

LEARNING WHAT NOT TO SAY: THE ROLE OF STATISTICAL PREEMPTION AND CATEGORIZATION IN *A*-ADJECTIVE PRODUCTION

JEREMY K. BOYD

University of Illinois at Urbana-Champaign

ADELE E. GOLDBERG

Princeton University

A persistent mystery in language acquisition is how speakers are able to learn seemingly arbitrary distributional restrictions. This article investigates one such case: the fact that speakers resist using certain adjectives prenominally (e.g. ??*the asleep man*). Experiment 1 indicates that speakers tentatively generalize or CATEGORIZE the distributional restriction beyond their previous experience. Experiment 2 demonstrates that speakers are sensitive to STATISTICAL PREEMPTION—that is, speakers learn not to use a formulation if an alternative formulation with the same function is consistently witnessed. Moreover, they are able to generalize the restriction to apply to other members of the category as well. Finally, experiment 3 finds evidence that speakers DISCOUNT a pseudo-preemptive context, rationally ignoring it as uninformative.*

Keywords: language acquisition, statistical learning, reasoning, negative evidence, preemption, entrenchment

1. INTRODUCTION. How do speakers learn not to use certain semantically sensible and syntactically reasonable formulations? For instance, the examples in 1–3 are perfectly interpretable, and the syntactic constructions involved are all licensed in English, yet each of the examples is decidedly odd.

- (1) ??The magician vanished the woman.¹
- (2) ??She explained him the news.
- (3) ??She considered to go to the store.

Occasionally, sentences such as those in 1–3 do occur in large corpora, but they are clearly dispreferred compared to the nearly synonymous examples in 4–6 (see Goldberg 2011 for ways to quantify this dispreference).

- (4) The magician made the woman vanish.
- (5) She explained the news to him.
- (6) She considered going to the store.

Another case of a nonpredictable restriction involves certain adjectives such as *asleep*, which resist prenominal attributive position (7), despite the fact that the vast majority of English adjectives readily appear in this position (8).

- (7) ??the asleep boy
- (8) the sleepy/absurd/active/tall boy

We detail the sense in which the restriction on these adjectives is not predictable in §2. Subsequently, we present a series of experiments that address the question of how such restrictions are learned.

1.1. THE (UN)RELIABILITY OF DIRECT NEGATIVE EVIDENCE. Clearly, the language that people hear does not come overtly marked with question marks or asterisks to indicate

* We thank Harald Baayen, Greg Carlson, and two anonymous referees for their very helpful comments. We are also grateful to lab groups or departments at several universities—Princeton, the University of Illinois, Harvard, MIT, and the University of Colorado—and at several meetings—in Vigo, Kiel, Kyoto, Paris, Os-nabrück, and at Brown—for helpful feedback. Useful analysis advice was provided by Harald Baayen, Scott Fraundorf, Tuan Lam, Gary Oppenheim, and Carolyn Anderson. This research was supported by an NICHD postdoctoral fellowship awarded to the first author via the University of Illinois Department of Psychology (2T32HD055272-11A2), and an NSF grant to the second author (0613227).

¹ We use ‘??’ instead of the traditional ‘*’ to indicate that judgments of ill-formedness are gradient and depend on many interacting factors.

unacceptability. Moreover, we know that speakers are not reliably overtly corrected for producing ill-formed utterances. That is, it is the rare parent who would utter anything like 9.

(9) Don't say 'the asleep boy', dear, say 'the boy that is asleep'.

Parents, and caregivers more generally, are much more interested in the content of children's utterances than the form. If a child utters the ungrammatical sentence in 10, for example, most mothers are likely to give the child a hug without correcting her grammar.

(10) Me loves you, Mommy.

This is not to say that explicit corrections never happen. There is a famous case of a caregiver who corrected her child's utterance ('Nobody don't love me') eight times without responding to the content of the utterance (cited in McNeill 1966:69), but this is (mercifully) uncommon (Brown & Hanlon 1970, Braine 1971, Baker 1979, Bowerman 1988, Pinker 1989, Marcus 1993). We return to the issue of caregiver RECASTS or reformulations in §1.6.

1.2. UNDERLYING FEATURES OR STRUCTURES. It is clear that appeals to unlearned constraints cannot account for examples such as 1–3 or 7, since the restrictions are lexically specific and do not follow from language-wide, let alone universal, constraints. That is, children have to learn that *vanish*, unlike *break*, cannot be used transitively as well as intransitively; that *explain*, unlike *tell*, cannot be used ditransitively as well as datively; that *consider*, unlike *want*, does not allow an infinitival VP complement; and that *asleep*, unlike *sleepy* or *acute*, disprefers appearing in attributive position. Even if we were to allow that each restricted lexical item might differ in having or lacking an underlying marker or some sort of underlying structure that distinguishes it from otherwise similar words, children would still be faced with the task of LEARNING which words have the invisible marker or structure and which do not (cf. also Pinker 1989). Thus, we would clearly be left with a version of our original question: how do children learn which words have the requisite underlying markers or structures?

1.3. ENTRENCHMENT. It is tempting to believe that the simple nonoccurrence of a given form is sufficient to render it unacceptable. In fact, several theorists have argued that the process of hearing a verb with sufficient token frequency in one argument-structure construction plays a key role in preventing it from appearing in other argument structures unless it has been witnessed in the other structures (e.g. Braine & Brooks 1995). The idea is basically that speakers only use words (particularly verbs) that they have heard used in particular constructions. This is the thrust of Braine and Brooks's 'unique argument-structure preference': once an argument-structure pattern has been learned for a particular verb, that argument-structure pattern tends to block the creative use of the verb in any other argument-structure pattern.

The idea that token frequency plays a role in constraining overgeneralization is supported by a finding by Brooks and colleagues (1999; see also Theakston 2004, Ambridge et al. 2008). Children are more likely to accept overgeneralizations of infrequent verbs (e.g. *vanish* used transitively) than overgeneralizations of frequent verbs (e.g. *disappear* used transitively). The suggestion has been that this is due to the fact that *disappear* has been heard in the simple intransitive construction much more often than *vanish*, and that since it is more entrenched intransitively, it is more difficult to creatively causativize it.

Unfortunately, the entrenchment proposal faces a problem in that it predicts that words that are used with high frequency in one construction should not be available for creative uses in other constructions. This prediction is contradicted by examples such as those in 11–14.

- (11) [She] prayed her way through the incomprehension of her atheist friends. (COCA)²
- (12) The python coughed her back out ... (www.rabbit.org/journal/3-7/snake-bite.html)
- (13) Aladar [a dinosaur] swam his friends to the mainland. (Zohfeld 2000)
- (14) He's right here at my feet, snoring his head off. (COCA)

Each of these verbs, *pray*, *cough*, *swim*, and *snore*, is very frequent (entrenched) in the intransitive construction, and is only exceedingly rarely, if ever, witnessed in the various transitive constructions in 11–14. As a result, an account based on entrenchment would inappropriately predict 11–14 to be unacceptable (Goldberg 2006).

1.4. NEGATIVE ENTRENCHMENT. Stefanowitsch (2008) argues for a slightly more subtle notion of entrenchment—'negative entrenchment'. He claims that learners can infer indirect negative evidence by calculating the expected likelihood of a given verb appearing in a given construction, based on the construction's and the verb's overall frequencies in the language, and comparing the expected frequency with the actual frequency with which a verb has been witnessed in a given construction. If a verb is expected to occur in a construction significantly more often than it actually occurs, learners could infer that it must be dispreferred in the construction. For example, he offers a proposal for how speakers might come to know that *say* is unacceptable in the ditransitive construction, as illustrated in 15.

- (15) ??Dad said Sue something nice. (cf. Dad told Sue something nice).

Stefanowitsch observes that the ditransitive construction occurs 1,824 times in his corpus, the verb *say* occurs 3,333 times, and that the total number of verbs in the corpus is 136,551. If the distribution of *say* across construction types were unrestricted, its expected token frequency in the ditransitive would be $1,824 \times 3,333 / 136,551$, or 44.52. But since its actual frequency in the ditransitive is zero, learners could use the discrepancy between the expected and observed values to infer that *say* cannot be used in the ditransitive.

Unfortunately, this account—exactly like the simple entrenchment account—predicts that highly frequent verbs should not be available for novel uses, particularly in constructions that are themselves fairly frequent. Note in the above calculation that as the frequencies of the verb and construction that one is attempting to test increase (the two terms in the numerator), one's estimate of the expected frequency of the verb in the construction also increases. So as either or both of these numbers go up, learners are increasingly likely to expect the verb and construction to cooccur. The absence of cooccurrence in the actual corpus, however, would mistakenly signal to learners that the combination is unacceptable. We have already seen that this prediction is falsified by examples like those in 11–14: highly frequent verbs ARE creatively used in constructions in which they normally do not appear, and such uses result in well-formed utterances.

In addition, speakers would not be able to use entrenchment, whether positive or negative, to learn that a low-frequency verb could NOT appear in a low-frequency construction. In this case the expected frequency would necessarily be low, and in many cases may not be significantly different from zero. That is, the learner may never expect to hear the verb in the construction. This would seem to predict that low-frequency constructions should be productively available with low-frequency verbs across the board, since there is no evidence to lead learners to infer that such combinations are ill-formed. Some low-frequency

² The Corpus of Contemporary American English (COCA) is a free, parsed, 400-million-word corpus of spoken and written texts made available online by Mark Davies; <http://www.americancorpus.org/>.

constructions are in fact quite productive (e.g. the *way* construction, Goldberg 1995), but others are markedly less so. Novel words are of course especially low frequency, since by definition they have not been heard before by native speakers. And yet we see below that there is some evidence that speakers understand restrictions on their distribution as well; we return to this point in the discussion of experiments 1 and 2 below.

Low-frequency constructions sometimes disallow low-frequency or novel verbs for semantic reasons (Zwicky 1971). For example, consider the relatively rare *up and V* construction, illustrated in 16 (a and b are from the COCA corpus).

- (16) a. ... Randy Travis sings the zillionth song about how his woman up and left him ...
 b. The science teacher up and quit.
 c. ??He up and sauntered across the room.

The *up and V* construction is unusual syntactically in that it conjoins a particle and a verb. It also has particular semantic restrictions, in that it is used to convey quick or unexpected actions. It is for this reason that *saunter* is not natural in the construction (16c), since *saunter* designates a relaxed and slow manner of motion. Here the notion of FIT between the verb and the construction clearly comes into play.

1.5. LEXICAL ITEMS MUST FIT OR ACCOMMODATE THE CONSTRAINTS OF CONSTRUCTION. Clearly, new uses of verbs must fit, or be able to ACCOMMODATE, the constraints of the constructions they appear in (Gropen et al. 1989, Pinker 1989, Gropen et al. 1991, Goldberg 1995, Coppock 2008, Ambridge et al. 2009). One way to view this is to recognize that speakers categorize instances of each construction, and thereby form generalizations about semantic, pragmatic, and phonological constraints (Tomasello 2003). New expressions are judged to be ill-formed if they do not fit the semantic and pragmatic properties of the constructions.

This idea has been offered as the singular explanation for how children retreat from overgeneralizations. Ambridge and colleagues (2009:1301), for example, state that ‘children acquire—in an incremental and probabilistic fashion—the meaning of particular constructions and particular verbs, rejecting generalizations where the incompatibility between the two is too great’. We agree with the fact that semantic, pragmatic, and phonological compatibility are necessary requirements for words to be used in constructions, and that children are often learning the meanings of words and constructions probabilistically and incrementally. However, while motivation for restrictions is often available, we need to keep in mind that this type of explanation does not predict the ill-formedness of the examples we set out to address in 1–3 or 7. This is evidenced by the fact that expressions that share closely related semantics, pragmatics, and phonological properties are fully acceptable.

- (17) The magician banished the woman. (cf. ??The magician vanished the woman.)
 (18) She told/guaranteed him the news. (cf. ??She explained him the news.)
 (19) the sleepy/absurd boy (cf. ??the asleep boy)

1.6. RECASTS OR REFORMULATIONS. While parents do not overtly correct children’s errors very often (recall §1.1), there is evidence that they do sometimes recast children’s utterances, implicitly correcting them when they are ill-formed. A reformulation holds the semantic and discourse function of the learner’s utterance constant and changes only the formal properties. When an utterance is reformulated, children can potentially directly compare what they said with what the adult said.

Evidence that children are exposed to morphological recasts has existed for some time (Hirsh-Pasek et al. 1984, Demetras et al. 1986, Moerk 1991, Saxton 1997). Moreover, at least within the domain of morphology, it seems that children are often quite responsive to recasts when they hear them (Saxton et al. 1988, Saxton 1997, Strapp et al. 2008), although this remains somewhat controversial (Morgan et al. 1995). Some studies have reported rates of self-correction of 10–50% after a recast, depending on the study and the child (Farrar 1992, Chouinard & Clark 2003).

The majority of research on recasts has been done in the domain of morphology, not syntax. This is perhaps for good reason, since discourse factors typically distinguish various constructions, and the discourse context typically shifts subtly from one moment to the next. For example, the following are perfectly normal hypothetical conversations. They both involve a shift from one construction to another, but they should NOT be construed as corrective recasts.

- (20) Child: Only boys who were tall made the team. (*tall* used predicatively)
 Adult: One day you'll be a tall boy. (*tall* used attributively)
- (21) Child: I gave the dog my sandwich. (*give* used ditransitively)
 Adult: You gave your sandwich to THE DOG!?! (*give* used datively)

What is needed for a recast to be successful is that the child recognize that the adult substituted a new formulation for her own formulation, not because of changing discourse demands, but because the new formulation was better suited to the original context. Recasts then can be viewed as a specific type of STATISTICAL PREEMPTION. Statistical preemption assumes that formulations that are witnessed compete with other formulations that might be expected to occur in a given context. Statistical preemption may be operative in situations far beyond those in which an adult reformulates a child's error immediately, since it requires only that the learner recognize that the adult's utterance differs from what might otherwise have been expected in the given context.

If recasts occurred with great regularity, one might question the need for the more general process of statistical preemption. And in fact, Chouinard and Clark (2003) found that reformulations were used in response to up to two-thirds of each of five two-year-olds' errors in phonology, morphology, OR syntax. One wonders, however, if this strikingly high percentage reflects what most parents do in most situations. Chouinard and Clark acknowledge that four out of five parents were college-educated and that this high rate of recasts might not be available in other social classes or in other cultures (Ochs & Schieffelin 1984).

We might also observe that the tape recording of the conversations may have led to an increase in recasts, in any one of several ways. The caregiver may have consciously or unconsciously aimed to help the transcriber determine what the child intended, even if the caregiver herself understood the child; or the caregiver may have been somewhat more doting in front of the experimenter, responding with more complete and informative utterances than when alone with the child. Finally, the caregivers, who were also the children's parents in three out of five cases, may have consciously or unconsciously wished to present the children in the best possible light by subtly encouraging the children to correct their errors. It is well worth investigating recasts further. At the same time, without a firm demonstration that recasts are routinely available to all children, it is worth exploring whether a more general process of statistical preemption can in fact help learners learn what not to say.

The following section reviews the notion of statistical preemption in more detail.

1.7. STATISTICAL PREEMPTION. A number of theorists have suggested that a process of statistical preemption plays a role in speakers' learning to avoid syntactic overgeneralizations (Pinker 1981, Bates & MacWhinney 1987, Clark 1987, Di Sciullo & Williams 1987, Goldberg 1993, 1995, 2006, Marcotte 2005, Foraker et al. 2007). Preemption is a particular type of indirect negative evidence that results from repeatedly hearing a formulation, B, in a context where one might have expected to hear a semantically and pragmatically related alternative formulation, A. Given this type of input, speakers implicitly recognize that B is the appropriate formulation in such a context, and that A is not appropriate.

Morphological preemption (or blocking) is already familiar from morphology: *went* preempts *goed*, and *feet* preempts *foots* (Aronoff 1976, Kiparsky 1982). But statistical preemption between two phrasal forms requires explanation, since expressions formed from distinct phrasal constructions are virtually never semantically and pragmatically identical, and therefore it is not clear that an instance of one phrasal pattern could preempt the use of another (Pinker 1989, Bowerman 1996). For example, the ditransitive construction is distinct, at least in terms of its information structure, from the dative (Green 1974, Erteschik-Shir 1979, Collins 1995, Goldberg 1995, 2006, Thompson 1995, Arnold et al. 2000, Wasow 2002, Bresnan et al. 2007). Thus, knowledge that the dative is licensed as in 22b should not in any simple way preempt the use of the ditransitive 22a. And in fact, a large number of verbs do freely appear in both constructions (e.g. *tell* as in 23a,b).

- (22) a. ??She explained someone the story.
 b. She explained the story to someone.
- (23) a. She told someone the story.
 b. She told the story to someone.

Nonetheless, preemption could play an important role in learning to avoid expressions such as 22a, once a speaker's expectations are taken into account in the following way. Learners may witness repeated situations in which the ditransitive might be expected because the relevant information structure suits the ditransitive at least as well as the prepositional paraphrase. If, in these situations, the prepositional alternative is systematically witnessed instead, the learner can implicitly infer that the ditransitive 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 of course be due to some subtle difference in the functions of the two formulations that favors formulation B. Or a single use may simply be due to an error by the speaker. But if an alternative formulation is consistently heard in relevant contexts, a process of statistical preemption predicts that speakers will learn to use the alternative.

Statistical preemption of phrasal forms has not received a great deal of attention in the experimental literature, with the exception of a few notable studies. It has been found that seeing novel intransitive verbs in periphrastic causative constructions significantly preempts children's use of them in 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 in a transitive construction 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. Thus, if a child had heard both *The cow is chamming* (intransitive) and *Ernie's making the cow cham* (periphrastic causative), they were less likely to respond to 'What did Ernie do to the cow?' with *Ernie*

chammed the cow (transitive) than 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 transitive, statistically preempting use of the latter (cf. also Tomasello 2003).

The preemptive process, unlike the notion of simple high token frequency, predicts that transitive expressions involving normally intransitive verbs (recall *pray*, *cough*, *swim*, and *snore* in 11–14) would not be preempted by intransitive uses because the expressions would not be used in the same contexts. For example, the meanings of causing a change of state (24a) and an involuntary intransitive action (24b) are not in competition.

- (24) a. And he sneezed the house in! (Robinette 1999)
 b. She sneezed.

As a result of their differing functions—and contra the predictions of an entrenchment account—witnessing many instances of *sneeze* in the construction type exemplified by 24b would not decrease the felicity of its use in 24a.

Like entrenchment, however, preemption takes token frequencies into account, but in a different way. Consider the sentence in 25.

- (25) ??She disappeared it.

According to the entrenchment account, this formulation is blocked because it is not witnessed in the input and because *disappear* is often used intransitively—that is, *It disappeared* is quite frequent. In contrast, preemption claims that 25 is not witnessed because it never occurs in the input, and because when a causative meaning is intended, *disappear* is highly frequent in the more complex periphrastic causative. It is thus the frequency of a verb in functionally comparable constructions (not in just any alternative construction) that predicts the strength of the negative evidence that is provided. Note that this fits nicely with the finding that high-frequency verbs are more resistant to overgeneralization (e.g. Brooks et al. 1999) when we note that an increase in the total token frequency of a verb is often accompanied by an increase in its frequency in a potentially preemptive construction. Thus, *disappear* is less acceptable than *vanish* in a transitive construction, not because it has higher token frequency in the intransitive, but because it has higher token frequency in a construction that has a comparable discourse function to the transitive—the periphrastic causative. In fact, a search of the COCA corpus confirms that *make NP disappear* is more frequent than *make NP vanish* by a factor of ten.

In this way, frequency plays a role in the process of statistical preemption because the strength of the negative evidence provided by witnessing a functionally equivalent alternative increases as more exemplars are seen. As learners are exposed again and again to one construction and not an alternative with a comparable function, they become more and more confident that only the first is conventional (see Goldberg 2011 for a proposal of how to quantify strength and confidence of statistical preemption). This presupposes that learners encode the frequencies of phrasal forms; there is in fact much evidence that they do (Bod 1998, Bannard & Matthews 2008, Arnon & Snider 2009, Gurevich et al. 2010, Tremblay & Baayen 2010).

The majority of the theoretical work on the learning of arbitrary restrictions has focused on argument-structure alternations, particularly the transitive construction. We aim to broaden the discussion with a case that shares the same basic property of being not wholly predictable on the basis of general semantic or syntactic factors: the case of *a*-adjectives.

2. *A*-ADJECTIVES. There is a class of English adjectives beginning with a syllabic schwa (written as *a-*) that resist appearing prenominally in attributive position. Examples include those in 26.

- (26) a. ??the/an afraid child
 b. ??the/an afloat ship
 c. ??the/an alive monster
 d. ??the/an ablaze building

The restriction is not a general semantic or general phonological prohibition, since near synonyms and words with closely related phonology readily appear prenominally. To get a sense of how the pattern is idiosyncratic, note first that the following near synonyms are fully acceptable prenominally.

- (27) a. the/a scared child
 b. the/a floating ship
 c. the/a living monster
 d. the/a burning building

Furthermore, the class of *a*-adjectives is defined beyond the initial syllabic schwa, insofar as each member is morphologically segmentable into *a-* plus a semantically related stem, as in the following examples.

- (28) a. afloat: *a* + float; cf. float
 b. alive: *a* + live; cf. live
 c. ablaze: *a* + blaze; cf. blaze
 d. afraid: *a* + fraid; cf. frighten, fraidy-cat

Other adjectives with similar phonology, but which are not segmentable, fall outside the *a*-adjective category (cf. also Coppock 2008). For example, *adult* is superficially similar to *a*-adjectives in that it consists of an initial syllabic schwa, but /dʌlt/ is not a recognizable English morpheme, much less one that is semantically related to *adult*. Along the same lines, while parsing *acute* as *a+cute* actually does result in a recognizable English morpheme—*cute*—this ‘stem’ has nothing to do with the meaning of *acute*. Similarly, the syllables /stʌt/ and /lʊf/ are clearly not morphemes in *astute* or *aloof*. Since these adjectives are not segmentable into *a-* plus a semantically related stem, they are not *a*-adjectives and are thus not subject to the prohibition on prenominal use.

- (29) a. the adult male
 b. the acute sickness
 c. the astute comment
 d. the aloof woman

Additionally, the adjectives in 29 provide further evidence against the hypothesis that phonology might somehow be solely responsible for *a*-adjectives’ dispreference for prenominal placement. Given that they are phonologically quite similar to those in 26, we can conclude that while *a*-adjectives are, in part, defined by the way that they sound, the category does not reduce to phonological characteristics.

The restriction against attributive use is motivated by the diachronic history of many *a*-adjectives as prepositional phrases (e.g. *asleep* comes from the Old English PP *on sleep*; Simpson & Weiner 1989). As prepositional phrases, it made sense that they would not occur in prenominal attributive position, since prepositional phrases do not occur prenominally (??*the on drugs man*). Today’s speakers, however, are generally unaware of the historical origin of these adjectives, and yet they implicitly recognize and respect their unusual distributional pattern. Since there is no general semantic or phonological

reason to treat the members of the class as anything other than adjectives, their unusual distribution poses a clear learnability challenge. How ever do speakers learn to avoid using *a*-adjectives preminally?

This article explores the roles that categorization and statistical preemption play in learning what not to say—that is, in constraining overgeneralization errors. The first question we address in experiment 1 is whether speakers are able to generalize the already recognized restriction to novel adjectives. In experiment 2, we investigate whether speakers make use of statistical preemption, and whether they can extend evidence gleaned from statistical preemption to other members of a category. Finally, in experiment 3, we determine whether speakers are able to rationally ignore pseudo-preemptive contexts.

3. MOTIVATION FOR THE EXPERIMENTAL STUDIES.

3.1. CATEGORIZATION (EXPERIMENTS 1 AND 2). We are constantly encountering new entities and situations that deviate in large and small ways from our previous experience. The next door we encounter may differ from previous doors in being larger or smaller, wooden or windowed, and may require pushing, pulling, or sliding to open. And yet we have no trouble recognizing this new door as a door. Rats recognize new foods arranged in new ways as food to eat; dogs recognize new dogs as dogs to mate with. Categorization is a process whereby related entities are treated alike for certain purposes. It is clearly ubiquitous in humans and throughout the animal kingdom.

We undoubtedly categorize linguistic elements as well. For example, we project that the past tense of a nonsense word *spling* might be *splang* on the basis of *spling*'s phonological similarity to *sing* and *ring* (Bybee & Moder 1983, Pinker 1999). What is less clear is whether people categorize potential RESTRICTIONS. We address this question by considering the distributional restriction on certain *a*-adjectives like *asleep*. Experiment 1 aims to investigate the extent to which the *a*-adjective category is productive. Do speakers avoid using NOVEL *a*-adjectives preminally? If novel *a*-adjectives are treated like familiar *a*-adjectives, it would indicate that speakers have formed a productive category of *a*-adjectives. If not, it would suggest that each items' restriction must be learned independently without reference to an implicit generalization.

3.2. STATISTICAL PREEMPTION AND CATEGORIZATION (EXPERIMENT 2). Experiment 2 investigates statistical preemption in the case of *a*-adjectives, and additionally asks whether the preemptive context is itself generalized. That is, if certain novel *a*-adjectives are heard used in a preemptive context, do speakers implicitly extend the restriction to other novel *a*-adjectives witnessed in the same experimental setting? We consider this question by exposing participants to two novel *a*-adjectives in relative clause structures as in 30, to determine whether this will preempt their usage in attributive structures like 31.

(30) The pig that was ablim moved to the star.

(31) The ablim pig moved to the star.

3.3. DISCOUNTING (EXPERIMENT 3). Recognizing a role for statistical preemption raises an important question. How savvy are speakers in determining whether a given utterance should qualify as a preemptive context? If the particular formulation used can be attributed to some other factor, do speakers readily DISCOUNT that context, essentially recognizing that the formulation is irrelevant to the potential alternative?

In experiment 3, we expose participants to a pseudo-preemptive context: that is, a context in which the novel *a*-adjective appears in an alternative construction (a relative clause), but use of the alternative is attributable to some other aspect of the utterance. In

particular, participants witnessed novel *a*-adjectives in relative clause constructions, but this time in the context of a conjoined complex modifier consisting of another adjective and its prepositional complement, as in 32. Note that complex adjective phrases such as *proud of itself* never appear prenominal, so a relative clause formulation is the only option. It does not depend on the identity of the other adjective in the conjunct (cf. 33 and 34). We are therefore asking whether speakers are sensitive to the fact that 32 represents a pseudo-preemptive context.

- (32) The fox that was ablim and proud of itself ...
 (33) ??The proud of itself fox ...
 (34) ??The red and proud of itself fox ...

3.4. EXPERIMENTAL OVERVIEW. We report three studies with adult native speakers of English that were designed to elicit adjectival productions in a naturalistic context. Participants were given the opportunity to use different adjectives either prenominal in an attributive construction, or postnominal in a relative clause construction. The basic protocol asked participants to view scenes in which one of two animals moved across a computer screen, then produce a description of what they saw. All animals were labeled onscreen with a word; on critical trials the words were always adjectives. This means, for example, that if a critical trial showed two cows (see Figure 1A)—one labeled with *vigilant* and the other labeled with *asleep*—and the cow with the *asleep* label moved toward an onscreen star, then participants should produce either an attributive construction like *The asleep cow moved to the star*, or a relative clause construction like *The cow that's asleep moved to the star*. Crucially, the choice of which construction to use was entirely up to the participants.

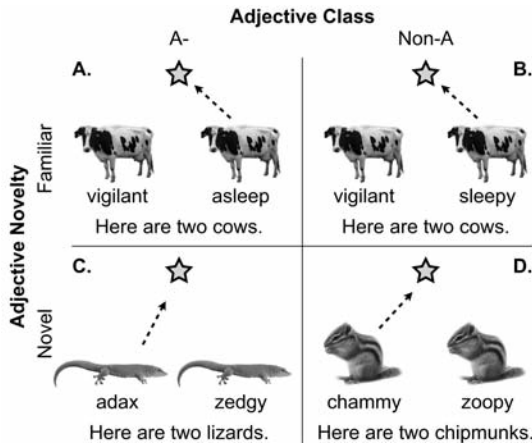


FIGURE 1. Critical trials involved describing scenes like those above, where the target label was either a familiar *a*-adjective (A), a familiar non-*a*-adjective (B), a novel *a*-adjective (C), or a novel non-*a*-adjective (D).

4. GENERAL METHODS. Each of the three experiments made use of similar materials and procedures and had the same structure. Participants began by watching the experimenter work through a number of trials in an EXPOSURE block. They then completed a series of trials themselves in a PRODUCTION block. The production block was identical across experiments; we manipulated the contents of the exposure block only, as a means of controlling participants' experience with novel *a*-adjectives like *ablim* and *adax*. In

experiment 1, the exposure block was designed to provide no experience with novel *a*-adjectives; speakers' behavior in the production block thus reflected their baseline production preferences. In experiment 2, the exposure block contained examples like 30 (*The pig that was ablim moved to the star*), with novel *a*-adjectives in relative clause structures. We hypothesized that speakers would treat these sorts of sentences as implicit negative evidence and would therefore reduce prenominal *a*-adjective uses later in the production block. In experiment 3, the exposure block contained sentences like 32 that were potentially pseudo-preemptive (e.g. *The fox that was ablim and proud of itself moved to the star*). If speakers rationally consider the nature of the evidence available to them, then these sentences should not reduce prenominal *a*-adjective use vis-à-vis the baseline from experiment 1.

4.1. MATERIALS. Figure 1 gives examples of four different critical trials from the production block used in experiments 1–3. Participants viewed slides in which one animal from a pair moved toward an onscreen star. The type of animal was identified in a sentence at the bottom of each slide, and the individuals in each pair were differentiated through the use of descriptive labels. For example, Fig. 1A is a slide that depicts an event involving two cows in which the cow associated with the label *vigilant* (the FOIL label) remains stationary, while the cow associated with the label *asleep* (the TARGET label) moves toward the star. After viewing an event of this sort, participants would read a centrally presented question on the following slide that prompted them to describe what they had just seen—for example, *Which cow moved to the star?*. Participants typically responded with the target label embedded in either an attributive structure or a relative clause structure.

4.2. CRITICAL TRIALS. The labels printed under animals were always adjectives in critical trials. We manipulated two characteristics of target adjectives by crossing adjective class (*a*- vs. non-*a*-) and adjective novelty (familiar vs. novel). This resulted in the four experimental conditions represented in the different panels of Fig. 1. We additionally controlled for semantic differences between familiar *a*- and non-*a*-adjectives by choosing non-*a*-adjectives that were near synonyms of the *a*-adjectives. *Asleep* (A) and *sleepy* (B) in Fig. 1 illustrate this relationship. There were four different target labels in each of the four conditions, resulting in a total of sixteen critical trials. Table 1 lists all target and foil labels used in critical trials.³ Note that since animals associated with foil labels never moved (see, for example, any of the animals in Fig. 1 with no arrows), participants never produced sentences containing foil labels. The sole purpose of the foils was to set up contrasts between animals that would make the use of attributive and relative clause descriptions discourse-appropriate. Additionally, for foil labels that occurred with novel *a*-adjectives, the foil always started with a *z*-, and was sometimes constructed to be highly similar to an existing adjective (e.g. *zunderful*). This was done as a means of keeping participants from paying particular attention to the phonology of the *a*-adjectives.

³ All novel non-*a*-adjectives ended in *-y* in order to facilitate the recognition of these novel labels as adjectives (rather than verbs), since *-y* is a common adjectival ending in English. This did not appear to influence the results, however, since subsequent work that included novel non-*a*-adjectives that did not end in *-y* (e.g. *bream*), as well as those that did, revealed no difference in rates of attributive usage based on the presence of a *-y* ending, $t(31) = 1.49, p = 0.15$.

FAMILIAR		NOVEL	
<i>a-</i>	NON- <i>a-</i>	<i>a-</i>	NON- <i>a-</i>
afloat (sinking)	floating (sinking)	ablim (zecky)	chammy (zoopy)
afraid (brave)	frightened (brave)	adax (zedgy)	flitzy (zappy)
alive (dead)	living (dead)	afraz (zibby)	gecky (zunderful)
asleep (vigilant)	sleepy (vigilant)	agask (zintesting)	slooky (zinky)

TABLE 1. Critical target adjective labels; foil labels are in parentheses.

4.3. FILLER TRIALS. Sixteen filler trials were interleaved with critical trials as a means of reducing the possibility that speakers might perseverate on the production of one construction type. Half of all fillers encouraged relative clause responses by using third-person singular verbs as target labels—for example, *smokes*, *runs*, *snowboards* (see Figure 2, left panel). Since participants were required to use EXACT labels in formulating their sentences (see §4.6), filler trials with verb labels invariably resulted in the production of relative clause structures. The remaining filler trials encouraged the production of prenominal attributive structures by using high-frequency adjectives that commonly occur in attributive position as target labels—for example, *fast*, *old*, *strong* (see Fig. 2, right panel).



FIGURE 2. Filler trials used either third-person singular verb labels to elicit a relative clause response (left panel), or high-frequency adjective labels to elicit an attributive response (right panel).

Filler trials thus served to encourage response variability and to guard against strategic responding. Given that half of the filler trials favored relative clause responses and the other half favored attributive responses, it was unlikely that a participant would be able to respond using one structure or the other throughout the course of the experiment. Instead, the presence of fillers increased the probability that participants would flexibly choose between the two structures on a trial-by-trial basis.

The type of filler that preceded a critical trial was counterbalanced across each condition to ensure that fillers did not unduly influence participants' responses—that is, half of all critical trials in any condition were preceded by fillers that were likely to elicit attributive responses, while the other half were preceded by fillers likely to elicit relative clause responses.

The production block used in experiments 1–3 was formed by interleaving sixteen filler trials with sixteen critical trials.

4.4. EXPOSURE TRIALS. Immediately prior to the production block, each participant watched while the experimenter worked through twelve exposure trials. Participants were told that the experimenter was simply demonstrating the task to them. These trials had the same structure as critical and filler trials—that is, one of two labeled animals performed an action—and were designed to give participants different kinds of exposure to *a*-adjectives. The nature of the exposure trials differed systematically across the three experiments as described below, but in each of the experiments, attributive struc-

tures were modeled in half of the trials, and relative clause structures were modeled in the other half.

4.5. SLIDESHOWS. All experimental materials were presented to participants as slideshows. Each slideshow consisted of a block of twelve exposure trials (the exposure block) followed by a block of sixteen critical trials plus sixteen fillers (the production block). As a means of guarding against order effects, participants were randomly assigned to view one of four slideshow versions in which critical and filler trials were presented in different pseudo-random orders; the order of the initial exposure trials were also varied pseudo-randomly across the versions. Exposure blocks had no more than two trials in a row in which *a*-adjectives were targets. Production blocks began with a filler trial, and alternated between critical and filler trials thereafter. Additionally, there were no more than two fillers of the same type (i.e. adjective labels vs. verb labels) in a row, there were no more than two critical trials from the same condition in a row, and there were no more than two trials featuring *a*-adjectives in a row. Each slideshow was balanced so that half of the target labels appeared on the left, and half on the right.

4.6. PROCEDURE. Each experiment started with the same instructions. Participants were briefed on the experimental procedure, and were told that in formulating their descriptions they must make use of the EXACT label associated with the target animal. This ensured that filler trials that were intended to result in relative clause responses actually did. For example, given the stimuli in Fig. 2, left panel, the most straightforward description that obeyed the exact-label constraint would be *The owl that smokes moved to the star*; changing the label was disallowed (e.g. *The smoking owl moved to the star*). The use of exact labels was emphasized in two practice trials with feedback: one in which a relative clause structure was produced, and one in which an attributive structure was produced. Note that the requirement that exact labels be used ensured that novel adjectives could only be used as adjectives. If for some reason participants had construed them as verbs, they would have had to add a verbal third-person singular inflection, as in *The cow that ablims moved to the star*. This never occurred.

Participants were also instructed that some trials might make use of words that they had never seen before. On these trials they were encouraged to avoid awkward utterances, and to instead formulate a description that sounded like something that a native speaker of English might say. Purportedly as a means of demonstrating this principle, the experimenter then worked through the exposure block while the participant observed. This involved the experimenter reading the labels on each slide aloud, then watching to see which animal moved, then formulating an utterance that described what happened.

At the end of the exposure block the experimenter asked participants if they had any questions about the procedure, then started the production block. As in the exposure block, the experimenter read the labels on each trial aloud. This ensured that participants would know the correct pronunciation for novel adjectives, which was especially important for novel *a*-adjectives, since class membership is in part contingent on phonological form (see §2). In contrast to the exposure block, however, participants formulated their own descriptions on each trial.

Each participant was debriefed at the end of their test session in order to identify any individuals who may have guessed the purpose of the experiment. Participants were asked whether they had noticed that certain types of adjectives in the experiment seemed to work well with some sentence types but not others. Two participants whose answers revealed explicit knowledge of the constraint against *a*-adjective use in attributive structures had their data excluded from consideration, as reported below.

Audio recordings were made of each session in order to facilitate later coding and analysis. Utterances in all critical trials were coded as either attributive responses (e.g. *The asleep cow moved to the star*), or relative clause responses (e.g. *The cow that's asleep moved to the star*). If participants initially responded with one structure but then changed to another—which happened on only three trials out of the 969 that were recorded (less than 1% of the time)—their final response was coded. There were no responses that did not fit clearly into one coding category or the other.

4.7. ANALYSIS. All statistical modeling and results graphics were done in the R computing language and environment (Wickham 2009, R Development Core Team 2010). We relied on logit mixed models as our primary modeling tool (Baayen 2008, Jaeger 2008, Quené & van den Bergh 2008). This type of model has become increasingly popular over the past several years for the analysis of linguistic data because it offers several advantages over traditional by-participants and by-items ANOVAs. Mixed models are robust to missing data, nonnormality, and heteroscedasticity, and allow for the combination of participant and item effects into a single unified analysis. They also avoid certain problems that stem from the transformation of categorical outcomes into proportions. In particular, it is possible for ANOVA models to generate predicted values that are uninterpretable because they are not bounded by zero and one. This can lead to potentially spurious interaction effects. For all mixed models appearing in the present work, model comparison starting with a fully saturated model was used to determine the reported set of parameters. That is, we began with an initial model that included all possible parameters, then removed them one at a time to identify models showing the best possible balance between fit and complexity.

Where indicated, we additionally performed follow-up tests using conditional inference trees (Strobl et al. 2009). Conditional inference is a nonparametric statistical method that originated in machine learning circles, and is especially well suited to investigating possible interaction patterns. It works by using correlations between predictor and response variables to recursively partition a dataset into homogeneous subsets. The result is a conditional inference tree, including *p*-value annotations indicating whether a correlation is significant. We explain the algorithm used to create conditional inference trees in more detail in §5.3.

5. EXPERIMENT 1: CATEGORIZATION. Experiment 1 investigated speakers' use of familiar and novel *a*-adjectives without prior exposure to any *a*-adjectives occurring in the exposure block. We asked two questions. First, do speakers in fact avoid using familiar *a*-adjectives (e.g. *asleep*) attributively? Intuitive judgments and corpus data (not reported here) suggest that they should; we therefore anticipated answering this first question in the affirmative. The second question—and the one that we were primarily interested in—asks whether speakers might also show a dispreference for using novel *a*-adjectives (e.g. *ablim*) attributively. If novel *a*-adjectives avoid attributive use in the same way that familiar *a*-adjectives are predicted to, then this would constitute evidence of generalization, which would suggest that speakers' grammars include an abstract *a*-adjective category that militates against the attributive use of ANY category member. If, by contrast, novel *a*-adjectives fail to pattern like familiar *a*-adjectives, then this would indicate that speakers' representations of *a*-adjectives are predominantly item-based. That is, it would suggest that the prohibition on attributive *a*-adjective use is learned on an adjective-by-adjective basis, and does not generalize to new items.

5.1. PARTICIPANTS. Twenty adult native speakers of English were recruited from the Princeton Department of Psychology subject pool and participated in experiment 1 in exchange for course credit.

5.2. EXPERIMENT 1 EXPOSURE. The exposure block in experiment 1 consisted of twelve trials, all with target labels that were novel non-*a*-adjectives that did not occur later in the production block. The lack of overlap between the two blocks was by design, so that speakers' responses in the production block would represent a baseline level of performance that could be compared to the results of experiments 2 and 3, where more active manipulations of the exposure block were planned.

Participants watched as the experimenter used two adjectives three times each in attributive structures, and two different adjectives three times each in relative clause structures. Since attributive and relative clause structures were evenly represented, there was nothing in the exposure block overall to suggest that one kind of structure was privileged over the other. Participants could thus reasonably be expected to anticipate that either an attributive or a relative clause response would be legitimate later on in the production block.

5.3. RESULTS. Responses were registered for 313 critical trials—a recording error led to the loss of seven trials from one participant. Additionally, all of the data from another participant were excluded from analysis because his answers to the debriefing questions revealed explicit knowledge of the prohibition against attributive *a*-adjective use. This left us with 297 analyzable data points, from which we summarized the experiment 1 results by calculating the proportion of attributive responses in each condition. The overall data pattern is shown in Figure 3; condition means are given in Table 2.

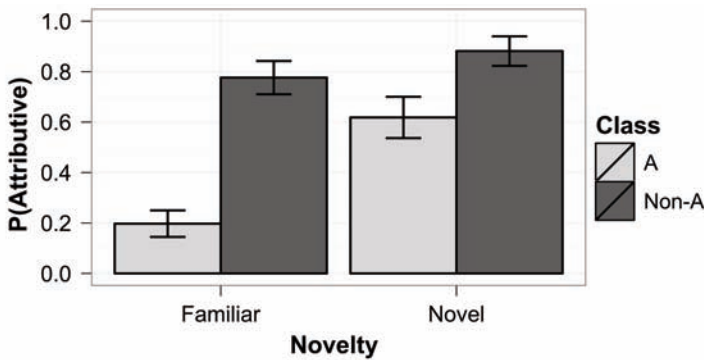


FIGURE 3. Experiment 1 results. Error bars indicate SEM.

	<i>M</i>	<i>SE</i>
FAMILIAR		
<i>a</i> -	0.20	0.05
non- <i>a</i> -	0.78	0.07
NOVEL		
<i>a</i> -	0.62	0.08
non- <i>a</i> -	0.88	0.06

TABLE 2. Proportions of attributive responses in experiment 1.

Comparison between different logit mixed models of the experiment 1 data indicated that the optimal balance between model fit and the number of parameters used was achieved when adjective class, adjective novelty, and their interaction were specified as fixed effects, and random intercepts were allowed for participants. Other random effects did not significantly improve model fit. The model is summarized in Table 3.

FIXED EFFECTS				
	estimate	SE	Z	p
intercept	0.75	0.32	2.36	0.018
adjective class = <i>a</i> -	-2.67	0.35	-7.69	< 0.0001
adjective novelty = novel	1.68	0.34	5.01	< 0.0001
interaction = <i>a</i> - and novel	1.26	0.66	1.90	0.058
RANDOM EFFECTS				
group	SD			
participant (intercept)	1.18			

TABLE 3. Mixed model parameters for experiment 1; parameter estimates indicate likelihood of an attributive response.

Parameter estimates for the likelihood of an attributive response are given in log-odds space. We refer to these when discussing statistical significance, but relate them to the proportions shown in Figure 4 in order to make our results more easily accessible. For example, Table 3 indicates that the estimate of -2.67 for adjective class is significantly different from zero. The negative value indicates that the likelihood of an attributive response was reliably lower for *a*-adjectives than non-*a*-adjectives. This is reflected in different proportions of attributive responses in the two conditions: 41% vs. 83%, respectively. Adjective novelty also played a significant role in determining participants' behavior. The parameter estimate of 1.68 indicates that the likelihood of an attributive response was higher for novel adjectives than familiar adjectives (75% vs. 49%). Further, there was a marginal interaction of adjective class and novelty: the class effect was larger for familiar than novel adjectives.

The reliability of the interaction was verified using the conditional-inference tree represented in Fig. 4. The conditional-inference algorithm works by looking for ways to split the dataset that will result in the creation of homogeneous subsets. Here, partitioning based on adjective class at node 1 results in the creation of a subset in which greater than 80% of trials were attributive responses (node 5: all non-*a*-adjectives). The node 1 *p*-value of less than 0.001 indicates that adjective class is a good predictor of whether participants will produce an attributive or relative clause structure. The subsequent partition at node 2 divides *a*-adjectives according to novelty. This allows for the creation of another nearly homogeneous subset in which most trials were relative clause responses (node 3: familiar *a*-adjectives). The low *p*-value at node 2 again represents good predictive strength. The fact that *a*-adjectives were partitioned according to novelty while non-*a*-adjectives were not is what indicates a significant interaction pattern: adjective novelty is predictive of whether participants will respond with an attributive or relative clause structure, but only for *a*-adjectives.

While it is clear from these results that speakers avoided the use of familiar *a*-adjectives in attributive structures, it is not yet clear whether there was significant attributive avoidance for novel *a*-adjectives. That is, while the presence of an interaction effect indicates that there was more attributive avoidance for familiar than novel *a*-adjectives, it does not say whether adjective class was significant when novel adjectives are considered alone. As outlined above, this issue is important because it bears on whether speakers' grammars specify an abstract *a*-adjective class. To answer this question, we again made use of a conditional inference test, but this time withheld all data points from familiar trials. This analysis confirmed that novel *a*-adjectives did indeed appear in attributive position significantly less often than novel non-*a*-adjectives ($p < 0.001$). So while it is true that familiar *a*-adjectives show stronger attributive avoidance than novel *a*-adjectives, it is also the case that speakers show a significant dispreference for the at-

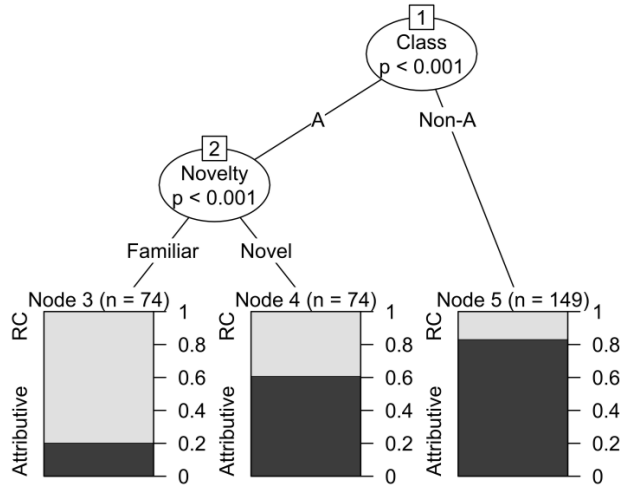


FIGURE 4. A conditional inference tree demonstrating a significant adjective class-by-novelty interaction in experiment 1 (RC = relative clause).

tributive use of novel *a*-adjectives. This behavior is consistent with the existence in the grammar of an abstract *a*-adjective representation.

5.4. EXPERIMENT 1 DISCUSSION. The hypothesis that adjectives like *a*float, *a*sleep, and *a*fraid resist attributive placement based solely on item-specific knowledge predicts that speakers should avoid using these adjectives attributively, but that they should show no such dispreference for *a*-adjectives that they have never heard before. The fact, however, that participants in experiment 1 were significantly less likely to use both familiar AND novel *a*-adjectives attributively relative to their non-*a* counterparts suggests generalization and categorization.

While the dispreference for the attributive use of novel *a*-adjectives indicates that they were treated as members of an abstract *a*-adjective category, such categorization appears to have been tentative. That is, novel *a*-adjectives patterned differently from familiar *a*-adjectives with respect to the DEGREE of dispreference shown. Speakers only subtly dispreferred the attributive use of novel *a*-adjectives, whereas they overwhelmingly dispreferred using familiar *a*-adjectives attributively. What accounts for this difference in behavior? One possibility is that membership in the *a*-adjective category is gradient, and the degree of dispreference that an adjective shows for attributive use is directly proportional to the degree to which it is viewed as an *a*-adjective. On this analysis, novel *a*-adjectives in experiment 1 may have been viewed as *a*-adjectives on the basis of phonology—they all consisted of a schwa followed by a stressed syllable—but not necessarily on the basis of morphology, where the segmentation suggested in §2 would have resulted in unrecognizable ‘stems’ like *blim* and *dax* (cf. *ablim* and *adax*). This hypothesis predicts that making these stems recognizable—providing them with meanings that are semantically related to those of *ablim* and *adax*—should make them less likely to be used attributively relative to a control condition where no exposure is given. This prediction is currently being tested in other experimental work.

Determining what sorts of criteria speakers use to judge *a*-adjective category membership is, however, orthogonal to our present goals. Regardless of how the category is defined in the minds of adult speakers, our interest is in how individuals learn to avoid

using *alive*, *asleep*, and *afraid* attributively. As discussed in the introduction, we hypothesize that speakers learn the restricted distribution through a process of statistical preemption. This predicts that giving speakers experience with novel *a*-adjectives—experience that demonstrates their consistent predicative use in relative clause structures and not in attributive structures—will cause them to avoid using novel *a*-adjectives attributively. In particular, participants who see words like *adax* and *ablim* modeled in relative clause structures but not attributive structures should begin to treat them like existing *a*-adjectives, since the input would suggest that *adax*, *ablim*, *alive*, and *asleep* all disprefer attributive use.

6. EXPERIMENT 2: STATISTICAL PREEMPTION. In order to determine whether seeing novel *a*-adjectives used in relative clause structures causes them to pattern more similarly to existing *a*-adjectives, a new set of participants was run on a slightly modified version of experiment 1. As before, participants described slides that showed one of two animals moving toward a star. This time, however, during the exposure block, half of the experiment's novel *a*-adjectives were modeled for participants in an alternative that was preemptive in the context of the experiment: the relative clause construction. According to the hypothesis outlined above, this should lead to reduced rates of attributive production for novel *a*-adjectives, since speakers would have indirect negative evidence against using them attributively.

Experiment 2 additionally provides a method for determining whether statistical preemption can have a general effect on a category, or whether it instead is effective only on individual items. That is, since only half of the experiment's novel *a*-adjectives are experienced in a preemptive relative clause context, we allow for the possibility of differential patterning based on exposure: novel *a*-adjectives that are represented in the exposure block (the EXPOSURE condition) might show greater attributive resistance than those that are not (the NO EXPOSURE condition). This outcome would be consistent with the hypothesis that statistical preemption operates only on items: each *a*-adjective's distribution would be learned on an item-by-item basis. By contrast, if all novel *a*-adjectives show increased attributive avoidance, then this would constitute evidence that learners are able to generalize indirect negative evidence: that statistical preemption can be applied to categories of items. This sort of outcome would corroborate the interpretation from experiment 1 that reduced attributive use for novel *a*-adjectives relative to novel non-*a*-adjectives signals a generalization of the restriction.

6.1. PARTICIPANTS. A new set of twenty adult native speakers of English was recruited from the same source as experiment 1. All participants took part in exchange for course credit.

6.2. EXPERIMENT 2 EXPOSURE. Recall that the exposure block in experiment 1 featured four novel non-*a*-adjectives that did not appear later in the production block. Two of these were modeled by the experimenter three times each in attributive structures, and two were modeled three times each in relative clause structures. The innovation in experiment 2 was to replace the latter two items with two novel *a*-adjectives that crucially WOULD appear later in the production block as target labels in critical trials. This provided participants with brief, potentially preemptive relative clause exposure to some novel *a*-adjectives under the guise of illustrating how the task was to be performed.

All other aspects of the materials and procedure were identical to experiment 1.

6.3. RESULTS. Responses were recorded for 320 critical trials. Since no participant indicated any explicit knowledge of the correlation between membership in the *a*-adjec-

tive class and the avoidance of attributive placement, all 320 trials were analyzed below. Summary proportion means for the four main experimental conditions are given in Table 4.

	<i>M</i>	<i>SE</i>
FAMILIAR		
<i>a</i> -	0.14	0.05
non- <i>a</i> -	0.90	0.03
NOVEL		
<i>a</i> -	0.24	0.06
non- <i>a</i> -	0.94	0.03

TABLE 4. Proportions of attributive responses in experiment 2.

We first considered whether there was an effect of exposure on the use of novel *a*-adjectives: did novel *a*-adjectives that were modeled for participants in the exposure block pattern differently from those that were not? All data points from the novel *a*-adjective condition were submitted to conditional inference tree analysis. The results show that while items in the exposure condition were produced attributively at numerically lower rates than items in the no exposure condition (20% vs. 28%), the conditional inference algorithm declined to posit a partition based on exposure. This indicates that novel *a*-adjectives patterned similarly at test, regardless of whether participants had witnessed them during the exposure block. Based on this result, subsequent analyses collapsed across exposure status.

As in the experiment 1 analysis, model comparison was used to identify a logit mixed model that balanced model fit with complexity. This resulted in a model that included adjective class and novelty as fixed effects, and random intercepts for participants. In contrast to the experiment 1 model, the addition of an adjective class-by-novelty interaction failed to significantly improve fit, and so was omitted. The experiment 2 model is summarized in Table 5. The effect of adjective class was similar to that reported for experiment 1: the parameter estimate of -4.23 is significantly different from zero, and indicates that *a*-adjectives were roundly dispreferred in attributive structures compared to non-*a*-adjectives (19% vs. 92%). The numbers for adjective novelty were in the same direction as in experiment 1—increased attributive usage for novel over familiar adjectives (59% vs. 52%)—but, as indicated by the *p*-value of 0.069, the effect was only marginally significant.

	FIXED EFFECTS			
	estimate	<i>SE</i>	<i>Z</i>	<i>p</i>
intercept	0.50	0.23	2.18	0.029
adjective class = <i>a</i> -	-4.23	0.38	-11.04	<0.0001
adjective novelty = novel	0.65	0.36	1.82	0.069
	RANDOM EFFECTS			
group	SD			
participant (intercept)	0.62			

TABLE 5. Mixed model parameters for experiment 2; parameter estimates indicate likelihood of an attributive response.

Figure 5 juxtaposes the outcomes of experiments 1 and 2. Recall that participants in experiment 2 witnessed novel *a*-adjectives being used in relative clause structures in the exposure block. This feature of the experimental design was intended to achieve a targeted reduction in attributive responding in the novel *a*-adjective condition. Visual in-

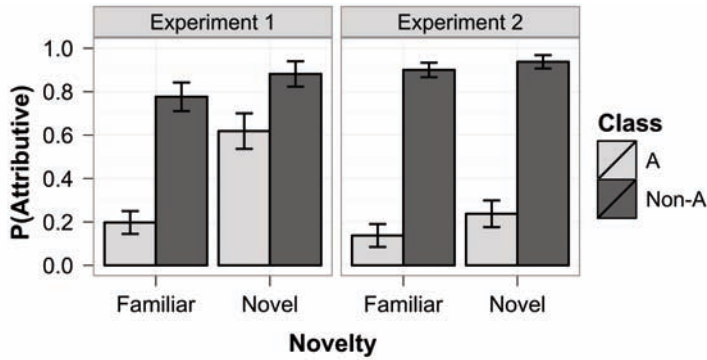


FIGURE 5. Graphical summary of results for experiment 1 (left panel) and experiment 2 (right panel). Error bars indicate SEM.

spection of the data (compare the left and right panels in Fig. 5) indicates that this was indeed the outcome: differences between experiments 1 and 2 were largely confined to the novel *a*-adjective condition. Further, whereas experiment 1 shows a classic interaction pattern in that the effect of adjective class was not equivalent across the different levels of novelty, no such interaction was detectable in experiment 2, even in a conditional inference tree analysis. Instead, novel *a*-adjectives patterned very similarly to familiar *a*-adjectives: both groups uniformly avoided attributive use. Perhaps most surprisingly, this was true even for novel *a*-adjectives that never occurred in the exposure block. This result in particular suggests that the experience that participants gained with just two novel *a*-adjectives in the exposure block generalized to all novel *a*-adjectives in the production block.

6.4. EXPERIMENT 2 DISCUSSION. The results of experiment 2 confirm that speakers can learn to avoid using a particular formulation if, despite the formulation being generally appropriate in a given context, an alternative formulation is witnessed instead. Moreover, the data indicate that the preemptive context itself need not be witnessed for every relevant item; speakers' treatment of the two novel *a*-adjectives that they did not witness in a preemptive context was indistinguishable from their treatment of the two novel *a*-adjectives that they had witnessed. This outcome bolsters the argument that speakers are generalizing the restriction both through a process of categorization (experiment 1), and through a process of statistical preemption (experiment 2). That is, in experiment 1, participants avoided the attributive use of never-seen-before *a*-adjectives to some extent, presumably due to their categorizing the new exemplars as members of the *a*-adjective category. In experiment 2, a second group of participants demonstrated that they generalized evidence gleaned from the statistical preemption of two novel *a*-adjectives to other novel *a*-adjectives: performance was statistically identical at test for novel *a*-adjectives that did and did not occur in the exposure block, and familiar and novel *a*-adjectives were indistinguishable from one another.

An unexpected outcome from experiment 2 was that a very small amount of preemptive exposure—seeing just two novel *a*-adjectives in relative clause structures three times apiece—had very large effects on speakers' tendency to avoid attributive use. Presumably when children initially learn the restriction against using *alive* and *asleep* attributively it takes more input than this. But children who are learning about the distributions of *alive* and *asleep* for the first time differ from the adult speakers of ex-

periments 1 and 2 in that they are in the process of learning an *a*-adjective category, whereas adults already have the category in place. For adults then, learning the distribution of new *a*-adjectives can be very fast: all that is required is that they implicitly recognize words like *ablim* and *adax* as *a*-adjectives. Once this is accomplished, *ablim* and *adax* automatically inherit the distributional statistics associated with the category, which leads to rates of attributive use that are indistinguishable from those of more familiar *a*-adjectives like *alive* and *asleep*.

It thus seems that statistical preemption—especially when combined with a preexisting category—is a powerful process. Indeed, one might be concerned that it appears to be TOO powerful. As noted in the initial discussion, words can typically appear in more than one phrasal construction. Witnessing a word repeatedly in one construction should not necessarily preempt its use in an alternative construction. In order for statistical preemption to work effectively, then, speakers should not be too quick to judge a context as preemptive. A preempted formulation would need to be at least as contextually appropriate as the preempting formulation in order for the latter to count as truly preemptive. But ARE speakers good at identifying truly preemptive contexts and ignoring or DISCOUNTING pseudo-preemptive contexts? This brings us to experiment 3.

7. EXPERIMENT 3: DISCOUNTING PSEUDO-PREEMPTIVE INSTANCES. Participants in experiment 2 witnessed certain novel *a*-adjectives in a preemptive construction—a relative clause—during the exposure block of the experiment. In the context of the experiment, it was clear that either the attributive construction or the relative clause construction would serve as a good descriptor to identify one of the two animals in the scene. Since the constructions' functions in the context of the experiment were more or less equivalent, choosing one was essentially choosing not to use the other. This is exactly where statistical preemption is supposed to apply.

But attributive and relative clause constructions are not always interchangeable—as noted at the outset, different constructions almost always have subtly different constraints on how they are used. What if participants were to witness novel *a*-adjectives in a relative clause structure, but in a context in which their attributive use would not be felicitous for some other reason—that is, for a reason having nothing to do with a potential restriction on the adjective? In experiment 3, we set out to investigate just such a situation.

7.1. PARTICIPANTS. A new set of twenty-one adult native speakers of English took part in experiment 3. As in experiments 1 and 2, these individuals were recruited from the Princeton Department of Psychology subject pool, and participated in experiment 3 in exchange for course credit.

7.2. EXPERIMENT 3 EXPOSURE. Recall that the exposure block in experiment 2 contained six trials in which two novel *a*-adjectives appeared three times each in relative clause structures. Experiment 3 differed subtly from experiment 2 in that these same trials were revised to involve labels in which a novel *a*-adjective was conjoined with a complex modifier composed of an adjective followed by a prepositional complement (e.g. *ablim and proud of himself*). The experimenter used these longer labels in the same way that the novel *a*-adjective labels of experiment 2's exposure block were used: they were incorporated into relative clause structures that described events of animal motion, as in *The hamster that's ablim and proud of himself moved to the star*. All other aspects of the materials and procedure were identical to experiments 1 and 2.

Participants who are exposed to novel *a*-adjectives in this context might respond in one of at least two ways. It is possible that simply seeing words like *ablim* or *adax* in

predicative position in a relative clause structure is enough to trigger preemption and lead to attributive avoidance. On this account, learners would be insensitive to other characteristics of the exposure context (e.g. the fact that *ablim* occurs in the conjoined modifier *ablim and proud of himself*), and the results of experiment 3 should pattern like those of experiment 2. By contrast, it may be the case that learners will fail to learn anything about *a*-adjective distributions from this sort of input. Instead, if they attribute the appearance of *ablim* and *adax* in relative clause structures to constraints on the placement of complex modifiers like *proud of himself* (which cannot appear prenominal), then the presence of novel *a*-adjectives in relative clause structures should fail to preempt their later attributive use, and the results of experiment 3 will pattern like those of experiment 1.

7.3. RESULTS. Responses were collected for 336 critical trials. All of the data from one participant were excluded, however, based on her responses to debriefing questions, which revealed explicit knowledge of the constraint on the attributive use of *a*-adjectives. This left us with 320 analyzable data points, from which the proportion of attributive responses in each condition was calculated as before. Summary means are provided in Table 6. Figure 6 shows the results of all three experiments.

	<i>M</i>	SE
FAMILIAR		
<i>a</i> -	0.31	0.07
non- <i>a</i> -	0.94	0.03
NOVEL		
<i>a</i> -	0.89	0.05
non- <i>a</i> -	0.99	0.01

TABLE 6. Proportions of attributive responses in experiment 3.

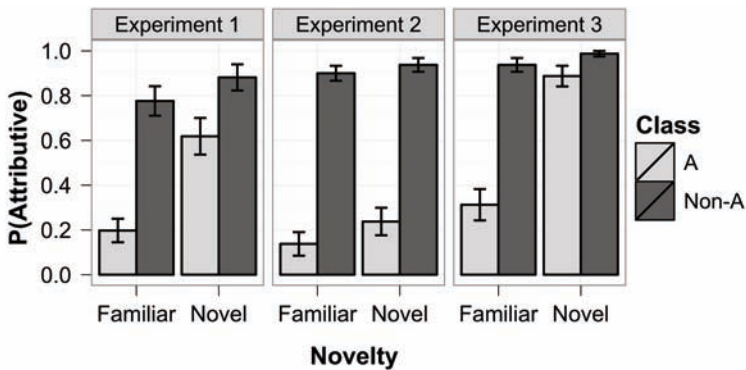


FIGURE 6. Graphical summary of results for experiment 1 (left panel), experiment 2 (middle panel), and experiment 3 (right panel). Error bars indicate SEM.

Visual inspection of the right panel in Fig. 6 suggests that experiment 3—like experiment 1 in the left panel—shows an interaction pattern in which speakers only strongly avoid using familiar *a*-adjectives attributively. The significance of this interaction was confirmed using the conditional inference tree given in Figure 7. In this way, experiment 3 patterns just like experiment 1. At the same time, the optimal logit mixed models for each dataset have different parameter specifications. In particular, the experiment 1 model (Table 3) contained a marginally significant interaction between

adjective class and novelty. In contrast, the optimal model for experiment 3 (see Table 7)—which was arrived at using the same model comparison procedure—achieves the best balance of fit and complexity without an interaction term. The effects of adjective class and novelty are similar to experiment 1. That is, the parameter estimate of -4.02 for adjective class indicates that *a*-adjectives are strongly dispreferred in attributive position relative to non-*a*-adjectives (60% vs. 96%), and the estimate of 3.37 for novelty demonstrates that novel adjectives are significantly more likely to occur attributively than familiar adjectives (94% vs. 63%). But the model is not improved by adding an interaction between adjective class and novelty.

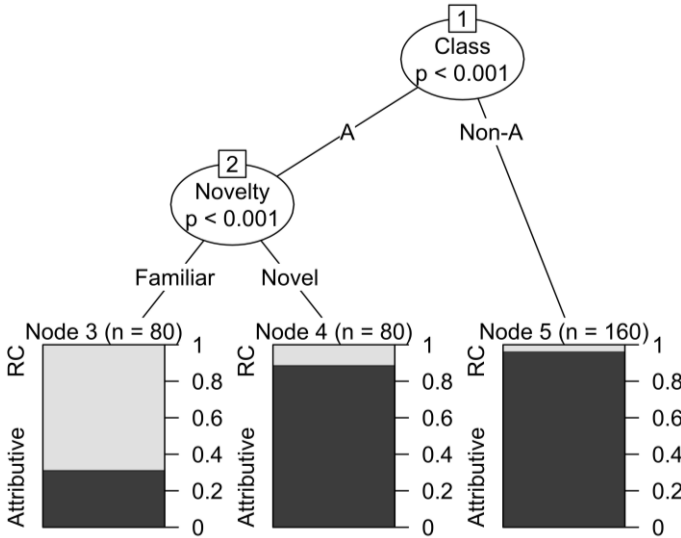


FIGURE 7. A conditional inference tree demonstrating a significant adjective class-by-novelty interaction in experiment 3 (RC = relative clause).

FIXED EFFECTS				
	estimate	SE	Z	p
intercept	2.75	0.40	6.80	< 0.0001
adjective class = <i>a</i> -	-4.02	0.55	-7.28	< 0.0001
adjective novelty = novel	3.37	0.47	7.12	< 0.0001
RANDOM EFFECTS				
group	SD			
participant (intercept)	1.15			

TABLE 7. Mixed model parameters for experiment 3; parameter estimates indicate likelihood of an attributive response.

How is it possible for the mixed model to accurately represent an obvious interaction pattern—as demonstrated visually in Fig. 6 (right panel) and statistically in Fig. 7—without an interaction term? The answer is that an interaction term is not needed when so many observations are at ceiling. In this case, the transformation from logits to the proportions displayed in Fig. 6 predicts an interaction pattern without the need for an explicit interaction term in the model.

Therefore, the summary of experiments given in Fig. 6 can be taken at face value. Both experiments 1 and 3 show larger adjective-class effects in the familiar than novel condition. As mentioned above in the discussion of experiment 1, this is presumably be-

cause participants were less likely to recognize novel *a*-adjectives as members of the *a*-adjective class. When, however, they were given small amounts of preemptive exposure in experiment 2, this counted as evidence that novel *a*-adjectives were dispreferred in attributive position, causing novel and familiar *a*-adjectives to show similar levels of attributive avoidance. Experiment 3 demonstrates that people are savvy about the preemption process: learners carefully consider the informativeness of the input they receive. When the input is less informative—for example, when the appearance of novel *a*-adjectives in relative clause structures can be explained by factors having nothing to do with the adjectives themselves—then preemption does not apply.

Two other aspects of the experiment 3 dataset merit attention. First, Fig. 6 indicates that the numerical difference between novel *a*- and non-*a*-adjectives is smaller in experiment 3 than in experiment 1. Nonetheless, a conditional inference test over data from experiment 3's novel condition demonstrated numerically small but statistically significant attributive avoidance for novel *a*-adjectives relative to novel non-*a*-adjectives ($p < 0.018$). This replicates a similar finding from experiment 1. Speakers avoid using novel *a*-adjectives attributively to some extent, even without truly preemptive exposure. This again provides evidence that speakers' grammars include a generalized *a*-adjective category.

Second, Fig. 6 also indicates that the proportion of attributive responses was numerically higher in every condition in experiment 3 compared to experiment 1 (overall, 78% vs. 62%). A conditional inference test demonstrates that this difference is significant ($p < 0.001$). The overall increase in attributive responses may provide further evidence of participants' rationally discounting relative clause uses during the exposure block, since in experiment 3 the uses were not informative. If participants essentially ignored relative clause structures, the remaining attributive uses during exposure may have taken on correspondingly more salience, resulting in increased overall attributive production via a process of structural priming (e.g. Bock 1986).

7.4. EXPERIMENT 3 DISCUSSION. Experiment 3 demonstrates that adult speakers apply statistical preemption rationally. They discount alternative formulations if they can attribute the alternative to some other factor.

A comparison of experiments 2 and 3 sheds light on the relationship between preemption and a hypothesized role for entrenchment. Recall that the difference between blocking by entrenchment and statistical preemption is that only the latter takes function into account. Blocking by entrenchment predicts that witnessing ANY alternative use of a particular word should discourage the word from being used in a target construction. Preemption, by contrast, predicts that only alternative uses that are potentially in competition with the target construction serve to discourage or preempt use of the target pattern. Note that in both experiments 2 and 3, two novel *a*-adjectives were witnessed three times apiece in relative clause constructions in the exposure block. If simple entrenchment were the key notion, we would expect that this exposure would have had similar effects across the two experiments: attributive responding should have been reduced in the production blocks of both experiments. And yet speakers were more likely to avoid using novel *a*-adjectives attributively in experiment 2 than in experiment 3. This is presumably because the attributive form was preempted only in experiment 2. In experiment 3, the attested relative clause uses did not provide indirect negative evidence because the relative clause was the only possible option for reasons that were irrelevant to the novel adjectives. That is, the attributive use was not in competition with the relative clause use that was witnessed. Experiment 3 therefore suggests that the rel-

evant process wherein speakers learn what not to say is statistical preemption and not simple entrenchment.

8. GENERAL DISCUSSION. The combined results from experiments 1–3 indicate that both categorization and statistical preemption play a role in restricting linguistic productivity. Participants extended the restriction against attributive use from familiar *a*-adjectives to novel *a*-adjectives to some degree, even without witnessing novel *a*-adjectives in a preemptive context (experiment 1). When two novel *a*-adjectives WERE witnessed in a preemptive context—a relative clause construction—attributive uses of all four novel *a*-adjectives dropped to rates indistinguishable from familiar *a*-adjectives (experiment 2). Finally, experiment 3 demonstrated that speakers do not apply preemption indiscriminately: they discounted witnessing novel *a*-adjectives in a relative clause structure when the relative clause use was attributable to some other factor—in this case, a complex modifier that could not be used attributively for independent reasons. Thus, people are adept at taking advantage of their input in order to categorize and generalize beyond their direct experience. In this way, they learn not only what they can say, but also what they cannot say.

Remarkably, our debriefing procedure determined that only a tiny minority of participants—two out of sixty-one—was consciously aware of *a*-adjectives' restricted distributions. These participants' data were excluded from analysis, so the phenomena reported here indicate implicit learning processes.

One might be concerned that the reasoning involved in statistical preemption is too complex for young children to apply when learning a language. But there is ample evidence from nonlinguistic contexts showing that this is not the case: even participants who are not normally noted for their reasoning abilities are able to apply the sort of reasoning that underlies participants' performance in experiments 2 and 3.

For example, Gergely and colleagues (2002) have shown that infants are capable of recognizing and discounting pseudo-preemptive contexts. Fourteen-month-olds were divided into groups whose input paralleled that given to participants in our experiments 2 (preemption) and 3 (discounting). In the preemption group, infants watched an experimenter with a blanket draped around her shoulders turn on a light using her head instead of her hands. When given the opportunity to manipulate the light themselves one week later, these infants largely copied the experimenter's action, illustrating that brief exposure to an alternative method of turning on a light (i.e. with one's head) can preempt the usual method of operating a light (i.e. with one's hands). In the discounting group, infants had the same input, but with one critical change. This time, prior to turning on the light, the experimenter pretended to be cold, pulled the blanket tight around her shoulders and then—with her hands occupied holding onto the blanket—proceeded as before to turn the light on with her head. Now when infants were given an opportunity to manipulate the light one week later the results were very different: the majority of infants in the discounting group attempted to turn the light on with their hands. In this case, infants' reasoning appeared to be: 'if the experimenter were able to use her hands she would have—but her hands were busy holding the blanket'. Thus, infants essentially discounted the implication that the light might appropriately be turned on with the head, because use of the head in this condition could naturally be attributed to an extraneous factor. It seems, therefore, that the type of implicit reasoning involved is accessible even to very young children.

Even more impressive, perhaps, is the surprising ability of other animal species to engage in the same sort of reasoning. Range and colleagues (2007) replicated the

Gergely et al. 2002 result by taking advantage of the fact that domestic canines prefer to operate an apparatus with their mouths rather than their paws. Dogs who witnessed a demonstrator dog use a dispreferred method of manipulating the apparatus (i.e. with its paw) tended to do the same when they were allowed to interact with the apparatus (pre-emption). But dogs who saw the demonstrator dog use its paw when its mouth was otherwise occupied holding a ball did not copy the demonstrator dog. Instead, they operated the apparatus with their mouths (discounting). The fact that both human infants and domestic canines do not mechanically copy everything that they see, but rather practice rational imitation, suggests a phylogenetically conserved reasoning process that is not specific to language, but can be used to arrive at conclusions about what forms are and are not preferred in a grammar.

9. CONCLUSION. Collectively, experiments 1–3 go some way toward establishing how speakers are able to learn arbitrary distributional restrictions in their language—that is, how they learn what not to say. Learners categorize their input. Familiar formulations can statistically preempt other formulations that have roughly the same function, and speakers are able to extend the information gleaned from preemptive contexts, applying it to other instances of the same category. Additionally, speakers are impressively adept at ignoring alternative formulations when they can be attributed to some irrelevant factor.

REFERENCES

- AMBRIDGE, BEN; JULIAN M. PINE; CAROLINE F. ROWLAND; REBECCA L. JONES; and VICTORIA CLARK. 2009. A semantics-based approach to the ‘no negative evidence’ problem. *Cognitive Science* 33.1301–16.
- AMBRIDGE, BEN; JULIAN M. PINE; CAROLINE F. ROWLAND; and CHRIS R. YOUNG. 2008. The effect of verb semantic class and verb frequency (entrenchment) on children’s and adults’ graded judgements of argument-structure overgeneralization errors. *Cognition* 106.87–129.
- ARNOLD, JENNIFER E.; THOMAS WASOW; ANTHONY LOSONGCO; and RYAN GINSTROM. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76.28–55.
- ARNON, INBAL, and NEAL SNIDER. 2009. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62.67–82.
- ARONOFF, MARK. 1976. *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- BAAYEN, R. HARALD. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- BAKER, C. L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10.533–81.
- BANNARD, COLIN, and DANIELLE MATTHEWS. 2008. Stored word sequences in language learning: The effect of familiarity on children’s repetition of four-word combinations. *Psychological Science* 19.241–48.
- BATES, ELIZABETH, and BRIAN MACWHINNEY. 1987. Competition, variation, and language learning. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 157–93. Hillsdale, NJ: Lawrence Erlbaum.
- BOCK, J. KATHRYN. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18.355–87.
- BOD, RENS. 1998. *Beyond grammar: An experience-based theory of language*. Stanford, CA: CSLI Publications.
- BOWERMAN, MELISSA. 1988. The ‘no negative evidence’ problem: How do children avoid constructing an overly general grammar? *Explaining language universals*, ed. by John Hawkins, 73–101. Oxford: Blackwell.
- BOWERMAN, MELISSA. 1996. Argument structure and learnability: Is a solution in sight? *Berkeley Linguistics Society* 22.454–68.

- BRAINE, MARTIN D. S. 1971. On two types of models of the internalization of grammars. *The ontogenesis of grammar*, ed. by Dan I. Slobin, 153–86. New York: Academic Press.
- BRAINE, MARTIN D. S., and PATRICIA J. BROOKS. 1995. Verb argument structure and the problem of avoiding an overgeneral grammar. *Beyond names for things: Young children's acquisition of verbs*, ed. by Michael Tomasello and William E. Merriman, 353–76. Hillsdale, NJ: Lawrence Erlbaum.
- BRESNAN, JOAN; ANNA CUENI; TATIANA NIKITINA; and R. HARALD BAAYEN. 2007. Predicting the dative alternation. *Cognitive foundations of interpretation*, ed. by Gerlof Boume, Irene Krämer, and Joost Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- BROOKS, PATRICIA J., and MICHAEL TOMASELLO. 1999. How children constrain their argument structure constructions. *Language* 75.720–38.
- BROOKS, PATRICIA J.; MICHAEL TOMASELLO; KELLY DODSON; and LAWRENCE B. LEWIS. 1999. Young children's overgeneralizations with fixed transitivity verbs. *Child Development* 70.1325–37.
- BROOKS, PATRICIA J., and OTTO ZIZAK. 2002. Does preemption help children learn verb transitivity? *Journal of Child Language* 29.759–81.
- BROWN, ROGER, and CAMILLE HANLON. 1970. Derivational complexity and order of acquisition in child speech. *Cognition and the development of language*, ed. by John R. Hayes, 11–53. New York: John Wiley and Sons.
- BYBEE, JOAN L., and CAROL L. MODER. 1983. Morphological classes as natural categories. *Language* 59.251–70.
- CHOUINARD, MICHELLE M., and EVE V. CLARK. 2003. Adult reformulations of child errors as negative evidence. *Journal of Child Language* 30.637–69.
- CLARK, EVE V. 1987. The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, ed. by Brian MacWhinney, 1–33. Hillsdale, NJ: Lawrence Erlbaum.
- COLLINS, PETER. 1995. The indirect object construction in English: An informational approach. *Linguistics* 33.35–49.
- COPPOCK, ELIZABETH. 2008. *The logical and empirical foundations of Baker's paradox*. Stanford, CA: Stanford University Press.
- DEMETRAS, M. J.; KATHRYN N. POST; and CATHERINE E. SNOW. 1986. Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language* 13.275–92.
- DI SCIULLO, ANNA MARIA, and EDWIN WILLIAMS. 1987. *On the definition of word*. Cambridge, MA: MIT Press.
- ERTESCHIK-SHIR, NOMI. 1979. Discourse constraints on dative movement. *Syntax and semantics, vol. 12: Discourse and syntax*, ed. by Talmy Givón, 441–67. New York: Academic Press.
- FARRAR, MICHAEL J. 1992. Negative evidence and grammatical morpheme acquisition. *Developmental Psychology* 28.90–98.
- FORAKER, STEPHANI; TERRY REGIER; NAVEEN KHETARPAL; AMY PERFORIS; and JOSHUA B. TENENBAUM. 2007. Indirect evidence and the poverty of the stimulus: The case of anaphoric *one*. *Proceedings of the twenty-ninth annual conference of the Cognitive Science Society*, ed. by D. S. McNamara and J. G. Trafton, 275–80. New York: Lawrence Erlbaum.
- GERGELY, GYÖRGY; HAROLD BEKKERING; and ILDIKÓ KIRÁLY. 2002. Rational imitation in preverbal infants. *Nature* 415.755.
- GOLDBERG, ADELE E. 1993. Another look at some learnability paradoxes. *Proceedings of the twenty-fifth annual Child Language Research Forum*, ed. by Eve V. Clark, 60–75. Stanford, CA: CSLI Publications.
- GOLDBERG, ADELE E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.
- GOLDBERG, ADELE E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- GOLDBERG, ADELE E. 2011. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, to appear.
- GREEN, GEORGIA M. 1974. *Semantics and syntactic regularity*. Bloomington: Indiana University Press.

- GROPEN, JESS; STEVEN PINKER; MICHELLE HOLLANDER; and RICHARD GOLDBERG. 1991. Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language* 18.115–51.
- GROPEN, JESS; STEVEN PINKER; MICHELLE HOLLANDER; RICHARD GOLDBERG; and RONALD WILSON. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65.203–57.
- GUREVICH, OLYA; MATT A. JOHNSON; and ADELE E. GOLDBERG. 2010. Incidental verbatim memory of language. *Language and Cognition* 2.45–78.
- HIRSH-PASEK, KATHY; REBECCA TREIMAN; and MAITA SCHNEIDERMAN. 1984. Brown and Hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language* 11.81–88.
- JAEGER, T. FLORIAN. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.434–46.
- KIPARSKY, PAUL. 1982. Lexical morphology and phonology. *Linguistics in the morning calm: Selected papers from SICOL-1981*, ed. by Linguistic Society of Korea, 3–91. Seoul: Hanshin.
- MARCOTTE, JEAN-PHILIPPE. 2005. Causative alternation errors as event-driven construction paradigm completions. *Constructions in acquisition*, ed. by Eve V. Clark and Barbara F. Kelly, 205–32. Stanford, CA: CSLI Publications.
- MARCUS, GARY F. 1993. Negative evidence in language acquisition. *Cognition* 46.53–85.
- MCNEILL, DAVID. 1966. Developmental psycholinguistics. *The genesis of language: A psycholinguistic approach*, ed. by Frank Smith and George A. Miller, 15–84. Cambridge, MA: MIT Press.
- MOERK, ERNST L. 1991. Positive evidence for negative evidence. *First Language* 11.219–51.
- MORGAN, JAMES L.; KATHERINE M. BONAMO; and LISA L. TRAVIS. 1995. Negative evidence on negative evidence. *Developmental Psychology* 31.180–97.
- OCHS, ELINOR, and BAMBI B. SCHIEFFELIN. 1984. Language acquisition and socialization: Three developmental stories and their implications. *Culture theory: Essays on mind, self, and emotion*, ed. by Richard A. Shweder and Robert A. LeVine, 276–320. Cambridge: Cambridge University Press.
- PINKER, STEVEN. 1981. Comments. *The logical problem of language acquisition*, ed. by C. L. Baker and John J. McCarthy, 53–63. Cambridge, MA: MIT Press.
- PINKER, STEVEN. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press/Bradford Books.
- PINKER, STEVEN. 1999. *Words and rules: The ingredients of language*. New York: Basic Books.
- QUENÉ, HUGO, and HUUB VAN DEN BERGH. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59.413–25.
- R DEVELOPMENT CORE TEAM. 2010. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- RANGE, FRIEDERIKE; ZSÓFIA VIRANYI; and LUDWIG HUBER. 2007. Selective imitation in domestic dogs. *Current Biology* 17.868–72.
- ROBINETTE, JOSEPH. 1999. *The trial of the big bad wolf*. Woodstock, IL: Dramatic Publishing.
- SAXTON, MATTHEW. 1997. The contrast theory of negative input. *Journal of Child Language* 24.139–61.
- SAXTON, MATTHEW; BELA KULCSAR; GREER MARSHALL; and MANDEEP RUPRA. 1988. Longer-term effects of corrective input: An experimental approach. *Journal of Child Language* 25.701–21.
- SIMPSON, JOHN, and EDMUND WEINER (eds.) 1989. *The Oxford English Dictionary*. Oxford: Oxford University Press.
- STEFANOWITSCH, ANATOL. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19.513–31.
- STRAPP, CHEHALIS M.; DANA M. BLEAKNEY; AUGUSTA L. HELMICK; and HAYLEY M. TONKOVICH. 2008. Developmental differences in the effects of negative and positive evidence. *First Language* 28.35–53.

- STROBL, CAROLIN; JAMES MALLEY; and GERHARD TUTZ. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14.323–48.
- THEAKSTON, ANNA. 2004. The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development* 19.15–34.
- THOMPSON, SANDRA A. 1995. The iconicity of 'dative shift' in English: Considerations from information flow in discourse. *Syntactic iconicity and linguistic freezes: The human dimension*, ed. by Marge E. Landsberg, 155–75. Berlin: Mouton de Gruyter.
- TOMASELLO, MICHAEL. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- TREMBLAY, ANTOINE, and HARALD BAAYEN. 2010. Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. *Perspectives on formulaic language: Acquisition and communication*, ed. by David Wood, 151–73. London: Continuum International.
- WASOW, THOMAS. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.
- WICKHAM, HADLEY. 2009. *ggplot2: Elegant graphics for data analysis*. New York: Springer.
- ZOEHFELD, KATHLEEN WEIDNER. 2000. *Aladar's story*. New York: Random House Disney.
- ZWICKY, ARNOLD M. 1971. In a manner of speaking. *Linguistic Inquiry* 2.223–33.

Boyd
University of Illinois at Urbana-Champaign
405 N. Mathews Ave.
Urbana, IL 61801
[jkboyd@illinois.edu]

[Received 13 August 2009;
revision invited 23 February 2010;
revision received 29 July 2010;
accepted 29 September 2010]

Goldberg
Department of Psychology
Princeton University
Princeton, NJ 08544
[adele@princeton.edu]