrequent verb for use in a new way as long as there is no readily available alternative way to express the intended meaning.

It is important to note that statistical preemption and a tendency toward conservatism are not necessarily at odds with one another, and we find evidence of this in our data (cf. also Ambridge et al. 2012c; Boyd, Ackerman, and Kutas, 2012). *Ceteris paribus,* people do prefer formulations they have witnessed before (they are somewhat conservative); there is a decrement in acceptability for verbs used in any new construction. Verbs used in *novel ways* are recognized as novel, but there is no indication that the sentences are judged in comparison to all other uses of the verbs. We interpret the results as follows: people accrue positive evidence of how verbs

(and other words) are used. Importantly, this evidence is not only about the verbs' formal distribution but also about the way the verbs' have been used in various situations such as causing-to-move type situations. When a speaker finds herself wishing to express a message that is relatively similar to one she has heard expressed before, she most naturally produces the same construction that had been witnessed previously for her intended message. In this way, speakers tend to be conservative. If one construction has consistently been witnessed in such a situation, other potential competitors become less available; speakers can essentially learn in this way that the witnessed construction is the conventional way to express the intended message and the potential competitors are not conventional. This is how statistical preemption works. When the speaker finds herself wishing to express a message of a type she has not previously witnessed, she does the only thing she can do: she creates a novel usage, one that does not have a readily available alternative. This situation not only allows verbs to be used creatively, but also allows nouns to be used as verbs (Clark & Clark 1979).

(10) We were wardrobing her for her nightly chat show.
http://www.slate.com/articles/life/doonan/2013/06/joan_rivers_80th_birthday_can_we_talk
(11) Verizon is trying to squirrel me into upgrading (Kim and Davies GloWbE-US)

Not all constructions lend themselves equally well to having readily available alternative formulations. The meanings of some constructions fill quite specific semantic niches for which there is no readily available alternative phrasing; and many constructions occur with such low frequency that they cannot be reasonably expected to occur (Goldberg, 1995). Intriguingly, many low frequency constructions or constructions that fill unique semantic niches appear to be quite productive, modulo semantic constraints. For example, almost any activity that can serve as the means of metaphorical motion despite difficulty can appear in the *way* construction as in (12), and almost any intentional activity that can be construed to be effortful and ongoing can appear in the construction in (13):

(12) She maneuvered/kicked/crawled her way down the street.
(13) She worked/studied/read/wrote/ran/jogged/drove her ass off.

It is possible that alternations in the sense of Levin (1993) offer the best situations for statistical preemption to take place. For example, the double-object construction and the *to*-dative are very similar in meaning and are used in very similar contexts. When a verb appears exclusively in one construction, such as *say* in the *to*-dative, it creates a prime opportunity for statistical preemption to take place. Construction-level alternations are not necessarily required, however; any situation in which a listener repeatedly witnessed a close alternative phrasing to the meaning he or she was expecting could constitute evidence for statistical preemption. Notice that one explanation of why (14) is less than fully acceptable is that it is statistically preempted by (15):

(14) ?She cried herself asleep.
(15) She cried herself to sleep.

In fact, there is only one example in COCA of *cry* <reflexive pronoun> *asleep*, while there are 119 examples of *cry* <reflexive pronoun> *to sleep*, even though *asleep* is slightly more frequent

overall than *to sleep.* It could also be that when the "edits" required to change a witnessed form to a conventional competing alternative are small, judgments of acceptability do not decrease as much as when they are larger. For example, the difference between *asleep* and *to sleep* is smaller than the difference between *explain me something* and *explain something to me* (since the latter requires a change in word order and the addition of a preposition). The competing alternatives used in the present Experiments 1 and 2 differed from the novel sentences in relatively large ways, involving both word order changes and prepositional changes. It could well be that the difference in acceptability between sentences that have a preemptive alternative and those that do not are mitigated when the differences between the novel and familiar formulations are smaller (see also Gibson et al. 2013; Theakston 2004).

Since statistical preemption requires that learners amass data about constructional uses in contexts, it requires more experience with particular formulations than if learners were simply observing the overall frequencies of verbs. Therefore statistical preemption of phrasal argument structure constructions as investigated here may be more relevant for older children and adults, than for young children. There is some evidence that this is the case; Brooks & Tomasello (1999), for example, found that children were not sensitive to statistical preemption in the case of the causative alternation until they were six – seven years of age. Moreover, there is reason to believe that non-native English speaking adults with less than fully proficient speaking ability are less able to take competing alternatives into account; in particular, Robenalt & Goldberg (to appear) use the stimuli and design reported here as Experiment 1 on non-native speakers as well as on a new group of native speakers. The effects reported here were replicated for native speakers, but non-native speakers do not prefer novel sentences with lower frequency verbs to those with higher frequency verbs for either hasCA or noCA sentences; in this way, non-native speakers show less evidence of using statistical preemption. That is, except at the highest levels of speaking proficiently, non-native speakers do not appear to take competing alternatives into account in their judgments the same way that native speakers do. At the highest proficiency levels, non-native speakers' judgments align with native-speakers.

For the researcher and the learner, statistical preemption requires nuanced awareness of constructional choices in context. The present study suggests that future work will be well served by recognizing the availability of alternative formulations as a factor in perceptions of acceptability. The present results indicate that speakers prefer to use familiar formulations, and that preference is stronger when the familiar formulations are especially familiar (more frequent). At the same time, we know speakers occasionally produce creative, novel utterances. We suggest that they do so under communicative pressure to convey a particular message when no familiar formulation exists. A lifeguard may swim or backstroke children to shore, because there is no readily available alternative way to express this (important!) message.

ACKNOWLEDGEMENTS

We would like to thank Jessica Hao and Laura K. Suttle for their assistance with paraphrase coding, and Ewa Dąbrowska, John Newman, and two anonymous reviewers of an earlier version of this paper for helpful and constructive feedback. Laura K. Suttle also provided valuable statistical help. We are also grateful for funding from the Einstein Foundation in Berlin to AEG.

REFERENCES

Ambridge, B., & Brandt, S. (2013). Lisa filled water into the cup : The roles of entrenchment , preemption and verb semantics in German speakers ' L2 acquisition of English locatives . *ZAA*, *6*(3), 245–263.

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.

Ambridge, B., Pine, J.M., & Rowland, C.F. (2012a). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, 123(2), 260-79.

Ambridge, B., Pine, J.M., Rowland, C.F., & Chang, F. (2012b). The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument structure overgeneralization errors. *Language*, 1-60.

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

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

Aronoff, M. (1976). Word Formation in Generative Grammar. Cambridge, MA: MIT Press.

Baker, C.L. (1979). Syntactic Theory and the Projection Problem. *Linguistic Inquiry*, 10(4), 533-581.

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*(02), 207-219.

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

Bowerman, M. (1996). Argument structure and learnability: Is a solution in sight? In J. Johnson, M. L. Juge, & J. L. Moxley (Eds.), *Proceedings of the Twenty-second Annual Meeting of the Berkeley Linguistics Society, February 16-19, 1996. General Session and Parasession on The Role of Learnability in Grammatical Theory* (pp. 454-468). Berkeley Linguistics Society.

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

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(1), 55-83.

Boyd, J. K., Ackerman, F., & Kutas, M. (2012, November). Adult learners use both entrenchment and preemption to infer grammatical constraints. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on* (pp. 1-2). IEEE.

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

Braine, M.D.S. & Brooks, P.J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W.E. Merriman (Eds.), *Beyond names for*

things: Young children's acquisition of verbs (pp. 353-376). Hillsdale, NJ: Lawrence Erlbaum Associates.
Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R.H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer & J. Zwarts (Eds), *Cognitive Foundations of Interpretation*, 69-94. Amsterdam: Royal Netherlands Academy of Science.
Brooks, P. & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language*, 75, 720-738.
Brooks, P.J. & Zizak, O. (2002). Does preemption help children learn verb transitivity? *Journal of Child Language*, 29, 759-781.
Brooks, P.J., Tomasello, M., Dodson, K., & Lewis, L.B. (1999). Young Children's Overgeneralizations with Fixed Transitivity Verbs. *Child Development*, 70(6), 1325-1337.
Brown, R. & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J.R. Hayes (Ed.), *Cognition and the development of language*. New York, NY: Wiley.
Chouinard, M.M. & Clark, E.V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30(3), 637-669.
Clark, E. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney, *Mechanisms of language acquisition*, Hillsdale, NJ: Erlbaum.
Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 767-811.
Collins, P. (1995). The indirect object construction in English: An informational approach. *Linguistics*, 33, 35-49.
Davies, M. The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present. 2008. *Available online a t http://www. americancorpus. org*.
DeLong, K., Troyer, M., & Kutas, M. (2014). Pre-Processing in Sentence Comprehension: Sensitivity to Likely Upcoming Meaning and Structure. *Language and Linguistics Compass*, *8*(12), 631–645. doi:10.1111/lnc3.12093
Di Sciullo, A.M. & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.
Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in cognitive sciences*, *8*(7), 301-306.
Erteschik-Shir, N. (1979). Discourse constraints on dative movement. In T. Givón (Ed), *Syntax and Semantics 12: Discourse and Syntax*, 441-467. New York, NY: Academic Press.
Foraker, S., Regier, T., Khetarpal, N., Perfors, A., & Tenenbaum, J.B. (2007). Indirect evidence and the poverty of the stimulus: The case of anaphoric one. In D.S. McNamara & J.G. Trafton, (Eds), *Proceedings of the Twenty-Ninth Annual Cognitive Science Society*, 275-80. New York: Lawrence Erlbaum Associates.
Gennari, S.P. & Macdonald, M.C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, 58(4), 161-187.
Gergely, György, Harold Bekkering & Ildikó Király. 2002. Rational imitation in preverbal infants. Nature 415.755.
Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051-8056.
Goldberg, A.E. (1993). Another look at some learnability paradoxes. In E.V. Clark (Ed.), *The Proceedings of the Twenty-Fifth Annual Child Language Research Forum*, 60-75. Stanford, CA: Center for the Study of Language and Information.

Goldberg, A.E. (1995). Constructions: A Construction Grammar Approach to Argument Structure. Chicago, IL: Chicago University Press.
Goldberg, Adele E. 2006. Constructions at Work: The Nature of Generalization in Language. Oxford: Oxford University Press.
Goldberg, A.E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, 22, 131-153.
Green, G.M. (1974). Semantics and Syntactic Regularity. Bloomington, IN: Indiana University Press.
Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, 203-257.
Gropen, J., Steven, P., Hollander, M., & Goldberg, R. (1991). Syntax and Semantics in the Aquisition of Locative Verbs. *Journal of Child Language*, *18*(1), 115–151.
Hovav, M.R. & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of Linguistics,* 44(1), 129.
Jaeger, T. Florian, and Neal E. Snider. "Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience." *Cognition* 127.1 (2013): 57-83.
Johnson, M. A., & Goldberg, A. E. (2013). Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes*, *28*(10), 1439-1452.
Johnson, Matt A., Nicholas B. Turk-Browne, and Adele E. Goldberg. "Prediction plays a key role in language development as well as processing." *Behavioral and Brain Sciences* 36.04 (2013): 360-361.
Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133-156.
Kiparsky, P. (1982). Lexical morphology and phonology. In S. Yang (Ed), *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, 3-91. Seoul: Hanshin.
Lakoff, G. (1970). *Irregularity in syntax*. Holt, Rinehart, and Winston.
Levin, B. (1993). English verb classes and alternations: a preliminary investigation. Chicago, IL: University of Chicago Press.
Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, *122*(1), 12–36.
Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193-198.
Marcotte, J. (2005). Causative alternation errors as event-driven construction paradigm completions. In E.V. Clark & B.F. Kelly (Eds), *Constructions in Acquisition*. Stanford, CA: Center for the Study of Language and Information.
Marcus, G.F. (1993). Negative evidence in language acquisition. *Cognition*, 46, 53-85.
Oehrle, R.T. (1976). *The grammatical status of the English dative alternation* (Doctoral dissertation, Massachusetts Institute of Technology).
Pinker, S. (1981). Comments. In C.L. Baker & J.J. McCarthy (Eds), *The Logical Problem of Language Acquisition*, 53-63. Cambridge, MA: MIT Press
Pinker, S. (1989). Learnability and cognition. Cambridge, MA: MIT Press.

Rainer, F. (1988). Towards a theory of blocking: The case of Italian and German quality nouns. In Geert E. Booij and Jaap Van Marle (Eds), *Yearbook of Morphology 1*, 155-185. Dordrecht: Foris.

Robenalt, C. and Goldberg, A.E. to appear. L2 learners do not take competing alternative expressions into account the way L1 learners do. *Language Learning*.

Saxton, M., Kulcsar, B., Marshall, G., & Rupra, M. (1998). Longer-term effects of corrective input: an experimental approach. *Journal of Child Language*, 25, 701-721.

Stephens, Greg J., Lauren J. Silbert, and Uri Hasson. "Speaker–listener neural coupling underlies successful communication." *Proceedings of the National Academy of Sciences* 107.32 (2010): 14425-14430.

Theakston, A.L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development*, 19, 15-34.

Thompson, S.A. (1995). The iconicity of "dative shift" in English: Considerations from information flow in discourse. In M.E. Landsberg (Ed), *Syntactic Iconicity and Linguistic Freezes: The Human Dimension*, 155-75. Berlin: Mouton de Gruyter.

Tomasello, M. (2003). Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press.

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

Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, *22*(1), 28-32.

Online Appendix

Online Appendix Table 1: Results From COCA Searches to Identify Novel Uses for Experiment 1

Sentence	Search String	Novel Tokens	All Verb Tokens	Novel Tokens/ All Uses
No Competing Alternative				
Jeff slept the afternoon away.	[sleep].[v*] the [nn*] away	14	39657	0.0004
Jeff napped the afternoon away.	[nap].[v*] the [nn*] away	1	927	0.0011
Megan smiled her boyfriend out the front door.	[smile].[v*] [p*] [i*] [smile].[v*] [a*] [nn*] [i*]	4	46336	0.0001
Megan grinned her boyfriend out the front door.	[grin].[v*] [p*] [i*] [grin].[v*] [a*] [nn*] [i*]	1	8233	0.0001
Terry's horse jumped her straight out of the saddle.	[jump].[v*] [p*] [i*] [jump].[v*] [a*] [nn*] [i*]	17	32456	0.0005
Terry's horse hopped her straight out of the saddle.	[hop].[v*] [p*] [i*] [hop].[v*] [a*] [nn*] [i*]	3	4728	0.0006
The chief will laugh you back to your desk job.	[laugh].[v*] [p*] [i*] [laugh].[v*] [a*] [nn*] [i*]	30	54646	0.0005
The chief will chuckle you back to your desk job.	[chuckle].[v*] [p*] [i*] [chuckle].[v*] [a*] [nn*] [i*]	0	4090	0.0000
The lifeguard swam the children to shore.	[swim].[v*] [p*] [i*] [swim].[v*] [a*] [nn*] [i*]	10	13019	0.0008
The lifeguard backstroked the children to shore.	[backstroke].[v*] [p*] [i*] [backstroke].[v*] [a*] [nn*] [i*]	0	76	0.0000
The performer sang the audience into another dimension.	[sing].[v*] [p*] [i*] [sing].[v*] [a*] [nn*] [i*]	27	47468	0.0006
The performer crooned the audience into another dimension.	[croon].[v*] [p*] [i*] [croon].[v*] [a*] [nn*] [i*]	0	710	0.0000
The shopkeeper shouted the teenagers out of the building.	[shout].[v*] [p*] [i*] [shout].[v*] [a*] [nn*] [i*]	9	19526	0.0005

The shopkeeper hollered the teenagers out of the building.	[holler].[v*] [p*] [i*] [holler].[v*] [a*] [nn*] [i*]	0	1678	0.0000
Has Competing Alternative				
Amber explained Zach the answer.	[explain].[v*] [p*] [a*] [n*] [explain].[v*] [a*] [nn*] [a*] [nn*]	3	91780	0.0000
Amber recited Zach the answer.	[recite].[v*] [p*] [a*] [nn*] [recite].[v*] [a*] [nn*] [a*] [nn*]	1	3647	0.0003
Anthony's merciless teasing cried his little sister.	[cry].[v*] [p*] [cry].[v*] [a*] [nn*]	0	33240	0.0000
Anthony's merciless teasing sobbed his little sister.	[sob].[v*] [p*] [sob].[v*] [a*] [nn*]	0	3181	0.0000
Daniel forced that Helen compete.	[force].[v*] that [p*] [v*] [force].[v*] that [a*] [nn*] [v*]	2	51051	0.0000
Daniel coerced that Helen compete.	[coerce].[v*] that [p*] [v*] [coerce].[v*] that [a*] [nn*] [v*]	0	1318	0.0000
Jacob fell the lamp.	[fall].[v*] [p*] -[v*] [fall].[v*] [a*] [nn*] -[v*]	2	110954	0.0000
Jacob tumbled the lamp.	[tumble].[v*] [p*] -[v*] [tumble].[v*] [a*] [nn*] -[v*]	34	4884	0.0070
Kayla's boss considered to give her a raise.	-[v*] [consider].[v*] to -be.[v*]	3	117036	0.0000
Kayla's boss contemplated to give her a raise.	-[v*] [contemplate].[v*] to [v*]	1	5643	0.0002
Please find a new pen to me.	[find].[v*] [p*] to [p*] [find].[v*] [p*] to [a*] [nn*] [find].[v*] [a*] [nn*] to [p*] [find].[v*] [a*] [nn*] to [a*] [nn*]	1	451285	0.0000
Please locate a new pen to me.	[locate].[v*] [p*] to [p*] [locate].[v*] [p*] to [a*] [nn*] [locate].[v*] [a*] [nn*] to [p*] [locate].[v*] [a*] [nn*] to [a*] [nn*]	0	24728	0.0000
The magician disappeared the rabbit.	[disappear].[v*] [p*] -[v*] [disappear].[v*] [a*] [nn*] -[v*]	26	26464	0.0010
The magician vanished the rabbit.	[vanish].[v*] [p*] -[v*] [vanish].[v*] [a*] [nn*] -[v*]	8	8256	0.0010

Online Appendix Table 2: List of Stimuli for Judgment Study in Experiment 1 (with Paraphrase Ratings)

	Exact Repetition	Max. Paraphrase Repetition
Novel-noCA		
Jeff slept/napped the afternoon away.	14	2
Megan smiled/grinned her boyfriend out the front door.	1	2
Terry's horse jumped/hopped her straight out of the saddle.	3	3
The chief will laugh/chuckle you back to your desk job.	11	1
The lifeguard swam/backstroked the children to shore.	13	4
The performer sang/crooned the audience into another dimension.	10	1
The shopkeeper shouted/hollered the teenagers out of the building.	5	4
Novel-hasCA		
Amber explained/recited Zach the answer.	1	16
Anthony's merciless teasing cried/sobbed his little sister.	0	13
Daniel forced/coerced that Helen compete.	1	16
Jacob fell/tumbled the lamp.	0	10
Kayla's boss considered/contemplated to give her a raise.	4	14
Please find/locate a new pen to me.	0	12
The magician disappeared/vanished the rabbit.	0	18

Online Appendix Table 2: All items used for the acceptability judgments in Experiment 1. Two paraphrase measures are shown for each item: the number of times a sentence was accepted without any revisions and the number of times the most frequent paraphrase for each sentence was repeated. The maximum value for both measures is 20.

Online Appendix Table 3: Sample Paraphrase Responses and Ratings

Sentence Category	Rating	Repetitions
Has Competing Alternative		
The magician disappeared the rabbit.		
The magician made the rabbit disappear.	P	18
The rabbit disappeared after the magician did a trick.	P	1
The rabbit disappeared via the magician.	P	1
No Competing Alternative		
The chief will laugh you back to your desk job.		
The chief will laugh you back to your desk job.	E	11
The chief laughed after you ended up back in your desk job.	P	1
The chief laughed as you went back to your desk job.	P	1
The chief will laugh you back to your desk.	P	1
The chief will send you back to your desk job.	P	1
The chief will take you back to your desk job, laughing.	P	1
You will get laughed back to your desk job by the chief.	P	1
The chief will laugh at your inability to be anything but a lowly peon and order you back to your desk job.	P	1
The chief will laugh at you behind your back.	O	1
The chief will laugh back at your desk job.	O	1

Online Appendix Table 3: Sample paraphrases responses for two items. In the Rating column, P stands for paraphrase, E stands for exact repetition, and O stands for Other. The number of times each response was repeated is shown in the rightmost column.

Online Appendix Table 4: List of Stimuli for Plausibility Norming from Experiment 1

Sentence Category	Sentence	Plausibility (High Freq.)	Plausibility (Low Freq.)
Baseline		**4.71 (0.67)**	**4.75 (0.62)**
	Will slept/napped on the sofa.	5.0 (0.00)	4.9 (0.32)
	Laurie smiled/grinned.	5.0 (0.00)	4.6 (0.97)
	The boys jumped/hopped on the trampoline.	4.8 (0.42)	4.9 (0.32)
	The students laughed/chuckled.	5.0 (0.00)	4.8 (0.42)
	Alex swam/backstroked to the dock.	4.8 (0.42)	4.9 (0.32)
	Taylor sang/crooned a lullaby to the baby.	4.6 (0.97)	4.5 (0.85)
	The coach shouted/hollered at the players.	4.7 (0.67)	5.0 (0.00)
	The professor explained/recited the assignment.	5.0 (0.00)	4.8 (0.42)
	Christina cried/sobbed when her hamster died.	4.8 (0.63)	5.0 (0.00)
	The spy forced/coerced the criminal to confess.	4.6 (0.70)	4.5 (1.08)
	Brandon fell/tumbled out of the tree.	5.0 (0.00)	4.5 (1.08)
	The family considered/contemplated going to Disneyland.	5.0 (0.00)	4.8 (0.42)
	Emily found/located the book she needed.	5.0 (0.00)	5.0 (0.00)
	Ashley disappeared/vanished into the darkness.	4.1 (0.74)	4.8 (0.42)
Novel-noCA		**4.06 (1.34)**	**3.70 (1.39)**
	Jeff slept/napped the afternoon away.	4.3 (1.25)	4.8 (0.42)
	Megan smiled at her boyfriend as he went out the front door/ Megan grinned as her boyfriend went out the front door.	4.9 (0.32)	4.9 (0.32)
	Terry's horse jumped so high that it threw her out of the saddle./ Terry's horse hopped her straight out of the saddle.	4.6 (0.70)	3.3 (1.16)
	The chief will laugh/chuckle you back to your desk job.	3.1 (1.10)	2.2 (1.23)
	The lifeguard swam/backstroked the children to shore.	4.3 (1.34)	3.8 (1.32)
	The performer sang/crooned the audience into another dimension.	1.7 (1.34)	2.6 (1.43)
	The shopkeeper shouted/hollered the teenagers out of the building.	4.1 (0.57)	3.7 (1.06)
Novel-hasCA	(most common paraphrased used)	**4.70 (0.68)**	**4.54 (0.83)**
	Amber explained/recited the answer to Zach.	5.0 (0.00)	4.8 (0.42)
	Anthony's merciless teasing made his little sister cry/sob.	4.8 (0.42)	4.7 (0.67)
	Daniel forced/coerced Helen to compete.	4.8 (0.42)	4.8 (0.42)

Jacob made the lamp fall/tumble over.	4.8 (0.42)	4.5 (0.85)
Kayla's boss considered/contemplated giving her a raise.	4.7 (0.67)	4.5 (1.27)
Please find me a new pen./ Please locate a new pen for me.	4.7 (0.67)	4.5 (0.53)
The magician made the rabbit disappear/vanish.	4.6 (0.70)	3.7 (1.06)

Online Appendix Table 4: Items used in Experiment 1 to collect plausibility ratings, along with the mean and standard deviation for each sentence's rating. The most common response from the paraphrase norming was substituted for each novel sentence. The high and low frequency verbs used for each sentence are underlined.

Online Appendix Table 5: Results From COCA Searches to Identify Novel Uses for Experiment 2

Sentence	Search String	Novel Tokens	All Verb Tokens	Novel Tokens/ All Uses
No Competing Alternative				
The sound rattled the bats out of their hiding place.	[rattle].[v*] [a*] [nn*] out [rattle].[v*] [nn*] out [rattle].[v*] [pp*] out	3	4995	0.0006
The sound reverberated the bats out of their hiding place.	[reverberate].[v*] [a*] [nn*] out [reverberate].[v*] [nn*] out [reverberate].[v*] [pp*] out	0	1104	0.0000
The lifeguard swam a pool toy to the kids.	[swim].[v*] [a*] [nn*] to [swim].[v*] [nn*] to [swim].[v*] [pp*] to	4	13019	0.0003
The lifeguard paddled a pool toy to the kids.	[paddle].[v*] [a*] [nn*] to [paddle].[v*] [nn*] to [paddle].[v*] [pp*] to	0	2066	0.0000
The teacher frowned a warning to the back of the class.	[frown].[v*] [a*] [nn*] to [frown].[v*] [nn*] to [frown].[v*] [pp*] to	0	6875	0.0000
The teacher glowered a warning to the back of the class.	[glower].[v*] [a*] [nn*] to [glower].[v*] [nn*] to [glower].[v*] [pp*] to	0	516	0.0000
The crowd cheered the reluctant candidate to the podium.	[cheer].[v*] [a*] [nn*] to [cheer].[v*] [nn*] to [cheer].[v*] [pp*] to	7	6470	0.0011
The crowd hollered the reluctant candidate to the podium.	[holler].[v*] [a*] [nn*] to [holler].[v*] [nn*] to [holler].[v*] [pp*] to	3	1678	0.0018
The editorial embarrassed the poor man out of town.	[embarrass].[v*] [a*] [nn*] out [embarrass].[v*] [nn*] out [embarrass].[v*] [pp*] out	1	2989	0.0003
The editorial mortified the poor man out of town.	[mortify].[v*] [a*] [nn*] out [mortify].[v*] [nn*] out [mortify].[v*] [pp*] out	0	526	0.0000
The lion roared the veterinarian out of the enclosure.	[roar].[v*] [a*] [nn*] out [roar].[v*] [nn*] out [roar].[v*] [pp*] out	0	5253	0.0000
The lion snarled the veterinarian out of the enclosure.	[snarl].[v*] [a*] [nn*] out [snarl].[v*] [nn*] out [snarl].[v*] [pp*] out	0	1238	0.0000
The editor smiled the new reporter into his office.	[smile].[v*] [a*] [nn*] into [smile].[v*] [nn*] into [smile].[v*] [pp*] into	1	46336	0.00002

The editor grinned the new reporter into his office.	[grin].[v*] [a*] [nn*] into [grin].[v*] [nn*] into [grin].[v*] [pp*] into	0	8233	0.0000
The woman screamed the children out of the ice cream store.	[scream].[v*] [a*] [nn*] out [scream].[v*] [nn*] out [scream].[v*] [pp*] out	1	19642	0.00005
The woman shrieked the children out of the ice cream store.	[shriek].[v*] [a*] [nn*] out [shriek].[v*] [nn*] out [shriek].[v*] [pp*] out	0	2083	0.0000
The magician fascinated the toddlers into a trance.	[fascinate].[v*] [a*] [nn*] into [fascinate].[v*] [nn*] into [fascinate].[v*] [pp*] into	0	4134	0.0000
The magician enthralled the toddlers into a trance.	[enthrall].[v*] [a*] [nn*] into [enthrall].[v*] [nn*] into [enthrall].[v*] [pp*] into	0	664	0.0000
Andrew insulted the potential member out of the club.	[insult].[v*] [a*] [nn*] out [insult].[v*] [nn*] out [insult].[v*] [pp*] out	0	2602	0.0000
Andrew derided the potential member out of the club.	[deride].[v*] [a*] [nn*] out [deride].[v*] [nn*] out [deride].[v*] [pp*] out	0	842	0.0000
Has Competing Alternative				
The scientist infected bacteria into the sample.	[infect].[v*] [a*] [nn*] into [infect].[v*] [nn*] into [infect].[v*] [pp*] into	0	5136	0.0000
The scientist corrupted bacteria into the sample.	[corrupt].[v*] [a*] [nn*] into [corrupt].[v*] [nn*] into [corrupt].[v*] [pp*] into	0	1114	0.0000
The children soiled mud onto the carpet.	[soil].[v*] [a*] [nn*] onto [soil].[v*] [nn*] onto [soil].[v*] [pp*] onto	0	1160	0.0000
The children splotched mud onto the carpet.	[splotch].[v*] [a*] [nn*] onto [splotch].[v*] [nn*] onto [splotch].[v*] [pp*] onto	0	74	0.0000
The designer decorated lace onto the invitation.	[decorate].[v*] [a*] [nn*] onto [decorate].[v*] [nn*] onto [decorate].[v*] [pp*] onto	0	6829	0.0000
The designer embellished lace onto the invitation.	[embellish].[v*] [a*] [nn*] onto [embellish].[v*] [nn*] onto [embellish].[v*] [pp*] onto	0	1067	0.0000
The dictator flooded propaganda into the city.	[flood].[v*] [a*] [nn*] into [flood].[v*] [nn*] into [flood].[v*] [pp*] into	4	6426	0.0006
The dictator inundated propaganda into the city.	[inundate].[v*] [a*] [nn*] into [inundate].[v*] [nn*] into [inundate].[v*] [pp*] into	0	1019	0.0000
The chef coated ranch dressing over the salad.	[coat].[v*] [a*] [nn*] over [coat].[v*] [nn*] over [coat].[v*] [pp*] over	0	4601	0.0000

The chef doused ranch dressing over the salad.	[douse].[v*] [a*] [nn*] over [douse].[v*] [nn*] over [douse].[v*] [pp*] over	0	934	0.0000
The housekeeper soaked bleach into the towel.	[soak].[v*] [a*] [nn*] into [soak].[v*] [nn*] into [soak].[v*] [pp*] into	6	6125	0.0010
The housekeeper drenched bleach into the towel.	[drench].[v*] [a*] [nn*] into [drench].[v*] [nn*] into [drench].[v*] [pp*] into	0	1072	0.0000
The nurse bound cotton over the wound.	[bind].[v*] [a*] [nn*] over [bind].[v*] [nn*] over [bind].[v*] [pp*] over	1	14745	0.00006
The nurse bandaged cotton over the wound.	[bandage].[v*] [a*] [nn*] over [bandage].[v*] [nn*] over [bandage].[v*] [pp*] over	0	530	0.0000
Natalie smacked a newspaper onto the mosquito.	[smack].[v*] [a*] [nn*] onto [smack].[v*] [nn*] onto [smack].[v*] [pp*] onto	2	3082	0.0006
Natalie swatted a newspaper onto the mosquito.	[swat].[v*] [a*] [nn*] onto [swat].[v*] [nn*] onto [swat].[v*] [pp*] onto	0	1049	0.0000
The landscaper surrounded rocks around the garden.	[surround].[v*] [a*] [nn*] around [surround].[v*] [nn*] around [surround].[v*] [pp*] around	1	25512	0.00004
The landscaper bordered rocks around the garden.	[border].[v*] [a*] [nn*] around [border].[v*] [nn*] around [border].[v*] [pp*] around	0	2464	0.0000
The camper blocked a heavy backpack into the entrance.	[block].[v*] [a*] [nn*] into [block].[v*] [nn*] into [block].[v*] [pp*] into	1	15831	0.00006
The camper obstructed a heavy backpack into the entrance.	[obstruct].[v*] [a*] [nn*] into [obstruct].[v*] [nn*] into [obstruct].[v*] [pp*] into	0	1510	0.0000

Online Appendix Table 6: List of Items from Experiment 2 (with results from norming studies)

Sentence	Acceptability (max=7)	Plausibility (max=7)	Manner (max=5)	Force (max=5)	End-State (max=5)
No Competing Alternative					
The sound rattled the bats out of their hiding place.	6.64	6.91	4.50	2.83	2.50
The sound reverberated the bats out of their hiding place.	4.00	6.38	2.00	2.00	2.50
The lifeguard swam a pool toy to the kids.	4.25	6.25	4.67	4.50	1.67
The lifeguard paddled a pool toy to the kids.	4.75	5.25	5.00	4.00	1.00
The teacher frowned a warning to the back of the class.	4.17	5.58	4.00	2.17	3.17
The teacher glowered a warning to the back of the class.	3.86	6.57	5.00	1.00	1.00
The crowd cheered the reluctant candidate to the podium.	6.13	6.25	3.40	3.40	2.20
The crowd hollered the reluctant candidate to the podium.	4.17	6.17	4.50	2.00	1.50
The editorial embarrassed the poor man out of town.	4.83	6.08	2.17	1.50	3.33
The editorial mortified the poor man out of town.	3.38	5.13	1.00	2.50	4.00
The lion roared the veterinarian out of the enclosure.	3.67	6.42	4.50	2.50	1.00
The lion snarled the veterinarian out of the enclosure.	3.36	6.45	4.83	3.00	3.00

The editor smiled the new reporter into his office.	1.91	5.91	3.50	1.00	3.00
The editor grinned the new reporter into his office.	1.83	5.08	3.83	2.33	3.83
The woman screamed the children out of the ice cream store.	3.00	5.92	4.50	2.00	2.50
The woman shrieked the children out of the ice cream store.	2.60	5.70	4.50	2.83	1.67
The magician fascinated the toddlers into a trance.	5.27	5.09	2.50	1.00	4.00
The magician enthralled the toddlers into a trance.	5.92	5.92	2.67	1.17	2.50
Andrew insulted the potential member out of the club.	5.00	6.58	2.50	2.50	3.50
Andrew derided the potential member out of the club.	4.89	5.56	2.25	2.75	3.00
Has Competing Alternative					
The scientist infected bacteria into the sample.	3.82	6.55	3.83	3.33	4.67
The scientist corrupted bacteria into the sample.	3.08	5.75	2.50	1.50	5.00
The children soiled mud onto the carpet.	3.25	6.83	2.67	2.17	3.83
The children splotched mud onto the carpet.	6.64	6.73	4.00	2.50	5.00
The designer decorated lace onto the invitation.	4.64	6.36	3.00	2.17	3.67
The designer embellished lace onto the invitation.	5.08	6.00	4.00	1.00	3.50

The dictator flooded propaganda into the city.	5.42	6.67	3.67	3.17	4.17
The dictator inundated propaganda into the city.	5.40	6.50	2.50	2.00	4.50
The chef coated ranch dressing over the salad.	4.91	6.82	4.33	2.50	4.67
The chef doused ranch dressing over the salad.	6.17	6.92	4.50	3.00	4.50
The housekeeper soaked bleach into the towel.	5.25	7.00	4.00	2.00	5.00
The housekeeper drenched bleach into the towel.	3.83	6.25	3.33	3.00	4.67
The nurse bound cotton over the wound.	6.00	6.92	4.50	4.50	5.00
The nurse bandaged cotton over the wound.	6.13	6.88	4.33	2.17	4.83
Natalie smacked a newspaper onto the mosquito.	4.88	6.63	4.00	4.00	3.00
Natalie swatted a newspaper onto the mosquito.	4.00	6.92	4.50	4.00	2.67
The landscaper surrounded rocks around the garden.	3.25	6.75	2.50	3.00	5.00
The landscaper bordered rocks around the garden.	4.88	6.63	2.83	1.67	3.33
The camper blocked a heavy backpack into the entrance.	2.75	5.75	3.50	3.00	5.00
The camper obstructed a heavy backpack into the entrance.	2.09	4.82	2.50	2.67	3.67