Samara, A., Wonnacott, E., Saxena, G., Maitreyee, R., Fazekas, J. & Ambridge, B. (in press). Learners restrict their linguistic generalizations using preemption but not entrenchment: Evidence from artificial language learning studies with adults and children. *Psychological Review*. https://osf.io/r7d2e

Learners restrict their linguistic generalizations using preemption but not entrenchment: Evidence from artificial language learning studies with adults and children.

Anna Samara
University of Birmingham
axs2179@student.bham.ac.uk

Elizabeth Wonnacott
University of Oxford
elizabeth.wonnacott@education.ox.ac.uk

Gaurav Saxena
University of Bristol
cp19962@bristol.ac.uk

Ramya Maitreyee
University of Liverpool
ramya.maitreyee@gmail.com

Judit Fazekas
University of Manchester
judit.fazekas@manchester.ac.uk

Ben Ambridge
University of Manchester
ESRC International Centre for Language and Communicative Development (LuCiD).
Ben.Ambridge@Manchester.ac.uk

Word count: 8,946 (main text)

Author Note

Anna Samara https://orcid.org/0000-0001-6503-5181; Elizabeth Wonnacott https://orcid.org/0000-0002-3261-7131; Gaurav Saxena https://orcid.org/0000-0002-9526-2550; Ramya Maitreyee https://orcid.org/0000-0002-8915-8757; Judit Fazekas https://orcid.org/0000-0002-6059-4109; Ben Ambridge https://orcid.org/0000-0002-9631-3671. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 681296: CLASS). Ben Ambridge is Professor in the International Centre for Language and Communicative Development (LuCiD) at The University of Manchester. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged. All data, and analysis code can be found at https://osf.io/4jdvf/ (DOI 10.17605/OSF.IO/4JDVF). The studies' design, hypotheses and analyses were preregistered: Experiment 1: https://rpubs.com/AnnaSamara/458000; Experiments 2-3: https://rpubs.com/AnnaSamara/856144; Experiment 4: https://rpubs.com/AnnaSamara/856146 Experiment 5: https://rpubs.com/AnnaSamara/539534. A preliminary version of this work was presented at the 2019 conference on Interdisciplinary Advances in Statistical Learning, San Sebastian, Spain. The authors declare no conflict of interests exists. We certify that we have complied with APA ethical standards in the treatment of our human sample. Correspondence concerning this article should be addressed to Ben Ambridge, Division of Psychology, Communication and Human Neuroscience, Coupland Building 1, University of Manchester, Manchester, UK, M15 6FH. Email: Ben.Ambridge@Manchester.ac.uk

Learners restrict their linguistic generalizations using preemption but not entrenchment: Evidence from artificial language learning studies with adults and children

A central goal of research into language acquisition is explaining how, when learners generalize to new cases, they appropriately RESTRICT their generalizations (e.g., to avoid producing ungrammatical utterance such as **The clown laughed the man*). The past 30 years have seen an unresolved debate between STATISTICAL PREEMPTION and ENTRENCHMENT as explanations. Under preemption, the use of a verb in a particular construction (e.g., **The clown laughed the man*) is probabilistically blocked by hearing that verb other constructions WITH SIMILAR MEANINGS ONLY (e.g., *The clown made the man laugh*). Under entrenchment, such errors (e.g., **The clown laughed the man*) are probabilistically blocked by hearing ANY utterance that includes the relevant verb (e.g., by *The clown made the man laugh* AND *The man laughed*). Across five artificial-language-learning studies, we designed a training regime such that learners received evidence for the (by the relevant hypothesis) ungrammaticality of a particular unattested verb/noun+particle combination (e.g., **chila+kem*; **squeako+kem*) via either preemption only or entrenchment only. Across all five studies, participants in the preemption condition (as per our preregistered prediction) rated unattested verb/noun+particle combinations as less acceptable for restricted verbs/nouns, which appeared during training, than for unrestricted, novel-at-test verbs/nouns, which did not appear during training; i.e., strong evidence for preemption. Participants in the entrenchment condition showed no evidence for such an effect (and in 3/5 experiments, positive evidence for the null). We conclude that a successful model of learning linguistic restrictions must instantiate competition between different forms only where they express the same (or similar) meanings.

Keywords: *preemption, entrenchment, restricting generalization, overgeneralization.*

Learners restrict their linguistic generalizations using preemption but not entrenchment: Evidence from artificial language learning studies with adults

A central goal of research into human language – one of the defining characteristics of our species – is explaining how learners form generalizations that allow them to produce utterances that they have never heard before (e.g., Chomsky, 1957). A particularly challenging aspect of this problem (e.g., Bowerman, 1998; Pinker, 1989) is explaining how, when producing these novel utterances, learners appropriately RESTRICT their generalizations to avoid producing ungrammatical utterances. For example, English speakers form generalizations which allow them to mark the past tense (e.g., *play→played*), plural (e.g., *dog→dogs*) and the reversal of an action (e.g., *button→unbutton*). How, then, do they learn not to overapply these generalizations to produce **sitted, *mans* or **unsqueeze* (as is conventional in the linguistics literature, '*' indicates an ungrammatical form); see, e.g., Li & MacWhinney, 1996; Köpke, 1998; Boyd & Goldberg, 2011; Ramscar, Dye, & McCauley, 2013),

Argument structure overgeneralization errors

The problem of appropriately restricting generalizations occurs in many linguistic domains – indeed, probably the majority – across different languages. However, one particularly clear and well-studied case is the domain of verb argument structure overgeneralization errors (e.g., **The clown laughed the man*). Researchers of all theoretical persuasions agree that children must form some generalized representation of the *transitive-causative* construction that allows them to produce novel causative sentences of the form [CAUSER] [VERB] [CAUSEE] (e.g., *The man broke the window*). That is, children can use verbs like *break* in transitive-causative sentences (as in the previous example) even if they have never heard the verb used in this way before (e.g. if they have only ever heard *break* in *intransitive-inchoative* sentences; e.g., *The stick broke*). Evidence that children are indeed forming such generalizations comes from utterances where they produce this construction with a novel verb (e.g., *The mouse tammed the ball;* e.g., Akhtar & Tomasello, 1997). How, then, do learners avoid applying this generalization to exception verbs such as *laugh* (e.g., **The clown laughed the man*), in some cases after a period of overgeneralization in which they do produce such errors (e.g., Bowerman, 1988; Pinker, 1989; Ambridge & Ambridge, 2020)?

At least part of the answer seems to lie with meaning. For example, Shibtani and Pardeshi (2002:89) propose that the English transitive causative construction (and its equivalent in other languages) is restricted to verbs whose meaning "entails a spatio-temporal overlap of the causer's activity and the caused event, to the extent that the two relevant events are not clearly distinguishable". For example, *The man broke the window* is a possible utterance because, in normal circumstances, the event that causes the window to break and the window breaking happen at more or less exactly the same time and in the same place. In contrast, **The clown laughed the man* is not a possible utterance because the causing event (e.g., the clown telling a joke) and the caused event (i.e., the man laughing) are two events that are relatively distinct in time and space. For evidence of the importance of meaning in restricting these types of generalizations see – for the transitive causative – Pinker (1989), Brooks and Tomasello (1999) Ambridge, Pine, Rowland and Young (2008), Ambridge, Pine, Rowland, Jones and Clark (2009), Ambridge, Pine and Rowland (2011), Ambridge, Maitreyee, Tatsumi, Doherty, Zicherman, Mateo-Pedro, et al (2020), Ambridge, Doherty, Maitreyee, Tatsumi, Zicherman, Mateo-Pedro et al (2022), Bidgood, Pine, Rowland, Sala, Freudenthal and Ambridge (2021) and – for other constructions – Ambridge, Pine and Rowland (2012), Ambridge (2013), Ambridge, Pine, Rowland, Freudenthal and Chang (2014), Bidgood, Ambridge, Pine and Rowland (2014), Blything, Ambridge and Lieven (2014). However, semantics alone is not a complete solution, since some restrictions seem to be semantically arbitrary (e.g., *manage to do; succeed in doing* vs **succeed to do; *manage in doing*).

Statistical preemption versus (conservatism via) entrenchment

As a solution to the problem of restricting these (often arbitrary) linguistic generalizations, researchers have proposed two statistical-learning mechanisms: *preemption* (e.g., Clark, 1987; Brooks & Tomasello, 1999; Brooks & Zizak, 2002; Goldberg, 1995, 2011, 2019), also known as *statistical preemption*, and *entrenchment* (e.g., Brooks, Tomasello, Dodson & Lewis, 1999; Ambridge et al, 2008; Stefanowitsch, 2008; Ambridge, Bidgood, Twomey, Pine, Rowland & Freudenthal, 2015), also known as *conservatism via entrenchment*). According to preemption, errors such as **The clown laughed the man* are probabilistically blocked by the occurrence in the input of utterances such as *The clown made the man laugh*. More generally, the more often a learner hears a verb in the *periphrastic causative* construction ([CAUSER] make [CAUSEE] VERB), the stronger her inference that that particular verb cannot be used in the transitive causative construction ([CAUSER] [VERB] [CAUSEE]). In other words, according to preemption, the use of a verb in a particular construction is probabilistically blocked by the use of that verb in another construction ONLY if those two constructions have very similar meanings. For example, **The clown laughed the man* would be probabilistically blocked by *Sue made John laugh*, but not by *John laughed*.

Such examples clearly illustrate the difference between preemption and entrenchment. According to entrenchment, errors such as **The clown laughed the man* are probabilistically blocked by the occurrence in the input of ANY utterance that includes the verb *laugh* (e.g., by BOTH *Sue made John laugh* and *John laughed*). More generally, the more often a learner hears a verb in the input, regardless of construction, the stronger her inference that that particular verb cannot be used in an unwitnessed construction ("otherwise I would have heard it by now"). In other words, according to entrenchment, the use of a verb in a particular construction is probabilistically blocked by the use of that verb in another construction REGARDLESS of whether or not those two constructions have similar meanings. Essentially, then, (a) preemption and (b) entrenchment predict, respectively, a negative correlation between the acceptability/production probability of a particular verb+construction combination and the input frequency of that verb (a) in near-synonymous constructions or (b) regardless of construction[1].

Readers who are not familiar with the details of this debate might reasonably wonder whether entrenchment is really conceptualized in such a simplistic fashion in the empirical literature. Yet this is exactly the version of entrenchment tested (and generally supported) in, for example, Brooks, Tomasello, Dodson and Lewis (1999), Theakston (2004), Ambridge & Brandt (2014), Blything, Ambridge & Lieven (2014) and Ambridge et al., (2008, 2009, 2011, 2020). Moreover, both Stefanowitsch (2008) and Ambridge et al., (2012, 2015) provided apparent evidence that entrenchment, in this conceptualization, outperforms preemption. Other studies provided evidence for preemption, without systematically comparing it against entrenchment (e.g., Brooks & Tomasello, 1999; Brooks & Zizak, 2002; Goldberg, 2011; Boyd & Goldberg, 2011).

Despite a great deal of research, however, the preemption-entrenchment debate remains unresolved: Studies that have attempted to differentiate the two accounts using one or more natural languages (e.g., Stefanowitch, 2008; Robenalt & Goldberg, 2015; Blything et al, 2014; Ambridge et al, 2012, 2014, 2015, 2018, 2020; Ambridge, 2013). do not clearly do so, since pre-emption counts are a subset of entrenchment counts and are inevitably highly correlated (e.g., Westfall &Yarkoni, 2016). Some also incorrectly interpret absence of evidence for an effect (e.g., for pre-emption in Stefanowitsch, 2008) as positive evidence of absence of the effect (e.g., Altman & Bland, 1995; Dienes, 2014).

A more promising approach for dissociating preemption and entrenchment, then, lies with artificial-language-learning studies, which allow the two to be manipulated independently. To the best of our knowledge, however, all such studies conducted to date – like many of the natural-language studies summarized above – have tested only preemption OR entrenchment, or a general statistical-learning mechanism that collapses the two (Wonnacott, Newport & Tanenhaus, 2008; Wonnacott, 2011; Wonnacott, Boyd, Thomson & Goldberg, 2011; Wonnacott, Brown & Nation, 2017; Perek & Goldberg, 2015, 2017; Robenalt & Goldberg, 2016; Harmon and Kapatsinski, 2017). Although two studies have demonstrated effects of preemption while holding entrenchment constant (e.g., Perek & Goldberg, 2015, 2017), none have directly compared – or even investigated – both putative learning mechanisms in a single study.

The present study

The aim of the present work is to move the field closer to a definitive conclusion to this long-running debate, by using an artificial-language-learning paradigm to better differentiate the preemption and entrenchment hypotheses. A particularly important aspect of these studies is the use of pre-registered Bayesian analyses with informed priors, which enable us to make inferences regarding the presence or – crucially – *absence* of preemption and entrenchment effects (or, if appropriate, to conclude that the data are inconclusive). First (Experiment 1), we report an adult study designed to map closely onto the English examples set out above. That is, we set up an artificial language in which sentences equivalent to **The clown laughed the man* are systematically absent. (Of course, since the language is novel, there is no sense in which a particular utterance is (un)grammatical, even probabilistically, outside of the context of the study). Within each study, we investigate whether we can, via training, induce participants to consider such utterances as *less* acceptable than other unwitnessed utterances; those with verbs introduced for the first time during the test session, as is predicted to be possible under the pre-emption and entrenchment hypotheses. This training consists of presenting the relevant verb (i.e., the artificial-language equivalent of *laugh*) in sentences with identical meaning to the systematically-absent sentences (equivalent to *The clown made the man laugh;* preemption condition) or with a different meaning (equivalent to *The man laughed*; entrenchment condition). This scenario is analogous to the real-world scenario by which (by hypothesis) English speakers come to consider perfectly-interpretable utterances such as **The clown laughed the man* as less than fully acceptable, as a function of implicit probabilistic learning (i.e., pre-emption / entrenchment), rather than by (even implicit) categorization or prohibition *per se*. Thus, under the entrenchment and preemption hypotheses, even natural languages do not draw a binary distinction between grammatical and ungrammatical utterances. Nevertheless, when talking about both natural and artificial languages, it is convenient to use the term "ungrammatical" as a shorthand for "less acceptable than equivalent utterances with other lexical items [here, untrained verbs/nouns].

To ensure reliability, we then replicate this study using (Experiment 2) the same materials and procedure, and (Experiment 3) increased training. Next, we report an analogous study in which the problem is mapped onto a simpler domain: noun plural marking

(Experiment 4), including a replication with children (Experiment 5). For all studies, the dependent measure reported in the present article is participants' grammatical acceptability judgment ratings (on a 5-point scale). As noted in the pre-registration (https://rpubs.com/AnnaSamara/458000), in addition to this "key" measure, we also collected production data as a "secondary…exploratory" test of our hypotheses. For reasons of space, the production task and results are reported in a separate companion article, which also reports a discriminative learning model that simulates both the judgment and production data (pre-print available at https://osf.io/4jdvf/).

Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow APA Journal Article Reporting Standards (Kazak, 2018). All data, analysis code, and research materials are available at https://osf.io/4jdvf/ (https://doi.org/10.17605/OSF.IO/4JDVF). For all studies, our planned data analyses, including decisions on participant exclusion, trial exclusion and criteria for sample size determination were pre-registered (see ***Preregistration and Data availability*** section below for links). Experiment 1 deviated partially from the pre-registration by adding tests which we realized retrospectively were key for testing our main hypotheses (those comparing unattested verb+particle combinations in restricted versus novel-at-test verbs). For this reason (and because replication is important) we ran Experiment 2 as a preregistered direct replication of Experiment 1.

Experiments 1-3: Adult verb argument structure

Method

Ethics.

All studies were approved by the Research Ethics Committee of the University of Liverpool (RETH001041 - Artificial Grammar Learning studies); all adult participants gave informed consent to their participation. All child participants gave verbal assent, informed consent having been provided by a parent or responsible caregiver.

Participants

With regard to diversity, it is important to note that although this and all subsequent studies were conducted with English speakers (the most convenient option given the study locations; the UK and India), since the trained languages are novel, our conclusions apply in principle to speakers of any language worldwide. Indeed, globally, the majority of languages – like our novel languages – use morphological suffixes to mark causality and/or plurality (Haspelmath, 1993; Shibatani & Pardeshi, 2003). Although, across studies, the majority of participants were monolingual English speakers, bilingual speakers were not systematically excluded; indeed, English/Hindi bilingual speakers constituted around half of the participants in the key final study with children (Experiment 5).

Full details of the recruitment, inclusion and exclusion criteria can be found in the supplementary online material at: https://osf.io/4jdvf/. For Experiment 1 and 2, the final sample size was N=80, divided equally between the preemption and entrenchment conditions. For Experiment 3 (identical to Experiments 1-2, except with twice the amount of training), the final sample size was N=113, 40 in the entrenchment condition and 73 in the preemption condition.

Training language and stimuli

The training language was designed to correspond to the English verb-argument examples outlined in the Introduction (although the general scenario whereby some verbs alternate between two constructions and others are restricted to a single construction applies more broadly, e.g., in at least 38 different languages; Shibatani & Pardeshi, 2002). The design is summarized in Figure 1, and set out in full in Appendix A. All participants were taught three novel verbs, with the same particle restrictions: One verb appeared 64 times with particle a only (e.g., *chila gos*), one verb appeared 64 times with particle b only (e.g., *tombat kem*), and one verb appeared 32 times with particle a (e.g*., coomo gos*) and 32 times with particle b (e.g., *coomo kem*). What differs between conditions is the nature of the evidence that participants can use to learn these restrictions. That is, while all participants heard **exactly the same sentences**, the conditions differed as to whether *gos* and *kem* both compete for the same, causal meaning[2] – as in the *preemption* condition – or whether one has a causal meaning and the other a noncausal meaning, as in the *entrenchment* condition. This maps onto the differing assumptions that ungrammatical forms such as **The clown laughed the man* are blocked **only** by appearances of *laugh* in a construction with the same meaning (e.g., *The clown made the man laugh*) – preemption – or **also** by appearances of *laugh* in a construction with a different meaning (e.g., *The man laughed*) – entrenchment.

Thus, each participant, regardless of condition, heard four sentences (e.g., *chila gos* x 64*, tomabt kem* x64*, coomo gos* x32*, coomo kem* x32), divided across 16 training blocks. What differed between the preemption and entrenchment conditions was solely the animations with which the sentences were paired.

As shown in Figure 1, in the preemption condition, all animations were causal (i.e., a boy causing a ball to bounce, roll etc.); in the entrenchment condition, one marker (e.g., *kem*) was always paired with causal animations (i.e., a boy causing a ball to bounce, roll etc.), the other (e.g., *gos*) was always paired with noncausal animations (e.g., a ball bouncing, rolling etc. of its own accord). The number of characters present (four) was held constant across animations, but none (noncausal animations) or one (causal animations) was shown interacting with the ball, with the remainder clearly far away in the background. All animations were created with Moho (https://moho.lostmarble.com). Note that because participants in the preemption and entrenchment conditions hear exactly the same sentences – with only the animations varying between conditions – this study constitutes a uniquely stringent test of the claim that preemption is "special", as compared to entrenchment; i.e., that constructions *competing for the same meaning* is crucial. If linguistic restrictions are learned purely in a surface-based manner, we would expect to see no difference between the preemption and entrenchment conditions, since the utterances heard are *identical* in the two conditions.

In terms of sentence construction, each took the form *VERB PARTICLE*, where an artificial verb (e.g., *chila*) was followed by one of two obligatory artificial particles (e.g., *kem*). The sentences used no character names (similar to argument-drop languages) in order to minimize the possibility of transfer effects from English. Particle words were always *kem* or *gos*. For each participant, three out of five nonwords (*chila, coomo, roosa, panjol, tombat*) were randomly drawn and used as training verbs, with the remaining two reserved for use as novel-at-test verbs in the subsequent test phase (one for the acceptability judgment test; one for the production test). Verbs took the meanings of *bounce*, *spin*, *drop*, *slide*, and *roll*, with the pairing between meaning and phonological form counterbalanced across participants. These meanings were chosen on the basis that all could be performed by the same item (e.g., *a ball bouncing*, *rolling*, *spinning* etc.) and all are semantically consistent with both causal and noncausal events in English[3] (we can say both *Someone bounced/rolled/spun the ball* and

The ball bounced/rolled/spun). Controlling for semantics in this way is crucial to ensure that participants can use only the distributional patterns in which each verb appears (i.e., preemption or entrenchment) to make inferences regarding the (un)grammaticality of particular utterances. Sentences were recorded by a female speaker (a near-native speaker of English) in a soundproof recording booth and the sound files normalised using Audacity (http://audacity.sourceforge.net/).

Procedure

The experiment was designed and run using the online experiment generation software Gorilla.sc (https://gorilla.sc/) (Anwyl-Irvine et al., 2020). Participants completed the experiment remotely on their own computers in a single session lasting between 45 and 65 minutes. They were introduced to Freddy the frog who "speaks another language" and were told that the aim of the study was to "learn how to say things in his language".

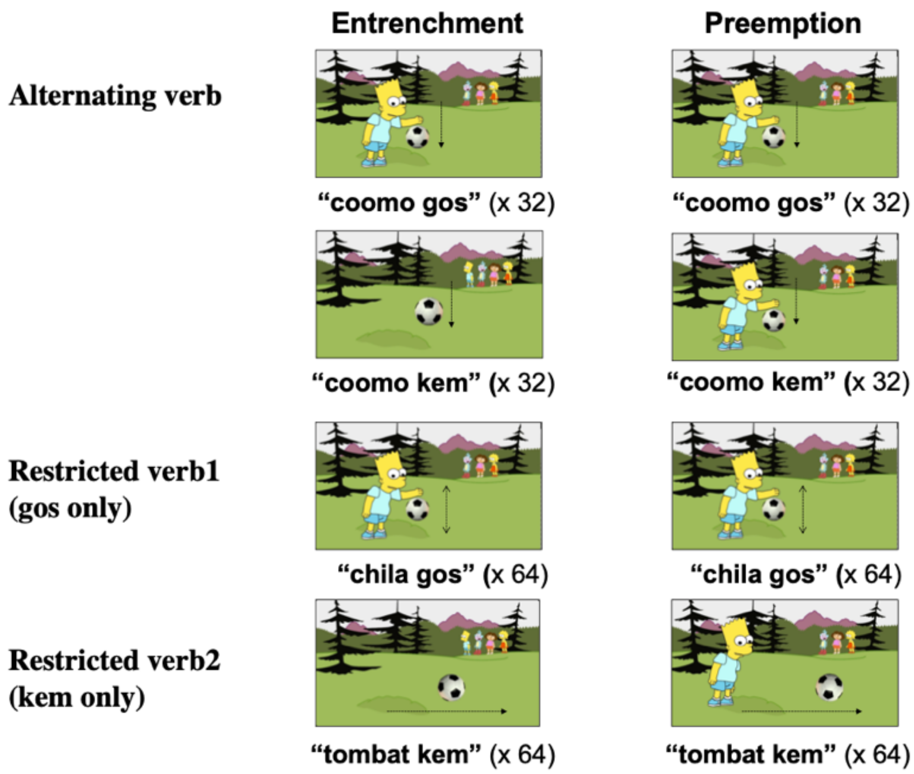

Figure 1. Snapshots of video scenes and accompanying sentences during training in the preemption and entrenchment conditions of Experiment 1. Note: The direction of motion (shown with an arrow) comes from the meaning of the novel verb. Causality is indicated by the presence of absence of a causing agent (the boy in the examples shown).

Language training. At the beginning of the experiment, participants were told that they were going to be trained in Freddie's language by hearing and copying his sentences ("copy-only" blocks) and/or helping him finish his sentences ("training-with-recast" blocks). Copy-only and training-with-recast blocks were alternated eight times, for a total of 16 training blocks.

During "copy-only" blocks (all odd number blocks from 1 to 15; 12 trials/block, n = 96), participants viewed the relevant animation and heard a sentence describing the video (e.g., *chila gos*), followed by brief presentation (2000 ms) of the written form of the sentence (in a speech bubble) to ensure that they had reference for producing accurate spellings of the novel words. On each "copy-only" trial, participants were required to type the sentence they had just seen and heard, in order to be able to move on to the next trial (though no feedback was given, and moving on was not contingent upon a correct response).

During "training-with-recast" blocks (even number blocks from 2 to 16), Freddie produced only the verb, with no particle (e.g., *chila…*) in both the audio and on-screen-text forms. Participants were required to "have a go at finishing Freddie's sentence" by typing the relevant particle, and then received both audio and on-screen-text feedback on "how Freddie would have said it". Again, although a response was required in order to move on to the next trial, the same feedback was given whether or not the participant's response was correct. Note that we did not score participants' responses during training. However, recall that we excluded and replaced participants who did not show above-chance performance on the tasks measuring vocabulary learning and (in the entrenchment condition) the semantics (causal/noncaual) associated with each particle (*gos/kem*).

Grammaticality judgment task. Participants were told that they would see some more animations and would be asked to rate, using a five-point smiley face scale, how well a sentence in Freddie's language describes what they see in the animation (see Figure 2). Participants received no feedback, except for 4 practice trials in which they rated grammatical versus ungrammatical English sentences (e.g., **His teeth man the brushed*). The grammaticality judgment test comprised 32 trials presented in randomized order. Twenty-four of these featured one of the three trained verbs (each encountered 8 times; four times with each of the two particles) and the remaining 8 tested performance on 1 novel verb, also presented 8 times (again, four times with each of the two particles). As during training, for participants in the preemption condition, all animations were two-participant (causal), while for participants in the entrenchment condition, half were two-participant (causal), while half were one-participant (noncausal).

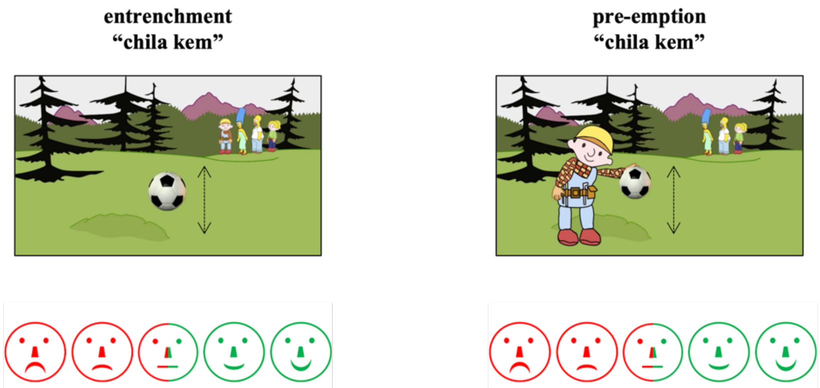

Figure 2. Grammaticality judgment test (Experiment 1): Snapshots of trials featuring unattested sentences (whereby *chila* was only attested with *gos* during training) and accompanying (semantically correct) video scenes. Participants were asked to click with the mouse cursor on the smiley face that expressed how well the sentence went with the video.

Analyses

Full details of the statistical approach, along with all data and analysis scripts can be found in the supplementary online material at: https://osf.io/4jdvf/. In brief, we built maximal (Barr et al, 2013) Bayesian Multi Level Models (MLMs) in the R package brms (Bürkner, 2016; R Core Team, 2023) and used the Bayes factors (*BF*) procedure outlined in Dienes (2008,2015) to compute the evidence for the experimental hypothesis (H1) over the null hypothesis (H0) given the data (see the supplementary material for details, including the priors used).

Experiments 1-3 (and the subsequent Experiments 4-5) tested three key predictions :

(1) Statistical preemption will constrain verb argument structure generalizations in adults.
(2) Entrenchment will constrain verb argument structure generalizations in adults.
(3) The effect of statistical preemption will be larger than the effect of entrenchment (based on the computational modelling studies of Ambridge et al, 2020; 2022; note that, as discussed in the Introduction, empirical studies remain unclear on this question).

Predictions (1) and (2) were both modelled using the *ratings participants gave to scenes featuring semantically-appropriate[4] yet unattested (i.e., notionally "ungrammatical") verb+particle combinations* in the grammaticality judgment test as the dependent variable, with verb type (restricted, novel-at-test) as a within-subject predictor. "Restricted" here means that, during training, a verb appeared either *only* with particle a (e.g., *kem*) or only with particle b (e.g., *gos*) and never both. Recall that, for all participants, two verbs were "restricted" in this sense, while one alternated between the two particles (i.e., 50% *kem*; 50% *gos*). "Novel-at-test" here refers to the fact that, for each participant, two of the verbs (+actions) created for the purposes of the experiment were never presented during training, and were reserved for use only in the test phase[5]. We use the rather long-winded phrase "novel-at-test" (rather than simply "novel") because all verbs in the study were novel, in the sense that they are not familiar English verbs. **Critically, if participants restricted their generalizations given the training input they received, there will be a main effect of verb-type such that unattested verb+particle combinations are rated lower (i.e., less acceptable) for restricted than novel-at-test verbs.**

In addition to the critical predictor of verb type (restricted, novel-at-test), some models – i.e., those for which it improved model fit – also included the control predictor of *tested scene* (scene a, scene b). For participants in the entrenchment condition, this corresponds to causal/noncausal; for participants in the preemption condition this distinction is arbitrary, since all scenes were causal (it is included only to ensure parity with the entrenchment condition, particularly in analyses that collapse across the two conditions).

Prediction (3) was tested by fitting the same model as for predictions (1) and (2) onto the pooled grammaticality judgment dataset, with the additional factor of condition (entrenchment, preemption). A significant verb type by condition interaction would suggest that participants restricted their generalizations more in the preemption than entrenchment condition (or vice versa).

Preregistration and Data availability

Data from this and all subsequent experiments are available on the Open Science Framework: https://osf.io/4jdvf/ (DOI: 10.17605/OSF.IO/4JDVF). This repository contains (a) an overall summary of the code and output across all experiments in R Markdown html format (https://osf.io/tzjm4) and (b) raw data and executable code for the adult and child experiments (https://osf.io/ur7de). The studies' design, hypotheses and analyses were preregistered: Experiment 1: https://rpubs.com/AnnaSamara/458000; Experiments 2-3: https://rpubs.com/AnnaSamara/856144; Experiment 4: https://rpubs.com/AnnaSamara/856146
Experiment 5: https://rpubs.com/AnnaSamara/539534.

Results (Experiments 1-3)

Research question 1. Does statistical preemption constrain verb argument structure generalizations in adults ?

Figure 3 (right side of each panel) shows preemption participants' acceptability ratings for trials (all semantically correct; see Footnote) featuring unattested verb+particle combinations for (red) restricted verbs that had been presented with either particle a or particle b (but never both) during training and (blue) novel-at-test verbs that had *never* been presented during training. Our pre-registered prediction was that, if statistical preemption constrains verb argument structure generalizations, participants will rate unattested verb+particle combinations as less acceptable **for restricted (blue) than novel-at-test and hence unrestricted verbs (red)**. Recall that this is a very strong test, since – given that the novel-at-test verbs had not been presented during training – *all* verb+particle combinations were previously unwitnessed; thus, any effect cannot be a mere familiarity preference[6]. Rather it must reflect an inference that *some* unwitnessed verb+particle combinations are unacceptable (i.e., for verbs that have been trained with a competing particle) and some are acceptable (i.e., for verbs that have not been trained at all).

For all three studies, the analysis found a reliable effect of verb type (restricted, novel) in the predicted direction, with Bayes Factors of approximately 236, 10 and 10 across Experiments 1, 2 and 3 respectively. **Experiment 1**: $b = 0.64$, $SE = 0.17$, $pMCMC < .001$, *CI*s [0.32, 0.97], $BF_{(0,0.575)} = 236.50$, RR[0.06, > 4]. **Experiment 2**: $b = 0.48$, $SE = 0.18$, $pMCMC = .023$, *CI*s [0.01, 0.74], $BF_{(0,0.65)} = 10.53$ (RR[0.1, > 3.01]). **Experiment 3** (double training): $b = 0.36$, $SE = 0.14$, $pMCMC = .008$, *CI*s [0.07, 0.64], $BF_{(0,0.65)} = 9.86$ (RR[0.07, 2.5]). Thus all three studies yielded strong evidence for preemption.

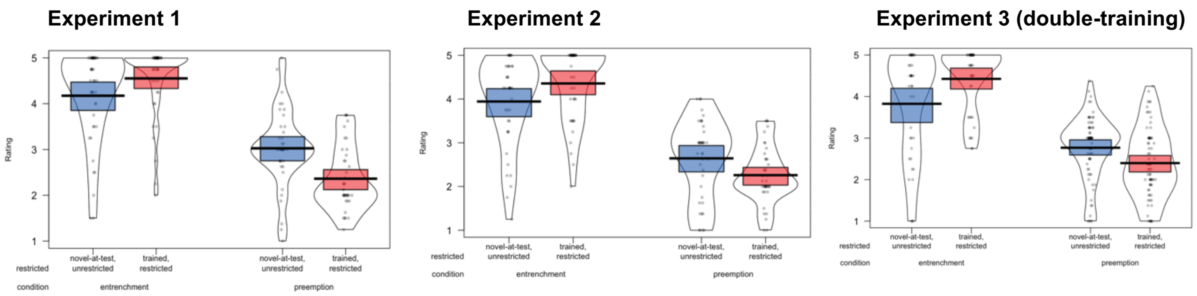

Figure 3 Participants' mean acceptability rating for (semantically correct) trials featuring unattested verb+particle combinations for (red) restricted verbs and (blue) novel-at-test verbs across **Experiments 1-3.** (5= Most acceptable, 1= Least acceptable). Dark black bars represent overall condition means, coloured bands denote 95% within-subjects confidence intervals around the means, and white beans represent the density of the distribution.

Research question 2. Does entrenchment constrain verb argument structure generalizations in adults ?

As can be seen in Figure 3 (left side of each panel), across experiments 1-3, participants in the entrenchment condition – counter to the pre-registered prediction of the entrenchment hypothesis – rated unattested verb+particle combinations as (numerically) *more* acceptable for restricted then novel-at-test verbs, with Bayes Factors of approximately 0.24, 0.25 and 0.23 across Experiments 1, 2 and 3 respectively. Thus, in each case, the data are approximately four times more likely under the null as H_1. Indeed, although the absence of a suitable prior precludes the calculation of a Bayes Factor for this unpredicted effect, the *pMCMC* values and credible intervals for the main effect of verb type (restricted, novel) are, in each case, consistent with a reliable effect in the opposite direction that predicted by the entrenchment hypothesis: **Experiment 1:** b = -0.36, SE = 0.17, $pMCMC$ = .985, *CI*s [-0.69, -0.04], $BF_{0,0.19.}$ = (RR[0.15, inf]). **Experiment 2**: b = -0.40, SE = 0.17, $pMCMC$ = .990, *CI*s [-0.73, -0.06], $BF_{(0,0.19)}$ = 0.25 (RR[0, inf]). **Experiment 3**: b = -0.57, SE = 0.19, $pMCMC$ = .999, *CI*s [-0.94, -0.20], $BF_{0.0.19}$ = 0.23 (RR[0, inf]). Thus, noted by an anonymous reviewer, “it could be that adults find it easier to extend familiar verbs than entirely new verbs”. Certainly, the findings of Experiments 1-3 constitute positive evidence against the possibility that entrenchment constrains verb argument structure generalizations in adults.

Research question 3. Is the effect of statistical preemption larger than entrenchment ?

Given that, across Experiments 1-3, participants showed the predicted effect of preemption but not entrenchment (and indeed, the data is consistent with an effect in the opposite direction), it is no surprise that each experiment showed a significant interaction between condition (preemption, entrenchment) and verb type (restricted, novel), such that the effect of preemption is larger than the effect of entrenchment. **Experiment 1**: b = 1.00, SE = 0.23, $pMCMC$ < .001, *CI*s [0.54, 1.44], $BF_{(0,1.455)}$ = 3157.35 (RR[0.06, > 4]). **Experiment 2**: b = 0.77, SE = 0.25, $pMCMC$ < .001, *CI*s [0.27, 1.25], $BF_{(0,1)}$ = 68.77 RR[0.1, > 4]. **Experiment 3**: b = 0.92, SE = 0.24, $pMCMC$ < .001, *CI*s [0.45, 1.40], associated with a $BF_{(0,1)}$ of 1055.52 (RR[0.07, > 4]).

Discussion (Experiments 1-3)

Across Experiments 1-3, learners showed clear evidence of preemption but not entrenchment[7]. Indeed, unexpectedly, the data were consistent with an effect in the opposite direction to that predicted by entrenchment[8], possibly suggesting that familiar verbs are generalized more easily to new constructions than are newly-learned verbs (e.g., Harmon & Kapatsinski, 2017). The fact that this pattern held across Experiment 3, which included twice the amount of training, suggests that the failure to find the effect predicted by the entrenchment hypothesis was not due to insufficient training. Neither was there any evidence for more entrenchment in Experiment 3 than Experiment 2. To sum up, at least in the domain of (novel) verb argument structure, Experiments 1-3 yielded strong evidence that adult learners constrain their generalizations via preemption, but not entrenchment. The next step is to investigate whether the same is true (a) for other types of linguistic generalization, and (b) for children. We therefore designed an analogous artificial language learning study in the domain of noun plural marking; a domain chosen because the semantics of singular versus plural nouns are simpler for children than those of causal versus non-causal events. Before proceeding to study children (Experiment 5), we first investigated whether the evidence for preemption – and for the zero effect of entrenchment – observed for Experiments 1-3 would generalize to this new domain for adults (Experiment 4).

Experiments 4-5: Noun studies

Given the need to create a more child-friendly artificial language, we designed a semi-artificial language (Wonnacott 2011; Feher, Ritt & Smith 2019) such that all novel nouns were based on familiar English onomatopoeia: *squeako* (‘mouse’), *oinko* (‘pig’), *moo-o* (‘cow’), *purro* (‘cat’), and *woofo* (‘dog’). For participants in the preemption condition, both particles (*bup* and *kem*) denoted plural. For participants in the entrenchment condition, *bup* indicated plural and *kem* singular (or vice versa, depending on counterbalance version); see

Appendix A. Analogous to Experiments 1-3, for all participants, one trained noun appeared with *bup* only, one with *kem* only, and one with both, with a novel noun introduced for the first time during the test session.

Method

Participants

The final samples consisted of *N*=80 adults (Experiment 4) and *N*=80 children (Experiment 5); in both cases, half allocated to the entrenchment condition, half to the preemption condition. The children comprised 37 monolingual English-speakers and 43 English/Hindi bilingual speakers, all aged 5;6-6;6, recruited from schools in England and India, respectively. Further details, including criteria for eligibility and optional stopping, as well as procedures surrounding recruitment, consent and compensation, were identical to those for Experiments 1-3 (see supplementary online material).

Design, materials and procedure

Experiments 4-5 used novel nouns based on familiar English onomatopoeia: *squeako* ('mouse'), *oinko* ('pig'), *moo-o* ('cow'), *purro* ('cat'), and *woofo* ('dog'). Particles were 'kem' and 'bup' (replacing 'gos' in Experiments 1-3) because they were the least frequently mispronounced monosyllabic nonwords used in a previous study (Samara et al, 2017). For participants in the preemption condition both particles (*bup* and *kem*) denoted plural. For participants in the entrenchment condition, *bup* indicated plural and *kem* singular (or vice versa, depending on counterbalance version). Accordingly, rather than video animations (Experiments 1-3) we used statistic images depicting plural or singular scenes, each presented for 1000 ms (see Figure 4). Singular scenes featured a single cartoon animal in the middle of the screen; plural scenes featured multiple cartoon animals (12/13/14/15 animals during training trials and 16/17/18/19 animals in test trials). Their location on screen was random, and the exact number of animals (for plural trials) was distributed randomly (to ensure that the particle indicated plurality, rather than a particular number.

For each participant, three out of the five novel nouns acted as training nouns (with one reserved for the judgment test). Each utterance presented during training consisted of a noun followed always by one of the particles (e.g*., squeako bup*; '[there is] one mouse'), accompanied by a relevant picture (e.g., a single mouse). Analagous to Experiments 1-3, one noun appeared 56 times with particle-a only (e.g., *squeako bup*), one noun appeared 56 times with particle-b only (e.g., *woofo kem*), and one noun appeared 28 times with particle-a (e.g., *oinko bup*) and 28 times with particle-b (e.g., *oinko kem*). Training was divided across 14 blocks – 7 "copy-only" blocks alternating with 7 "training-with-recast" blocks) – which worked in the same way as for Experiments 1-3. Crucially, just an Experiments 1-3, all participants heard **identical utterances**; what differs between conditions is the nature of the evidence that participants can use to learn these restrictions. That is, the conditions differed as to whether *bup* and *kem* both compete for the same, plural meaning[9] – as in the *preemption* condition – or whether one has a singular meaning and the other a plural meaning, as in the *entrenchment* condition. This maps onto the differing assumptions that ungrammatical forms such as **The clown laughed the man* are blocked **only** by appearances of *laugh* in a construction with the same meaning (e.g., *The clown made the man laugh*) – preemption – or **also** by appearances of *laugh* in a construction with a different meaning (e.g., *The man laughed*) – entrenchment.

Test trials were conducted in the same way as for Experiments 1-3, with one novel-at-test noun introduced in the judgment test (e.g., *moo-o*, 'cow'). However, unlike in

Experiments 1-3, the judgement test included only semantically appropriate trials (i.e., participants in the entrenchment condition never saw several animals and heard the singular particle, or vice versa). (Though recall that, in Experiments 1-3, semantically inappropriate trials were excluded from the main analyses and used to check only if participants in the entrenchment condition had learned the particle semantics). The elimination of semantically-inappropriate judgment trials in Experiment 3 allowed us to increase the number of semantically appropriate judgment trials from 2 per cell (Experiments 1-3) to 3 per cell (Experiments 4-5). Thus, this change reduced the overall length of the judgment task from 32 to 24 trials (making it more child friendly), while at the same time increasing the amount of data available for the main analysis.

Other than the use of different materials, the procedures were the same as for Experiments 1-3: Exposure to Freddie's language (language training) was followed by the judgment test, and finally the baseline vocabulary test. For adults, the experiment was run online in a single session lasting approximately 40 minutes. Children completed the experiment over 3 sessions (delivered over consecutive days). Days 1 and 2 consisted of training only (6 blocks; 10-15 minutes in total each day); Day 3 consisted of the final two blocks of training followed by the production, judgment, and baseline vocabulary tests (20-25 minutes in total). All UK-based children were tested face-to-face, while all India-based children were tested online via Zoom, due to Covid-19 restrictions.

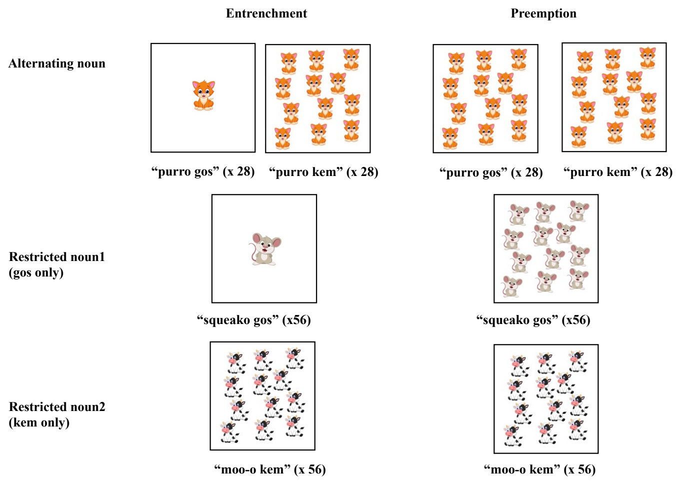

Figure 4. Snapshots of picture materials and accompanying sentences used during training in the preemption and entrenchment conditions of Experiment 4.

Results

All aspects of the Bayesian MLM analyses were identica l to previous experiments (where verb → noun), with priors drawn from Experiments 1-3 (see online supplementary material for details).

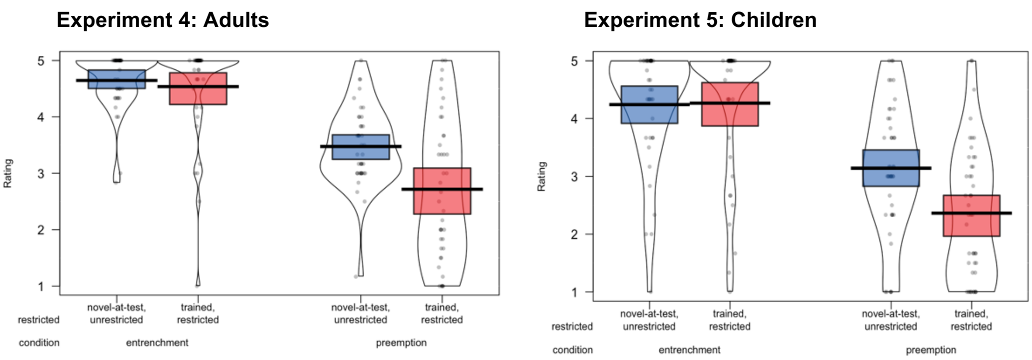

Figure 5. Participants' mean acceptability rating for semantically correct trials featuring unattested restricted noun forms versus unattested novel noun forms across Experiment 4 (adults) and Experiment 5 (children) (5= Most acceptable, 1= Least acceptable). Dark black bars represent overall condition means, coloured bands denote 95% within-subjects confidence intervals around the means, and white beans represent the density of the distribution

Research question 1. Does statistical preemption constrain morphological noun generalizations?

Figure 5 (left side of each panel) shows preemption participants' acceptability ratings for unattested novel+particle combinations for (red) restricted nouns that had been presented with either particle-a or particle-b (but never both) during training and (blue) novel-at-test nouns that had *never* been presented during training. Visual inspection of this figure suggests that, for both adults (Experiment 4) and children (Experiment 5), the data are consistent with our pre-registered prediction: if statistical preemption constrains noun argument structure generalizations, participants will rate unattested noun+particle combinations as less acceptable for restricted (blue) than novel-at-test and hence unrestricted nouns (red). As for Experiments 1-3, this is a very strong test, since – given that the novel-at-test nouns had not been presented during training – *all* noun +particle combinations were previously unwitnessed; thus, any effect cannot be a mere familiarity preference. Rather it must reflect an inference that *some* unwitnessed noun+particle combinations are unacceptable (i.e., for nouns that have been trained with a competing particle) and some are acceptable (i.e., for noun that have not been trained at all).

Indeed, as predicted by the preemption hypothesis, Bayesian mixed-effects model showed reliable effects of noun type (restricted, novel-at-test), for both adults, **Experiment 4:** $b = 0.72$, $SE = 0.22$, $pMCMC < .001$, *CI*s [0.29, 1.14], $BF_{(0,0.65)} = 78.22$ RR[0.13, > 4] and children, **Experiment 5**: $b = 0.75$, $SE = 0.21$, $pMCMC < .001$, *CI*s [0.34, 1.14], and the $BF_{(0, 0.300)}$ was 82.43 (RR[0.07, > 4]). Although we did not preregister any predictions in this regard, it is interesting to note that adults and children showed a very similar magnitude of effect.

Research question 2. Does entrenchment constrain morphological noun generalizations

Inspection of Figure 5 (left side of each panel) suggests very little difference between acceptability ratings for unattested noun+particle combinations with restricted than novel-at-test nouns. Indeed, the Bayesian models showed that the effect of noun type (restricted, novel-at-test) was inconclusive, for both adults, **Experiment 4:** $b = 0.11$, $SE = 0.14$, $pMCMC = 226$, *CI*s [-0.18, 0.39], $BF_{(0,0.19)} = 1.07$ (RR[0.01, 0.88]) and children, **Experiment 5**: $b = -0.02$, $SE = 0.22$, $pMCMC = .552$, *CI*s [-0.44, 0.42]. The $BF_{(0,0.173)}$ was 0.737 (RR[0.01, 0.54]). For children, the numerical effect is (very narrowly) in the opposite direction to that predicted. For adults, supplementary analyses (see the R Markdown file at the OSF project site) suggest that, assuming that the error term would reduce in proportion to $\sqrt{SE}$), over 200 participants would be required to establish evidence for entrenchment (i.e., for H_1 over H_o). Thus while, unlike Experiments 1-3, these findings do not provide strong evidence against entrenchment, they do demonstrate that any such effect is likely to be too small to be meaningful. Indeed, for children, note that the robustness region for entrenchment – which tells us the range of estimates of H_1 we could have used and still seen ambiguous evidence – *excludes* the estimate of the size of the preemption effect shown by children (0.75). Thus, if we had hypothesized an entrenchment effect of the same size as the observed preemption effect, we *would* have seen substantial evidence against an effect of entrenchment for children.

Research question 3. Is the effect of statistical preemption larger than entrenchment ?

As for Experiments 1-3, the Bayesian models yielded the predicted interaction between noun-type (restricted, novel-at-test) and condition (preemption, entrenchment), for both adults, **Experiment 4:** $b = 0.57$, $SE = 0.25$, $pMCMC = .01$, *CI*s [0.08, 1.06], $BF_{(0,1)} = 5.52$ ([RR [0.15, 2.12]) and children, **Experiment 5:** $b = 0.73$, $SE = 0.29$, $pMCMC = .007$, *CI*s [0.16, 1.30], associated with a $BF_{(0,0,523)}$ of 10.79 (RR[0.15, >4]). Thus for both age groups, the effect of preemption was larger than the (effectively zero) effect of entrenchment.

This interaction demonstrates that participants' dispreference for restricted noun+particle combinations as compared to unrestricted (novel-at-test) noun+particle combinations was larger in the preemption than entrenchment condition. Thus, just as in Experiments 1-3, the effect of preemption was larger than the (close to zero) effect of preemption.

Discussion (Experiments 4-5)

Having translated the methodological approach of Experiments 1-3 (verb argument structure generalizations) into the domain of morphological noun plural marking and run this study with adults (Experiment 4), we investigated the crucial question of how *children* restrict their linguistic generalizations (Experiment 5). In fact, despite the differences in age and (c.f., Experiments 1-3) domain, the findings were very similar: Preemption (a) constrained learners' generalizations and (b) did so to a larger extent than entrenchment. One small difference across studies is that while Experiments 1-3 found positive Bayesian evidence *against* entrenchment, Experiments 4-5 simply found no positive evidence *for* entrenchment (i.e., the Bayes Factor was inconclusive). However, recall from the robustness-region analysis that, for Experiment 5 (children), we do have strong Bayesian evidence against the possibility of an entrenchment effect as large as a preemption effect.

General Discussion

The present article has reported five artificial language learning studies designed to address the question of how learners restrict their linguistic generalizations (e.g., **The clown laughed the man*), while retaining the ability to produce grammatically acceptable novel utterances. Under ***statistical preemption***, the use of a verb in a particular construction is probabilistically blocked by the use of that verb in another construction ONLY if those two constructions have very similar meanings. For example, **The clown laughed the man* would be probabilistically blocked by *Sue made John laugh*, but not by *John laughed*. Under ***entrenchment***, errors such as **The clown laughed the man* are probabilistically blocked by the occurrence in the input of ANY utterance that includes the verb *laugh* (e.g., by BOTH *Sue made John laugh* and *John laughed*). Across the five studies reported above, we designed a training regime such that learners received in-principle evidence for the ungrammaticality of a particular unattested verb/noun+particle combination via either **preemption only** or **entrenchment only**.

In each study, we tested the key prediction of the preemption/entrenchment hypotheses: Participants will rate unattested verb/noun+particle combinations as less acceptable for **restricted** verbs/nouns, which appeared during training, than for **unrestricted**, novel-at-test verbs/nouns, which did not appear at all during training. This is a very strong test since it cannot reflect a mere familiarity preference: None of the crucial verb/noun+particle combinations rated at test had *ever* been presented during training. Thus, any effect of preemption/entrenchment must reflect an inference that *some* unwitnessed verb/noun+particle combinations are unacceptable (i.e., those for which preemption/entrenchment has occurred) and some are acceptable (i.e., for verbs/nouns that have not been trained at all).

Despite some minor differences in results the same overall pattern was observed across all five studies: substantial/strong Bayesian evidence for preemption but not entrenchment (and, in some cases, substantial Bayesian evidence *against* entrenchment). In a companion paper (Ambridge et al, in prep.), we show that very similar findings are observed using a production-based task, and present a simple discriminative-learning model (adapted from the model in Ramscar et al, 2013) that can account for these findings (and that, in turn, makes new predictions that are confirmed by further analyses of the present findings). Space precludes discussion of the model here, but essentially it simulates the present findings by instantiating competition between lexical items and particle semantics to predict the occurrence of the relevant particle (e.g., *gos/kem*) on each trial. Even setting aside the modeling, since the observed pattern was robust across five studies in different domains (verb argument structure/noun morphology) and with different populations (adults/children), we feel confident in our conclusion that preemption, and not entrenchment, is the means – or at least the major means – by which speakers restrict their linguistic generalizations.

To loop back to the natural language examples discussed at the start of the paper, the present findings suggest that learners infer the ungrammaticality of forms such as **sitted, *mans, *unsqueeze, *the asleep boy and *the clown laughed the man* because of the competition that arises when there are competing forms such as *sat, men, let go, the boy who's asleep* and *the clown made the man laugh*[10] which express the same message (i.e., via preemption); rather than by repeatedly hearing various forms of *sit, man, squeeze, asleep* and *laugh* in the absence of the unwitnessed forms (i.e., via entrenchment). Indeed, when focussing squarely on the domain of inflectional morphology (*sat/*sitted, men/*mans*), this conclusion already is uncontroversial, since speakers of highly-inflected languages must be able to extend inflectional morphemes marking (amongst other things) person, number and case to unwitnessed verbs and nouns. Yet, as we noted in the Introduction, in the domain of syntax, studies including Stefanowitsch (2008) and Ambridge et al., (2015, 2015) provided apparent evidence that entrenchment outperforms preemption. In contrast, the present findings in our view constitute evidence for preemption over entrenchment in the domains of both morphology (Studies 4-5) and syntax (Studies 1-3). This is because while the markers *gos* and *kem* could be analysed an inflectional morphemes, they are "morphosyntactic" in that they perform functions that, in many languages are performed by syntax (e.g., intransitive / transitive-causative / periphrastic causative in English).

A potential objection to our conclusion of an effect of preemption but not entrenchment is that the studies reported here used artificial languages, and that the findings therefore cannot be extrapolated to real language learning. In response to this objection, we would first note that, although the languages taught were artificial rather than natural, they were certainly *naturalistic*. Languages in which causality is marked by morphological suffixes on the verb (e.g., *chila+gos*; Experiments 1-3) are common globally; indeed, languages that (like English) lack such morphology are in the minority (Haspelmath, 1993; Shibatani & Pardeshi, 2003). The same could be said for languages that use morphological markers or particles to mark singular versus plural, as in Experiments 4-5. For example, like the languages used in these studies, German makes use of a variety of different singular (*der/die/das*) and plural markers (*[e]n/-e/[e]r/s*), the choice of which depends on the noun. n summary, then, although the linguistic systems we trained happened to be artificial, they do not differ in any important sense from naturally-occurring systems that we could equally well have trained instead The fact that our findings were remarkably consistent across each of these domains lends weight to the generality of our findings and suggests that they do not result from a methdological quirk.

Of course, as in any artificial-language-learning study, the novel language taught to participants was a highly simplified and idealized language that included (across the training and test sessions) only four lexical items (verbs/nouns, depending on the study) and two particle-based constructions (e.g., *kem/gos*). Ideally, we would have been able to include more lexical items, and more constructions, which would have made the scenario closer to real-world language-learning, and allowed for further, even more stringent, tests of the preemption and entrenchment hypotheses (e.g., by having multiple items per construction, that occur with different frequencies). This was not possible in the present studies, since even the simple languages we created were at the limit of learnability in the time available (indeed, around 15-20 participants per study were excluded and replaced for failing to learn one or more aspect of the novel language; see the online supplementary materials for details). Developing a paradigm that allows for the training of larger, more naturalistic languages – perhaps over a much longer time-period – is an important methodological goal for future research.

At the same time, if entrenchment really were an important mechanism in language acquisition, we would have expected to see at least some minimal evidence of entrenchment even in the present setup, as opposed to (in Experiments 1-3) Bayesian evidence against the effect (and, indeed, potentially an effect in the opposite direction to that predicted by entrenchment). Likewise, we cannot see any reason to believe that the preemption effects observed were somehow mere artifacts of the particular artificial-language set-up.

A second concern could be that the fact that our experiments are conducted with older learners limits their relevance to first language acquisition. To some extent, the fact that the effect holds equally for 5-6 year-olds mitigates against this concern; for example, it makes it unlikely that the results are a consequence of adult explicit strategy use. Still, we acknowledge that even our child participants come to the task with considerable prior knowledge of an existing language (English or Hindi), including knowledge of causal and plural marking, and that, if the methodological challenges could be overcome, studies with novice infant language learners could in principle allow for even stronger conclusions.

In the meantime, the findings of the present studies suggest that preemption and not entrenchment provides the best description of how human learners restrict their linguistic generalizations. These findings do not address the question of exactly how preemption effects should be integrated into an overall model of language learning. It could be the case, for example, that effects of preemption occur naturally as a consequence of underlying mechanisms comprising simple learning principles of error-driven learning and cue competition (e.g., Ramscar, Dye & Klein, 2013). Such effects could also, in principle, fall naturally out of the types of learning instantiated in state-of-the-art large language models (e.g., Ambridge & Blything, submitted). It is also important to emphasize that identifying a key role for preemption does not provide a full account of the factors impacting on the balance between generalization and learning of lexical restrictions. As noted in the Introduction, there is copious evidence from natural languages that generalization is also affected by the fit between verb and construction semantics. Artificial language studies also suggest that the extent of generalization may be affected by the distribution of lexical items across structures, with the balance between generalization and conservatism affected for example by skew (Casenhiser, & Goldberg, (2005), Wonnacott Brown & Nation (2017) and type-frequency (Gomez, 2002; Wonnacott, Boyd & Goldberg 2011).

Setting aside these wider considerations, what the present studies have clearly shown is that whether we look at verbs or nouns, adults or children, speakers use preemption and not entrenchment to restrict their linguistic generalizations. This pattern of preemption without entrenchment is now – in our view – sufficiently well established that it constitutes a finding that any successful theory of language acquisition must be able to explain.

References

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*(6), 952-965.

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *Bmj*, *311*(7003), 485.

Ambridge, B. (2013). How do children restrict their linguistic generalizations?: an (un-)grammaticality judgment study. *Cognitive Science, 37*(3), 508-543.

Ambridge, B. & Ambridge, C. (2020). The retreat from transitive-causative overgeneralization errors: A review and diary study. In Rowland, C., Theakston, A., Ambridge, B., & Twomey, K. (Eds). *Current Perspectives on Child Language Acquisition: How children use their environment to learn*. John Benjamins. (pp.113`-130). https://doi.org/10.1075/tilar.27.05amb

Ambridge, B., & Blything, R. P. (2016). A connectionist model of the retreat from noun argument structure overgeneralization. *Journal of Child Language, 43(6)*, 1245-1276. https://doi.org/10.1017/S0305000915000586

Ambridge, B., Barak, L., Wonnacott, E., Bannard, C., & Sala, G. (2018). Effects of both preemption and entrenchment in the retreat from verb overgeneralization errors: Four reanalyses, an extended replication, and a meta-Analytic synthesis. *Collabra: Psychology, 4*(1), 23. DOI: http://doi.org/10.1525/collabra.133

Ambridge, B., Doherty, L., Maitreyee, R., Tatsumi, T., Zicherman, S., Mateo-Pedro, P., Kawakami, A., Bidgood, A., Pye, C., Narasimhan, B., Arnon, I., Bekman, D., Efrati, A., Pixabaj, S.F.C, Peliz, M.M., & Mendoza, M.J, Samanta, S., Campbell, S., McCauley, S., Berman, R., Sharma, D.M., Nair, R.B. & Fukumura, K (2022). Testing a computational model of causative overgeneralizations: Child judgment and production data from English, Hebrew, Hindi, Japanese and Kiche, *Open Research Europe, 1*(1) https://doi.org/10.12688/openreseurope.13008.1

Ambridge, B., Maitreyee, R., Tatsumi, T., Doherty, L., Zicherman, S., Mateo-Pedro, P., Bannard, C., Samanta, S., McCauley, S., Arnon, I., Bekman, D., Efrati, A., Berman, R. Narasimhan, B., Sharma, D.M., Nair, R.B., Fukumura, K., Campbell, S., Pye, C., Pixabaj, S.F.C, Peliz, M.M., & Mendoza, M.J., (2020). The Crosslinguistic acquisition of causative sentence structure: Computational modeling and grammaticality judgments from adult and child speakers of English, Japanese, Hindi, Hebrew and Kiche, *Cognition, 202*,104310. https://doi.org/10.1016/j.cognition.2020.104310

Ambridge, B., Pine, J. M, Rowland, C. F., Freudenthal, D., & Chang, F. (2014). Avoiding dative overgeneralization errors: semantics, statistics or both? *Language, Cognition and Neuroscience, 29(2)*, 218-243. https://doi.org/10.1080/01690965.2012.738300

Ambridge, B., Pine, J. M., Rowland, C. F., Jones, R.L, & Clark, V. (2009). A semantics-based approach to the 'no negative-evidence' problem *Cognitive Science*, 33(7), 1301-1316.

Ambridge, B., Pine, J.M. & Rowland, C.F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition, 123*(2) 260–279.

Ambridge, B., Pine, J.M. & Rowland, C.F., (2011). Children use verb semantics to retreat from overgeneralization errors: A novel verb grammaticality judgment study. *Cognitive Linguistics, 22*(2), 303–323.

Ambridge, B., Pine, J.M., Rowland, C.F. & Young, C.R. (2008) The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgements of argument-structure overgeneralization errors. *Cognition, 106*(1), 87-129.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52(1)*, 388-407. https://doi.org/10.3758/s13428-019-01237-x

Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2015). ndl: Naive discriminative learning. https://cran.r-project.org/web/packages/ndl/ndl.pdf

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, M., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bidgood A., Pine, J.M., Rowland, C.F., Sala, G., Freudenthal, D.T. & Ambridge, B. (2021). Verb argument structure overgeneralisations for the English intransitive and transitive constructions: Grammaticality judgments, production priming and a meta-analytic synthesis. *Language and Cognition, 13* (3), 397-437. https://doi.org/10.1017/langcog.2021.8

Bidgood, A., Ambridge, B., Pine, J.M. & Rowland, C.F. (2014). The retreat from locative overgeneralisation errors: A novel verb grammaticality judgment study. *PLoS ONE, 9*(5), e97634.

Blything, R.P. Ambridge, B., & Lieven, E.V.M. (2014). Children use statistics and semantics in the retreat from overgeneralization. *PLoS ONE, 9*(10) e110009.

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

Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, *87*(1), 55-83.

Braine, M. D. S., & Brooks, P. J. (1995). Verb argument strucure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: young children's acquisition of verbs* (pp. 352-376). Hillsdale, NJ: Erlbaum.

Brooks, P. J., & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*(4), 720-738.

Brooks, P. J., & Zizak, O. (2002). Does preemption help children learn verb transitivity? *Journal of Child Language, 29*, 759-781.

Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development, 70*(6), 1325-1337.

Bükner, P. C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1-28.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8*(6), 500–508. https://doi.org/10.1111/j.1467-7687.2005.00441.x

Chomsky, N. (1957). Syntactic Structure. The Hague: Mouton.

Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language aquisition* (pp. 1–33). Lawrence Erlbaum Associates, Inc.

Dienes, Z. (2008). Understanding psychology as a science: An introduction to scientific and statistical inference. Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Dienes, Z. (2015). How Bayesian statistics are needed to determine whether mental states are unconscious. In M. Overgaard (Ed.), *Behavioural Methods in Consciousness Research* (pp. 199-220). Oxford: Oxford University Press.
discriminative learning. R package version 0.2.17. https://CRAN.R-project.org/package=ndl.

Dienes Z. (2019) How do I know what my theory predicts? *Advances in Methods and Practices in Psychological Science*, 2(4), 364-377.

Dienes, Z. (2023). Testing theories with Bayes factors. In Austin Lee Nichols & John E. Edlund (Eds), *Cambridge Handbook of Research Methods and Statistics for the Social and Behavioral Sciences, Volume 1*: Building a program of research, pp 494-512. Cambridge University Press.

Goldberg, A. E. (2011). Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics*, *22*(1), 131-153.

Goldberg, A., E. (2019). *Explain Me This*: *Creativity, Competition, and the Partial Productivity of Constructions.* Princeton University Press.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22-44.

Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. In B. Comrie and M. Polinsky (Eds). *Causatives and Transitivity.* John Benjamins, (pp. 87-120).

Jeffreys, H. (1961). *Theory of Probability.* Oxford University Press.

Kamin, L. J. (1968). “Attention-like” processes in classical conditioning. Miami symposium on the prediction of behavior: Aversive stimulation, 9–31.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. Punishment and Aversive Behavior.

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist, 73*(1), 1–2. https://doi.org/10.1037/amp0000263

Köpcke, K. M. (1998). The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language*, *25*(2), 293-319.

Li, P., & MacWhinney, B. (1996). Cryptotype, overgeneralization and competition: A connectionist model of the learning of English reversive prefixes. *Connection Science*, *8*(1), 3-30.

Perek, F., & Goldberg, A. E. (2015). Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language*, *84*, 108-127.

Perek, F., & Goldberg, A. E. (2017). Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition*, *168*, 276-293.

Pinker, S. (1989). *Learnability and Cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

R Core Team (2022). R: A language and environment for statistical computing.

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*(6), 1017-1023.

Robenalt, C., & Goldberg, A. E. (2015). Judgment evidence for statistical preemption: It is relatively better to vanish than to disappear a rabbit, but a lifeguard can equally well backstroke or swim children to shore. *Cognitive Linguistics*, *26*(3), 467-503.

Robenalt, C., & Goldberg, A. E. (2016). Nonnative speakers do not take competing alternative expressions into account the way native speakers do. *Language Learning*, *66*(1), 60-93.

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review, 21*, 301-308. Doi: 10.3758/s13423-014-0595-4

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology, 94*, 85-114. https://doi.org/10.1016/j.cogpsych.2017.02.004.

Shibatani, M. & Pardeshi, P. (2002). The causative continuum. In *The Grammar of Causation and Interpersonal Manipulation*, Masayoshi Shibatani (ed.), pp. 85–126. Amsterdam: John Benjamins.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the Rescorla–Wagner model. Psychonomic Bulletin & Review, 3, 314–321.

Silvey, C., Dienes, Z., & Wonnacott, E. (2021). Bayes factors for mixed-effects models. PsyArXiv. https://doi.org/10.31234/osf.io/m4hju

Stefanowitsch, A. (2008). Negative evidence and preemption: A constructional approach to ungrammaticality. *Cognitive Linguistics, 19*(3), 513-531.

Theakston, A. L. (2004). The role of entrenchment in children's and adults' performance on grammaticality judgement tasks. *Cognitive Development, 19*(1), 15-34.

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language, 65(1)*, 1-14. https://doi.org/10.1016/j.jml.2011.03.001

Wonnacott, E., Boyd, J.K, Thomson. J.J., & Goldberg, A.E. (2012) Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of Memory and Language, 66*, 458-478.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing noun argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56(3)*, 165-209. https://doi.org/10.1016/j.jml.2011.03.001.

APPENDIX: Study Schematics

Experiments 1-3: Verb argument structure. For the particular counterbalance condition shown, *chila*='bounce', *tombat*='roll', *coomo*='drop', *panjol* (novel at judgment test) = 'spin', roosa (novel at production test) = 'slide'. For the entrenchment condition, *gos*=noncausal, *kem*=causal (for the preemption condition, both indicate causal).

Training (greyed out rows are never presented)					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted verb 1 – gos only	chila gos x 64	Someone bounced the ball	gos=causal a	The ball bounced	gos=noncausal
	*chila kem x 0	Someone bounced the ball	kem=causal b	Someone bounced the ball	kem=causal
Restricted verb 2 – kem only	*tombat gos x 0	Someone rolled the ball	gos=causal a	The ball rolled	gos=noncausal
	tombat kem x 64	Someone rolled the ball	kem=causal b	Someone rolled the ball	kem=causal
Alternating verb –gos+kem	coomo gos x 32	Someone dropped the ball	gos=causal a	The ball dropped	gos=noncausal
	coomo kem x 32	Someone dropped the ball	kem=causal b	Someone dropped the ball	kem=causal

Grammaticality judgment test (Bold rows relate to key predictions; Grey rows are excluded from the statistical analyses)[1]					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted verb 1 – gos only	chila gos x 2	Someone bounced the ball	gos=causal a	The ball bounced	gos=noncausal
	***chila kem x 2**	**Someone bounced the ball**	**kem=causal b**	**Someone bounced the ball**	**kem=causal**
Restricted verb 2 – kem only	***tombat gos x 2**	**Someone rolled the ball**	**gos=causal a**	**The ball rolled**	**gos=noncausal**
	tombat kem x 2	Someone rolled the ball	kem=causal b	Someone rolled the ball	kem=causal
Alternating verb –gos+kem	coomo gos x 2	Someone dropped the ball	gos=causal a	The ball dropped	gos=noncausal
	coomo kem x 2	Someone dropped the ball	kem=causal b	Someone dropped the ball	kem=causal
Novel-at-test verb	**panjol gos x 2**	**Someone spun the ball**	**gos=causal a**	**The ball spun**	**gos=noncausal**
	panjol kem x 2	**Someone spun the ball**	**kem=causal b**	**Someone spun the ball**	**kem=causal**

Key prediction for both preemption (Research Q1) and entrenchment (Research Q2): unattested verb+particle combinations for restricted verbs (*chila kem; *tombat gos) will be rated as less acceptable than unattested verb+particle combinations for novel-at-test verbs (panjol kem, panjol gos). Note that all four verb+particle combinations (*chila kem, tombat gos, panjol kem, panjol gos) are previously unwitnessed. (Research Q3) Is any such effect larger for preemption than entrenchment?

Secondary prediction (Appendix C, E, G) for both preemption and entrenchment: Restricted verbs yield more acceptable utterances when combined with the attested than unattested particle (i.e., [chila gos, tombat kem] > [*tombat gos *chila kem])

(Secondary) production test (Experimenter produces verb – chila/tombat/coomo/roosa – only)[2]					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted verb 1 – gos only	chila (gos) x 4	Someone bounced the ball	gos=causal a	The ball bounced	gos=noncausal
	***chila (kem) x 4**	**Someone bounced the ball**	**kem=causal b**	**Someone bounced the ball**	**kem=causal**
Restricted verb 2 – kem only	***tombat (gos) x 4**	**Someone rolled the ball**	**gos=causal a**	**The ball rolled**	**gos=noncausal**
	tombat (kem) x 4	Someone rolled the ball	kem=causal b	Someone rolled the ball	kem=causal
Alternating verb –gos+kem	coomo (gos) x 4	Someone dropped the ball	gos=causal a	The ball dropped	gos=noncausal
	***coomo (kem) x 4**	**Someone dropped the ball**	**kem=causal b**	**Someone dropped the ball**	**kem=causal**
Novel-at-test verb	roosa (gos) x 4	Someone slid the ball	gos=causal a	The ball slid	gos=noncausal
	***roosa (kem) x 4**	**Someone slid the ball**	**kem=causal b**	**Someone slid the ball**	**kem=causal**

The x4 indicates the number of responses of each type we would expect to see if participants were choosing gos/kem solely on the basis of the scene (entrenchment condition) or randomly with equal probability (preemption condition); corresponding to the comparisons against chance reported in the relevant analyses (see Appendix B, D, F).

Predictions for both preemption and entrenchment (note that for alternating and novel-at-test verbs, the designation of one marker (indicated by an asterisk) as "unattested" is arbitrary (see Appendix B, D, F for explanation).
Model 1: Participants will produce a lower proportion of unattested (vs attested) verb+particle combinations for restricted verbs than for novel-at-test verbs; i.e., [chila kem (vs gos), tombat gos (vs kem)] < [roosa kem (vs gos)].
Model 2: Participants will produce a lower proportion of unattested (vs attested) verb+particle combinations for restricted verbs than for alternating verbs; i.e., [chila kem (vs gos), tombat gos (vs kem)] < [coomo kem (vs gos)].

[1]Note that this table shows only the 16 judgment trials for which all rated utterances were semantically appropriate for all participants regardless of condition, and hence which were included in the main analyses. The 16 judgment trials featuring utterances that (for participants in the entrenchment condition) were semantically inappropriate are shown separately below (see *Have participants in entrenchment condition learned particle semantics? (Research Q0): Judgment test*).

[2]The experimenter produced 8 trials with each verb (chila/tombat/coomo/roosa), paired with either all causal scenes (preemption condition) or half causal and half noncausal scenes (entrenchment condition), as shown in the relevant columns. The particles (gos/kem) are shown in brackets in the "Utterance" column to indicate that the participant (rather than the experimenter) chooses which particle to produce on a given trial.

Baseline Vocabulary test (greyed out rows show foil animations)					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted verb 1 – gos only	chila gos x 1	Someone bounced the ball Someone rolled the ball Someone dropped the ball	gos=causal a	The ball bounced The ball rolled The ball dropped	gos=noncausal
	*chila kem x 1	Someone bounced the ball Someone rolled the ball Someone dropped the ball	kem=causal b	Someone bounced the ball Someone rolled the ball Someone dropped the ball	kem=causal
Restricted verb 2 – kem only	*tombat gos x 1	Someone bounced the ball Someone rolled the ball Someone dropped the ball	gos=causal a	The ball bounced The ball rolled The ball dropped	gos=noncausal
	tombat kem x 1	Someone bounced the ball Someone rolled the ball Someone dropped the ball	kem=causal b	Someone bounced the ball Someone rolled the ball Someone dropped the ball	kem=causal
Alternating verb –gos+kem	coomo gos x 1	Someone bounced the ball Someone rolled the ball Someone dropped the ball	gos=causal a	The ball bounced The ball rolled The ball dropped	gos=noncausal
	coomo kem x 1	Someone bounced the ball Someone rolled the ball Someone dropped the ball	kem=causal b	Someone bounced the ball Someone rolled the ball Someone dropped the ball	kem=causal
Prediction: If participants have learned the verb vocabulary (here, *chila*='bounce', *tombat*='roll', *coomo*='drop'), they should choose the non-greyed-out option in each cell.					

Have participants in entrenchment condition learned particle semantics? (Research Q0): Production test[3]				
		Preemption condition	**Entrenchment condition**	
Trial Type	**Utterance**		**Meaning**	**Scene**
Alternating verb –gos+kem	coomo (gos) x 4	NA	The ball dropped	gos=noncausal
	coomo (kem) x 4		Someone dropped the ball	kem=causal
Novel-at-test verb	roosa (gos) x 4		The ball slid	gos=noncausal
	roosa (kem) x 4		Someone slid the ball	kem=causal

Have participants in entrenchment condition learned particle semantics? (Research Q0): Judgment test[4]					
		Preemption condition	**Entrenchment condition**		
Trial Type	**Utterance**		**Semantically correct?**	**Meaning**	**Scene**
Restricted verb 1 – gos only	chila gos x 2	NA	YES	The ball bounced	gos=noncausal
	*chila kem x 2		YES	Someone bounced the ball	kem=causal
Restricted verb 2 – kem only	*tombat gos x 2		YES	The ball rolled	gos=noncausal
	tombat kem x 2		YES	Someone rolled the ball	kem=causal
Alternating verb –gos+kem	coomo gos x 2		YES	The ball dropped	gos=noncausal
	coomo kem x 2		YES	Someone dropped the ball	kem=causal
Novel-at-test verb	panjol gos x 2		YES	The ball spun	gos=noncausal
	panjol kem x 2		YES	Someone spun the ball	kem=causal
Restricted verb 1 – gos only	chila gos x 2		NO	Someone bounced the ball	gos=causal
	*chila kem x 2		NO	The ball bounced	kem=noncausal
Restricted verb 2 – kem only	*tombat gos x 2		NO	Someone rolled the ball	gos=causal
	tombat kem x 2		NO	The ball rolled	kem=noncausal
Alternating verb –gos+kem	coomo gos x 2		NO	Someone dropped the ball	gos=causal
	coomo kem x 2		NO	The ball dropped	kem=noncausal
Novel-at-test verb	panjol gos x 2		NO	Someone spun the ball	gos=causal
	panjol kem x 2		NO	The ball spun	kem=noncausal

[3]The experimenter produced 8 trials with each alternatig/novel-at-test verb (coomo/roosa), half paired with noncausal scenes (for which gos would count as a correct production), half paired with casual scenes (for which kem would count as a correct production). The particles (gos/kem) are shown in brackets in the "Utterance" column to indicate that the participant (rather than the experimenter) chooses which particle to produce on a given trial. The numbers (x4) reflect idealized performance (i.e., perfect learning of the semantics) for a given participant.

Experiments 4-5: Noun plural marking. For the particular counterbalance condition shown, *bup*=singular, *kem*=plural, for the entrenchment condition (for the preemption condition, both indicate plural).

Training (greyed out rows are never presented)					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted noun 1 – bup only	squeako bup x 56	Several mice	bup=plural a	One mouse	bup=singular
	*squeako kem x 0	Several mice	kem=plural b	Several mice	kem=plural
Restricted noun 2 – kem only	*woofo bup x 0	Several dogs	bup=plural a	One dog	bup=singular
	woofo kem x 56	Several dogs	kem=plural b	Several dogs	kem=plural
Alternating noun –bup+kem	oinko bup x 28	Several pigs	bup=plural a	One pig	bup=singular
	oinko kem x 28	Several pigs	kem=plural b	Several pigs	kem=plural

Grammaticality judgment test (Bold rows relate to key predictions; Grey rows are excluded from the statistical analyses)[1]					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted noun 1 – bup only	squeako bup x 3	Several mice	bup=plural a	One mouse	bup=singular
	***squeako kem x 3**	**Several mice**	**kem=plural b**	**Several mice**	**kem=plural**
Restricted noun 2 – kem only	***woofo bup x 3**	**Several dogs**	**bup=plural a**	**One dog**	**bup=singular**
	woofo kem x 3	Several dogs	kem=plural b	Several dogs	kem=plural
Alternating noun –bup+kem	oinko bup x 3	Several pigs	bup=plural a	One pig	bup=singular
	oinko kem x 3	Several pigs	kem=plural b	Several pigs	kem=plural
Novel-at-test noun	**moo-o bup x 3**	**Several cows**	**bup=plural a**	**One cow**	**bup=singular**
	moo-o kem x 3	**Several cows**	**kem=plural b**	**Several cows**	**kem=plural**
Key prediction for both preemption (Research Q1) and entrenchment (Research Q2): unattested noun+particle combinations for restricted nouns (*squeako kem; *woofo bup) will be rated as less acceptable than unattested noun+particle combinations for novel-at-test nouns (moo-o kem, moo-o bup). Note that all four noun+particle combinations (*squeako kem, woofo bup, moo-o kem, moo-o bup) are previously unwitnessed. (Research Q3) Is any such effect larger for preemption than entrenchment? Secondary prediction (Appendix C, E, G) for both preemption and entrenchment: Restricted nouns yield more acceptable utterances when combined with the attested than unattested particle (i.e., [squeako bup, woofo kem] > [*woofo bup *squeako kem])					

(Secondary) production test (Experimenter produces noun – squeako/woofo/oinko/purro – only)[2]					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted noun 1 – bup only	squeako (bup) x 4	Several mice	bup=plural a	One mouse	bup=singular
	***squeako (kem) x 4**	**Several mice**	**kem=plural b**	**Several mice**	**kem=plural**
Restricted noun 2 – kem only	***woofo (bup) x 4**	**Several dogs**	**bup=plural a**	**One dog**	**bup=singular**
	woofo (kem) x 4	Several dogs	kem=plural b	Several dogs	kem=plural
Alternating noun –bup+kem	oinko (bup) x 4	Several pigs	bup=plural a	One pig	bup=singular
	***oinko (kem) x 4**	**Several pigs**	**kem=plural b**	**Several pigs**	**kem=plural**
Novel-at-test noun	purro (bup) x 4	Several cats	bup=plural a	One cat	bup=singular
	***purro (kem) x 4**	**Several cats**	**kem=plural b**	**Several cats**	**kem=plural**
The x4 indicates the number of responses of each type we would expect to see if participants were choosing bup/kem solely on the basis of the scene (entrenchment condition) or randomly with equal probability (preemption condition); corresponding to the comparisons against chance reported in the relevant analyses (see Appendix H, J, L). Predictions for both preemption and entrenchment (note that for alternating and novel-at-test nouns, the designation of one marker (indicated by an asterisk) as "unattested" is arbitrary (see Appendix H, J, L for explanation). Model 1: Participants will produce a lower proportion of unattested (vs attested) noun+particle combinations for restricted nouns than for novel-at-test nouns; i.e., [squeako kem (vs bup), woofo bup (vs kem)] < [purro kem (vs bup)]. Model 2: Participants will produce a lower proportion of unattested (vs attested) noun+particle combinations for restricted nouns than for alternating nouns; i.e., [squeako kem (vs bup), woofo bup (vs kem)] < [oinko kem (vs bup)].					

[1]Note that, unlike in Experiments 1-3, we ran only semantically appropriate judgment trials (i.e., participants in the entrenchment never saw several animals and heard the singular determiner, or vice versa). (Though recall that, in Experiments 1-3, semantically inappropriate trials were excluded from the main analyses and used to check only if participants in the entrenchment condition had learned the particle semantics). The elimination of semantically-inappropriate judgment trials in Experiment 3 allowed us to increase the number of semantically appropriate judgment trials from 2 per cell (Experiments 1-3) to 3 per cell (Experiments 4-5).

[2]The experimenter produced 8 trials with each noun (squeako/woofo/oinko/purro), paired with either all plural scenes (preemption condition) or half plural and half singular scenes (entrenchment condition), as shown in the relevant columns. The particles (bup/kem) are shown in brackets in the "Utterance" column to indicate that the participant (rather than the experimenter) chooses which particle to produce on a given trial.

Baseline Vocabulary test (greyed out rows show foil animations)					
		Preemption condition		**Entrenchment condition**	
Trial Type	**Utterance**	**Meaning**	**Scene**	**Meaning**	**Scene**
Restricted noun 1 – bup only	squeako bup x 1	Several mice Several dogs Several pigs	bup=plural a	One mouse One dog One pig	bup=singular
	*squeako kem x 1	Several mice Several dogs Several pigs	kem=plural b	Several mice Several dogs Several pigs	kem=plural
Restricted noun 2 – kem only	*woofo bup x 1	Several mice Several dogs Several pigs	bup=plural a	One mouse One dog One pig	bup=singular
	woofo kem x 1	Several mice Several dogs Several pigs	kem=plural b	Several mice Several dogs Several pigs	kem=plural
Alternating noun –bup+kem	oinko bup x 1	Several mice Several dogs Several pigs	bup=plural a	One mouse One dog One pig	bup=singular
	oinko kem x 1	Several mice Several dogs Several pigs	kem=plural b	Several mice Several dogs Several pigs	kem=plural
Prediction: If participants have learned the noun vocabulary (for all participants, *squeako*='mouse', *woofo*='dog', *oinko*='pig'), they should choose the non-greyed-out option in each cell.					

Have participants in entrenchment condition learned particle semantics? (Research Q0): Production test[3] NB: for Experiments 4-5, this question is NOT addressed in the judgment task, as it included no semantically-inappropriate trials				
		Preemption condition	**Entrenchment condition**	
Trial Type	**Utterance**		**Meaning**	**Scene**
Alternating noun –bup+kem	oinko (bup) x 4	NA	One pig	bup=singular
	oinko (kem) x 4		Several pigs	kem=plural
Novel-at-test noun	purro (bup) x 4		One cat	bup=singular
	purro (kem) x 4		Several cats	kem=plural

[3]The experimenter produced 8 trials with each alternating/novel-at-test noun (oinko/purro), half paired with singular scenes (for which bup would count as a correct production), half paired with casual scenes (for which kem would count as a correct production). The particles (bup/kem) are shown in brackets in the "Utterance" column to indicate that the participant (rather than the experimenter) chooses which particle to produce on a given trial. The numbers (x4) reflect idealized performance (i.e., perfect learning of the semantics) for a given participant.

Footnotes

---

[1] In practice, some studies replace simple corpus frequency measures with proportional frequency or contingency measures designed to reflect the fact that some dispreferred verb+construction combinations do nevertheless occur in the relevant corpora with some non-trivial frequency. However, if we restrict our consideration to verb+construction combinations that are unattested in the corpora used to derive the measures (e.g., *laugh* in the transitive causative construction), these various ways of calculating preemption and entrenchment are equivalent.

[2] Of course, we have no way of knowing precisely what meaning participants in the preemption condition attached to each marker: Since they only saw causal scenes, they may well not have associated the markers with causality at all. The crucial point is whatever meaning they attached to them was – unlike in the entrenchment condition – *the same for both markers*.

[3] All participants reported speaking English as their primary language. Although we did not systematically exclude bilingual speakers (which would be unrepresentative, since bi-/multilingualism is the norm globally), the semantics of causal and noncausal actions are broadly similar crosslinguistically (e.g., Shibatani & Pardeshi, 2002).

[4] For participants in the entrenchment condition, half of the grammaticality judgment test trials were semantically appropriate (i.e., the trained causal marker appearing with a causal scene; the trained noncausal marker appearing with a noncausal scene), while half were semantically incorrect (the trained causal marker appearing with a noncausal scene; the trained noncausal marker appearing with a causal scene). Thus for participants in the entrenchment condition semantically incorrect trials (16/32) were excluded from the analysis. For participants in the preemption condition, no trials were semantically incorrect, because both markers appeared with causal scenes during both training and test. However, given that our chief goal was to compare entrenchment and preemption, it would have been highly undesirable to have included in the statistical analysis twice as much data in the preemption condition as in the entrenchment condition. We therefore excluded from the preemption condition the 16 trials corresponding to the semantically incorrect trials in the entrenchment condition. Thus, for both the entrenchment and preemption conditions, all key statistical analyses were conducted across 16 semantically-appropriate test trials.

[5] For experiment 1, we did not pre-register these key analyses testing prediction 2,3 & 4 for unattested combinations for restricted versus novel-at-test versus. Instead, we (erroneously) pre-registered only the analyses comparing grammaticality ratings for attested versus unattested combinations for trained verbs. We realized retrospectively that this was not a strict enough test, and thus it is moved to secondary analyses in the reporting. (From Experiments 2 onwards, we appropriately pre-registered the stricter analyses with novel verbs as the key hypotheses and the other as secondary)].

[6] It is for this reason that all our preregistered tests of preemption/entrenchment compare acceptability ratings across **unattested verb+particle combinations only**, with “alternating” verbs (i.e., those that appeared with both particles during training) excluded from the analysis. Unsurprisingly, participants do show high acceptability ratings for previously

witnessed verb+particle combinations (around 4.5 on the 5-point scale; see Supplementary Figure 1 of the supplementary materials at the OSF project site), but this is consistent with a mere familiarity preference. Hence higher ratings for these witnessed than unwitnessed forms should not be taken as evidence for preemption or entrenchment.

[7] Looking across studies provides for another (albeit unplanned, and between-subjects) test of the entrenchment hypothesis. Since Experiment 3 was a double-training version of Experiment 2 (and Experiment 1), the entrenchment hypothesis would seem to predict that the trained-restricted verb will be rated – in the entrenchment condition – as less acceptable in the unwitnessed condition of Experiment 3 (higher-frequency due to the double training) than in the unwitnessed condition of Experiment 2 or Experiment 1 (lower frequency). However, it is clear from Figure 3 that there is not even a numerical trend in this direction.

[8] An anonymous reviewer suggest that we test for this effect statistically, but this is not straightforward within the current framework, given that the lack of a Bayesian prior means that there is no way to calculate the Bayes Factor. Note however from Figure 3 that the non-overlapping Credible Interval (Highest Density Intervals) are consistent with the presence of a statistically reliable effect in the opposite direction to that predicted.

[9] As in Experiments 1-3, we have no way of knowing precisely what meaning participants in the preemption condition attached to each particle: Since they only saw plural scenes, they may not have associated the particles with plurality at all. The crucial point is whatever meaning they attached to them was – unlike in the entrenchment condition – *the same for both particles*.

[10] As well as via the learning of semantically-based restrictions that we do not consider here.