Modeling the Partial Productivity of Constructions

Libby Barak and **Adele E. Goldberg**
Psychology Department
Princeton University
NJ, USA 08540

Abstract

People regularly produce novel sentences that sound native-like (e.g., *she googled us the information*), while they also recognize that other novel sentences sound odd, even though they are interpretable (e.g., ? *She explained us the information*). This work offers a Bayesian, incremental model that learns clusters that correspond to grammatical constructions of different type and token frequencies. Without specifying in advance the number of constructions, their semantic contributions, nor whether any two constructions compete with one another, the model successfully generalizes when appropriate while identifying and suggesting an alternative when faced with overgeneralization errors. Results are consistent with recent psycholinguistic work that demonstrates that the existence of competing alternatives and the frequencies of those alternatives play a key role in the partial productivity of grammatical constructions. The model also goes beyond the psycholinguistic work in that it investigates a role for constructions' overall frequency.

1 Introduction

Native speakers of a language generalize beyond the sentences they witness in order to produce new utterances, while at the same time they generally avoid overgeneralizations that would sound unnatural. For example, as soon as *google* became a household term, English speakers freely used the verb to *google* in the double-object (DO) construction (*Google me the instructions*). At the same time, native speakers restrict their creative potential to avoid producing seemingly parallel examples, which in fact sound odd to proficient English speakers (?*Explain me the instructions*). The fact that constructions tend to be partially but not fully productive has been puzzled over for decades (Braine 1971; Bowerman 1990; Pinker 1989; Yang and Montrul 2017).

Several psycholinguistic studies have made some headway on the problem as experimental work has investigated possible factors that enable people to learn and use language creatively without overgeneralizing to produce utterances that sound odd. Three factors in particular have been the focus of much work: whether the verb and construction are semantically compatible; the frequencies of verbs, constructions, and verbs-in-constructions; and whether there exists a readily-available competing alternative way to express the intended message.

Recent work has found that when there exists a readily available (agreed upon) conventional way to express an intended message, it tends to *statistically preempt* the use and acceptability of a novel formulation (Boyd and Goldberg 2011; Goldberg 1995; 2006; 2011; Robenalt and Goldberg 2015). That is, participants prefer a familiar witnessed formulation over a novel formulation that is intended to express the same message. For example, when presented with the novel formulation, ?*She explained him the news* and asked to paraphrase it, people tend to agree on the paraphrase, *She explained the news to him*. This has been attributed to the fact that *explain* has been previously witnessed in the "caused-motion" (CM) paraphrase, far more often than in the DO, even in discourse contexts in which the DO would have been appropriate (for other verbs) (Goldberg 2011). At the same time, if there does not exist a readily available (agreed upon) alternative, people are more likely to accept the novel formulation. For example, since the novel utterance, *She sang him into another dimension*, does not lend itself to an agreed upon, more conventional alternative paraphrase, speakers are relatively free to use this formulation, combining *sing* and the CM in a novel way (Robenalt and Goldberg 2015).

Other work has argued that people generally do not stray far from the input, regardless of what message they intend to convey. This view predicts that novel pairings of a familiar verb with a construction should be less and less acceptable as the overall frequency of the verb increases. That is, high frequency verbs should be particularly resistant to novel uses, since they have been observed frequently enough that their conventional distributions would be highly entrenched. A key difference between this *conservatism via entrenchment* proposal, and *statistical preemption* is that conservatism is not directly influenced by the intended message being conveyed (Ambridge et al. 2012; Stefanowitsch 2008; 2011).

Psycholinguistic work directly comparing *statistical preemption* and *conservatism via entrenchment* has found conflicting results (Ambridge and Blything 2015; Ambridge et al. 2012). In a recent paper aimed at clarifying these apparent contrastive findings, (Ambridge, Barak, and Wonnacott 2018) point to an underlying methodological difficulty in teasing apart the various distributional measures in the corpus-based analysis framework. By utilizing the analytical power of a computational model, we are able to overcome these limitations by directly manipulating the input to the model to represent different hypothesized scenarios.

The incremental Bayesian computational framework described here simulates human behaviour in the context of complex naturalistic data. It provides a useful testbed to investigate possible influences of semantics, frequencies, and competition. The model's framework allows for an evaluation of whether the model can recognize when there exists a readily-available agreed-upon competing formulation of a novel utterance type (Robenalt and Goldberg 2015). In particular, Robenalt and Goldberg asked native English speakers to provide a paraphrase of various novel combinations of verbs and constructions. When speakers tended to converge on the same paraphrase for a novel sentence, the sentence is considered to have a readily-available agreed-upon competing alternative (has-CA); when speakers were more likely to vary in their paraphrases of a sentence, the sentence is considered not to have a readily-available agreed-upon competing alternative (no-CA). Robenalt and Goldberg found that speakers rated equally novel sentences as more acceptable when there was no competing alternative (Robenalt and Goldberg 2015). The present model replicates this result. Specifically, provided with a verb and an intended message, our model identifies the most likely syntactic pattern to be used to express that message. Results demonstrate that the existence of a readily available competing alternative essentially leads to a rejection of a novel formulation and the substitution of the more familiar formulation of the intended message. The model additionally replicates the the psycholinguistic finding that the frequency of a verb plays a role when there is a competing alternative formulation (Robenalt and Goldberg 2015): the more frequently an alternative formulation has been witnessed, the more strongly it is preferred.

Importantly, the present model goes beyond psycholinguistic work in that it illuminates a possible role for the overall construction frequency. That is, the model is able to compare the role of token frequency of a verb in the competing construction, with the frequency of the competing construction across all verbs. The model replicates the finding, just mentioned above, that a familiar formulation is more likely to be predicted over a given novel formulation when the specific verb is more frequent in the competing construction; the model also demonstrates a secondary role for the overall frequency of the construction.

2 Computational Model

2.1 Previous models

Several computational models have been applied to explain human-like performance on production and judgment of combinations of verbs and constructions. Many of these have focused on the ability to generalize beyond the observed data in the case of verb alternations (Barak, Goldberg, and Stevenson 2016; Ambridge and Blything 2015; Barak, Fazly, and Stevenson 2014a; Parisien and Stevenson 2011; Perfors, Tenenbaum, and Wonnacott 2010). But this approach is limited in its coverage since a verb alternation involves only two distinct constructions, each with a similar meaning. Other work has looked at additional aspects of restricting over-generalization behaviour. Freudenthal et al. (2007) and Connor, Fisher, and Roth (2012) present computational frameworks that make minimal assumptions about the learner's knowledge, while showing some ability to retreat from overgeneralization errors. However, these models are limited in the ability to represent fine-grained properties of the events since neither framework includes semantic properties.

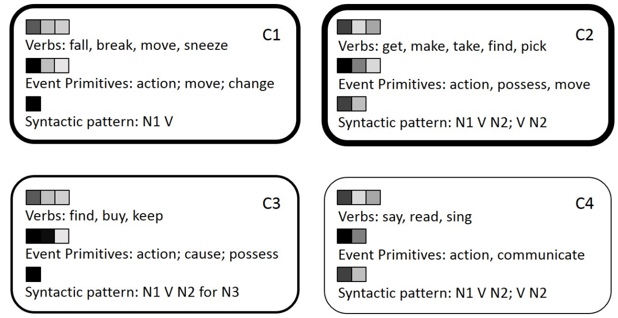

Figure 1: An illustration of a snapshot of the clusters learned by the model. The model assigns each input frame to a single cluster by counting the number of times each feature value was observed; e.g., the first cluster, marked *C1*, recorded the values *fall*, *break*, and *move* for the *verb* feature. The observations of each value in each cluster are marked by darker shades for frequently observed values. The size of the cluster is illustrated by the weight of the cluster's borders in the figure. Note, that the semantic and syntactic properties are simplified in the illustration.

The model of Alishahi and Stevenson (2008) offers an incremental Bayesian clustering framework that incorporates semantic and syntactic properties into a learning process designed to simulate natural learning mechanisms. It uses an incremental process that facilitates the replication of child over-generalization behaviour (e.g., ? *He fell the ball*). The original analysis was based on a small set of mostly high-frequency action verbs, but Barak, Fazly, and Stevenson (2013) extended the data provided to the model with additional verbs, a wider range of semantic classes, and a wider range of verb frequencies. When the incremental Bayesian clustering model was enriched with additional semantic features, it was found to be superior at replicating human judgments than otherwise comparable models (e.g., Ambridge and Blything 2015), due to the fact that it takes advantage of incremental clustering of semantic and syntactic properties (Barak, Goldberg, and Stevenson 2016).

In Section 2.2, we give an overview of the model of Alishahi and Stevenson (2008), followed by a more precise description of the mathematical formulation of the model (see Section 2.3).

2.2 The present model

The present model simultaneously and incrementally learns clusters that resemble grammatical constructions (Alishahi

head predicate	*FIND*
Syntactic Features:	
syntactic pattern	N1 *V* N2 *for* N3
argument count	3
complement type	none
Semantic Features:	
event primitives	$\{$ *physical*, *cause*, *possess* $\}$
event participants	$\{$ *agent*, *animate*, *cause* $\}$ $\{$ *theme*, *changed* $\}$ $\{$ *beneficiary* $\}$

Table 1: An example **usage-event**. The Syntactic features reflect an utterance such as *He found a book for Danny*: i.e., syntactic pattern 'N1 *V* N2 *for* N3', 3 arguments, and no sentential complement. The semantic features reflect a corresponding conceptualized FINDING event with a physical action described as ($\{physical, cause, possess\}$) whose 'N1' participant ($\{agent, animate, cause\}$) locates the 'N2' ($\{theme, changed\}$) for 'N3' ($\{beneficiary\}$).

and Stevenson 2008). See Figure 1 for an illustration of the clusters formed by the model. Input to the model consists of a sequence of *usage-events* representing a verb with its semantics, in a particular construction with its syntactic and semantic features, including the use of *prepositions*, *argument count*, *semantic participants of the event*, and *semantic features of the event*. For these annotations, we rely on the parsed and tagged data from Barak, Fazly, and Stevenson (2014a) as the training input for the model. Table 1 presents a sample usage-event illustrating values for these features.

At each step, the model assigns a new usage-event to a cluster that is most similar to the usage-event in feature values. Thus clusters grow incrementally with every addition of a usage-event. If none of the existing clusters is similar enough, the model uses the current usage-event to create a new cluster. The clustering decision is permanent in the sense that a usage-event cannot be removed from a cluster, and clusters cannot be deleted, merged, or divided over the course of training.

Since each usage-event involves both a verb and a construction, the input to the model does not distinguish between the contribution of each. And yet, constructional generalizations emerge as different usage-events involve distinct verbs with overlapping properties in their syntactic configuration and event participants.

The model can use the clusters to associate the meaning of a novel verb in *She's gorping him something* with a transfer event and can deduce that a sentential complement syntactic pattern is the most appropriate to convey a mental meaning, and so on.[1] This is due to the fact that Figure 1, each cluster includes similar verb usages that co-occur with the same semantic and syntactic values. The learning mechanism thus allows the model to generalize over the input in order to bootstrap from the observed input to select appropriate semantics or syntax for a message that is underspecified (for either semantics or syntax).

[1] We use the framework developed by Barak, Fazly, and Stevenson (2012) which includes a mechanism to address the use of sentential complements. Since the current study concerns other constructions, we focus on the relevant properties of the model. See Barak, Fazly, and Stevenson (2012) for details.

2.3 Learning Clusters

The model incrementally learns clusters from a sequence of usage-events (Us). Importantly, the number of clusters and their values are not determined in advance. For each usage-event, the model identifies the best cluster by maximizing over the similarity in values of the semantic and syntactic features of the frame and the clusters:

$$\mathrm{BestCluster}(U) = \underset{k \in Clusters}{\mathrm{argmax}}\ P(k|U) \tag{1}$$

where k ranges over all existing clusters and a new one. Using Bayes rule:

$$P(k|U) = \frac{P(k)P(U|k)}{P(U)} \propto P(k)P(U|k) \tag{2}$$

$P(k)$ is estimated based on the relative size of the cluster given all observed frames. In this way, the model gives a higher probability to bigger clusters. The probability of a new cluster is estimated as a cluster with a single usage-event. In early stages, when most clusters do not record many verb usages, the relative size of a new cluster (of 1) is similar enough to existing clusters to encourage the creation of more new clusters than at later stages of learning. The similarity of the event usage U and the cluster k is measured by their feature values assuming independence of the features. Formally, the likelihood $P(U|k)$ is estimated as:

$$P(U|k) = \prod_{f_i} P_i(v|k) \tag{3}$$

where f_i refers to the i^{th} feature of U and v refers to its feature value, and $P_i(v|k)$ is calculated using a smoothed version of:

$$P_i(v|k) = \frac{\mathrm{count}_i(j,k)}{n_k} \tag{4}$$

where $\mathrm{count}_i(j,k)$ is the number of times the value v was observed in cluster k for the f_i feature out of all the frames clustered to k denoted by n_k. Note that as the learning process progresses the model prefers bigger clusters that record more information despite having the ability to create a new cluster directly corresponding to the given frame by using Eq. (3).

3 Experimental Setup

3.1 Input and Training

We aim to train the model on rich linguistic data that represent the key distributional properties available to child learners. The frequencies of each verb, construction, and verb-in-construction are represented by the input stream, which was created to be naturalistic, as described here. A specific input stream of usage-events (i.e., verbs in constructions) is automatically generated following the methodology used by Alishahi and Stevenson (2008) and Barak, Fazly, and Stevenson (2014a). The generation of the input is based on an *input-generation lexicon* that contains an entry for each of the 71 verbs included in the data of Barak, Fazly, and Stevenson (2014a). The input covers a wide range of verbs with varying semantic properties, syntactic constructions, and frequency ranges (e.g., *go*, *fall*, *want*, *see*, *believe*).

The lexicon consists of entries for each of the 71 verbs denoting its overall frequency in a collection of eight corpora from CHILDES (MacWhinney 2000).[2] Furthermore, the lexicon specifies the frequencies of each verb in each construction, and the intended abstract interpretation of each usage (e.g., whether the event is one of transfer, caused-motion, desire, etc.). The construction frequencies for each verb are estimated based on a manual annotation of a random sample of 100 uses of each verb in the above corpora. The syntactic constructions are identified directly from the corpora, while the semantic properties are adapted from several resources, including Alishahi and Stevenson (2008), Kipper et al. (2008), and Dowty (1991).

To create a single input stream of 10000 event usages for the model, we randomly pick a sequence of verbs, one verb at a time, from the set of all verbs based on the overall frequency of the verbs. Since the model is sensitive to the order of presentation, we generate 100 individual input streams using this process. Each input stream captures the same distributional properties observed in CDS, while the order of presentation of verbs varies across the 100 simulations. The reported results are averaged, given all simulations for each of the experimental sections.

3.2 Set-up of Simulations

Our experiments are designed to analyze how various distributional properties of the input influence the acceptability of a novel usage-event. For this purpose, we begin by simulating the paraphrase task reported by Robenalt and Goldberg (2015). In the behavioral study, participants were presented with a novel use of a verb in a construction that it does not normally occur with, with verb-in-construction novelty confirmed using the Corpus of Contemporary American English (COCA) corpus (Davies 2008). In order to operationalize whether a novel sentence had a competing alternative (Has-CA) or not (No-CA), Robenalt and Goldberg (2015) asked a group of participants to paraphrase the sentence if they could think of a better way to express the intended meaning with the same verb; if not, participants were to simply rewrite the sentence as it was. Each sentence was considered to have a readily available competing alternative (has-CA), when more than half of participants converged on the same paraphrase; if fewer than half of participants agreed on a paraphrase, the use of the verb was considered to not have a competing alternative (no-CA). For example, participants tended to agree that the use of *find* in ?*Find a book to me* is more conventionally expressed by *Find a book for me*. On the other hand, *He sang them into another dimension* had no agreed upon paraphrase involving the verb *sing*: instead, participants suggested a wide range of paraphrases for this sentence.

We simulate this paraphrase task in our model as follows. First, the model is trained on a randomly ordered input corpus of 10000 verbs-in-constructions that naturally occurred in the corpora. Then, we present the model with a novel usage-event for a particular verb, e.g. *Find the book to me* or *Sing them into another dimension* (see Figure 1 for a full list of semantic and syntactic features composing a usage-event). After the model sees the novel usage-event exactly once, the model is queried with a *test usage-event* which consists of the same usage-event minus the specification of a syntactic pattern. That is, the model is required to predict the likelihood of each syntactic pattern, when queried about a combination of a fixed verb and semantic features.

The model's choice is considered as the syntactic pattern assigned the highest likelihood. This is then compared with whether the sentence (corresponding to a usage-event) was determined to have a competing alternative or not. The model's decision about whether an utterance has (or does not have) a competing alternative is determined by whether the same (or a different) syntactic pattern is considered most likely for the intended meaning and verb. If both participants and the model agreed on the availability of a competing alternative, we considered whether the model predicts the same paraphrase that human participants had as the most likely syntactic pattern.

At the end of each test query, we remove the occurrence of the *novel usage-event* from the learned clusters to prevent the model from developing an association of the verb to the feature values in that novel frame; i.e., the model "forgets" the test usage before it receives the next test usage.

Formally, we calculate the likelihood of each of the possible values v for the syntactic pattern given a test event usage U_{test}, as in:

$$P(v|U_{\text{test}}) = \sum_{k \in Clusters} P_{\text{main}}(v|k)P(k|U_{\text{test}}) \quad (5)$$

where $P_{\text{main}}(v|k)$ is the probability of the main predicate feature having the value v in cluster k, calculated as in Eq. (4), and $P(k|U_{\text{test}})$ is calculated as in Eq. (2) (see Section 2 for details).

4 Experiments and Analysis of Results

We evaluate the likelihood of the model to predict the most appropriate syntactic construction for a given test novel usage as explained in the section above. In Section 4.1, we first analyze the performance of the model by attempting to replicate the results reported by Robenalt and Goldberg (2015) for novel usages with and without a competing alternative. We then evaluate how various factors related to frequency interact within the model and how those relate to psycholinguistic findings; to do this, we directly manipulate the frequency of a construction, or separately, the frequency of a verb-in-construction, to consider test case scenarios in the context of the same simulation task (see Section 4.2).

4.1 Novel Construction Judgment

Here we focus on whether the model appropriately differentiates sentences that have a competing alternative (has-CA)

[2](Brown 1973; Suppes 1974; Kuczaj 1977; Bloom, Hood, and Lightbown 1974; Sachs 1983; Lieven, Salomo, and Tomasello 2009).

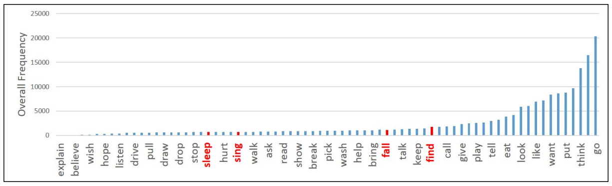

Figure 2: The overall frequency values observed for all verbs in our lexicon as measure in the CHILDES data (see Section 3.1.

	Overall raw of verb	Relative Frequency of V in CA
find	1749	2%
fall	1115	50%
sing	738	0%
sleep	704	0%

Table 2: The 4 verbs included in the testing data, along with their overall frequency, and their relative frequency with the CA, if available, as determined by the CHILDES data. The CA is taken to be the most likely syntactic pattern; e.g., "N1 find N2 for N3" for *find*, and "N2 fall" for *fall*.

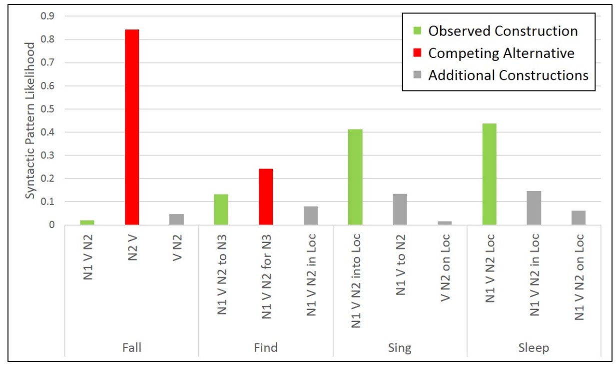

Figure 3: The model's likelihoods for the top three values for the *syntactic pattern* for each of four verbs: two has-CA, *find* and *fall* on left; and 2 no-CA, *sing* and *sleep* on right.

or not (no-CA) according to human judgments. That is, we evaluate whether the test usage-event specifying a verb and an interpretation is predicted by the model as most likely with the same or different syntactic pattern than that which had been witnessed once. If the model predicts a different syntactic frame, we compare the predicted frame to the alternative which was produced by humans.

Two verbs with similar frequencies were used in usage-events that were determined to have-CA: *Find a pen to me* and *Fell the lamp*. That is, participants converged on the preferred paraphrases, *Find a pen for me* and *The lamp fell*. Two other verbs with similar frequencies represent the No-CA scenario: *Sing the audience into another dimension* and *Sleep the afternoon away*. That is, participants did not converge on any shared paraphrase and often chose to simply repeat the novel sentence as it was (see Figure 2 for overall frequency information and Table 2 for CA frequency information).

Results replicate the finding of Robenalt and Goldberg (2015), as the model prefers a distinct formulation for the usage-events which corresponded to sentences that have-CA, while allowing novel usage-events when the corresponding sentences were determined to be No-CA (see Figure 3). In both cases, the usage-events were novel, so the model did not already have a cluster including the verb with the presented syntactic pattern and semantic properties; therefore, the model had to select the best cluster by maximizing the shared syntactic and semantic properties specified by the novel usage-event.

In the Has-CA scenario, the model identifies the novel usage as most similar to the biggest cluster that associates the given verb with the intended meaning, despite the mismatch in the novel usage of a syntactic pattern and observed usages of the same verb. The likelihood calculation is strongly affected by the cluster in making the prediction. Because the cluster has recorded previous uses of the same verb with the CA syntactic pattern and intended meaning, the CA wins over the novel syntactic pattern which has contributed only a single occurrence of the novel syntactic pattern value (since it had been novel). Importantly, the model considers *all* clusters in predicting the most likely syntactic pattern value for the test usage-event, as shown in Eq. 5. Although other clusters would include previous usage-events of the novel syntactic pattern, these clusters would not be as influential in predicting the winning syntactic pattern if the recorded values for other features did not match the test usage-event. Similarly, the likelihood is not as biased by clusters that record the novel syntactic pattern with verbs and semantic properties that do not match those in the novel usage.

In the No-CA scenario, the model correctly predicts the novel formulation is the most likely choice. We observe that the model predicts the novel syntactic pattern with similar confidence for both *sing the audience into another dimension* and *sleep the afternoon away*. Due to the lack of an existing cluster that associates the novel meaning for the given verbs in this syntactic pattern, the model adds the novel training usage to a cluster that records the use of these syntactic patterns for other verbs with the closest semantic properties available. For example, the use of *sing the audience into another dimension* is added to a cluster that includes usages of verbs such as *run* or *drop* with the same syntactic pattern and a partial overlap in semantic features (the caused-motion aspect is shared). At the time of prediction, the model considers clusters that match the test usage-event in any subset of values. Other clusters that recorded usage-events of the verb did not match as well in syntactic and semantic feature values.

The model mirrors human judgments in another respect as well. A number of studies have found that the acceptability of verbs used in novel ways correlates inversely with the frequency of the verb; i.e., higher overall verb frequency correlates with lower judgment scores on the novel sentences (e.g., Theakston 2004). Robenalt and Goldberg (2015) note that previous work finding this effect had included only novel sentences that had a competing alternative. They tested both types of sentences and found that verb frequency only correlated (negatively) with judgments on novel sentences that had a competing alternative. In the case of novel sentences without a clear competing alternative, no effect of verb frequency was found. Therefore, it may be the frequency of the verb in its CA that correlates inversely with the frequency of the novel sentence, not the overall frequency of the verb.

The modeling results are consistent with this interpretation of the judgment data. Notice that here is a difference in the degree of likelihood predicted for each of the two usage-events in the has-CA scenario, in that the model finds that the CA for *Fall the lamp* has a much higher likelihood than the CA for *Find a new pen to me*. Interestingly, even though *find* has slightly higher frequency than *fall*.

the frequencies of their CAs are markedly different. Only 2% of the usages of *find* occur in its CA (*Find something for someone*) compared with 50% of the usages of *fall* with its CA (*Something fell*). With both verbs having an overall frequency of about 1500 occurrences, the single novel pairing of *find* with a novel syntax is compared against about 30 usages (2% of 1500), while the novel form for *fall* is compared against roughly 750 usages (50% of 1500). Thus, the likelihood of the competing alternative increases with its frequency. The prediction that verb frequency is not relevant when there is no-CA is not tested in the model as described, *sing* and *sleep* occur with roughly equal frequency in the corpus, and are predicted to occur in the novel construction with roughly the same likelihood.

An additional factor that the model allows us to clarify is whether the overall frequency of the CA grammatical construction matters or whether the only relevant frequency is that of the specific verb in the CA construction. Recall that the model creates cohesive clusters that represent the typical meaning expressed by a syntactic pattern across several verbs; e.g., a cluster that includes e.g., *fall*, *break*, *move*, is associated with intransitive syntax (Figure 1). The number of usage-events within a given cluster is the (token) frequency of a construction.

The model takes each construction's frequency into account in determining how likely a new usage-event is to be associated with a given construction (higher frequency correlates with higher likelihood). This then influences the likelihood of the intransitive cluster attracting a novel usage-event such as *He fell the lamp* because the high frequency of the cluster increases the likelihood of that cluster being selected. The high frequency intransitive cluster fits best for the test usage-event despite the novel syntactic pattern *he fell the lamp*, because the this cluster includes occurrences of the verbs and significant overlap in semantic attributes, e.g., additional verbs describing *motion*.

The relatively lower likelihood of *find* with the CA can be thus explained in two ways. First, the relatively lower frequency of "N1 V N2 for N3" in the input (compared with the intransitive syntax) may result in a less strongly entrenched cluster for this syntax (represented by the weaker line around *C3* in Figure 1). Second, an alternative explanation can be found in the lower frequency of *find* with the CA compared

	Overall freq.	Observed freq. in CA	Modified freq. of find	Modified freq. of construction
ask	818	45	45	186
buy	1442	103	103	401
change	553	13	13	58
cut	885	20	20	92
draw	615	15	15	68
find	**1749**	**37**	**747**	**37**
keep	1392	38	38	170
leave	1214	31	31	141
make	4165	65	65	306
sing	738	13	13	62
write	650	23	23	102
Total	14221	404	1113	**1622**

Table 3: All verbs that occur with "N1 V N2 for N3" pattern in the lexicon, along with (i) their overall frequencies as measured in the CHILDES data , (ii) their number of occurrences with the CA, "V NP to NP", (iii) frequencies modified for *find* only, and (iv) frequencies of the construction modified, but not for experiment (modified for all verbs but *find*). Total number of occurrences of verbs overall and with the construction is given in the bottom line.

with the higher frequency of *fall* with its CA.

In the following section, we analyze these two possible factors–whether overall frequency of the CA or the frequency of the particular verb in the CA–determine the degree of likelihood of selecting a syntactic pattern.

4.2 Construction Frequency

Does the overall token frequency of the CA, *across all participating verbs* play a role in determining the likelihood of the CA? And if so, how does it compare to the frequency of the particular verb in its CA? In this section, we investigate these questions. In particular, we investigate the model's predictions for *find a pen for me* as the CA for *find a pen to me*, when we systematically vary a) the overall frequency of the CA construction (the number of tokens of all verbs occurring in "V NP for NP") or b) the frequency of the CA with the verb *find* in particular "find NP for NP" (see Table 3).

In Figure 4, the first panel shows the result we have already seen: the model's prediction for *find* being used in its CA with the original frequencies as found in CHILDES. The center panel in Figure 4 shows the predictions when the CA construction is given a higher token frequency for all verbs except *find*: this leads to a higher likelihood of the CA, compared with the original scenario. That is, the higher number of occurrences of verbs other than *find* with "N1 V N2 for N3" increases the likelihood of *find* being used in the same pattern. As the model observes more usages of the CA, the relevant cluster grows. Our prediction method thus raises the probability of this cluster for the sake of prediction (see Eq. (5)).

Importantly, this influence is made possible by the co-occurrence of verbs with the CA that have semantic properties that are similar to those of *find*. If the increased number of usages of the CA had recorded unrelated meanings, the cluster would have become *less* well suited to express the

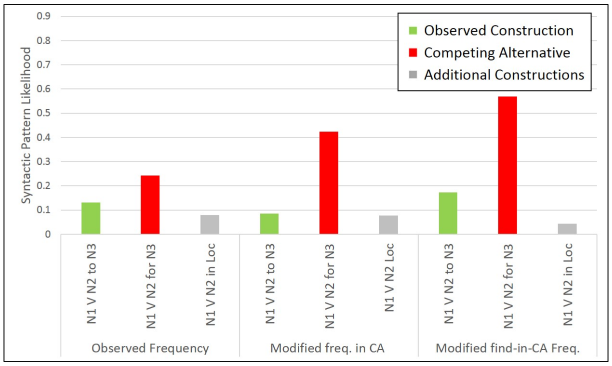

Figure 4: The model's likelihoods for the top three values for the *syntactic pattern* for 3 CA frequency scenarios for the verb *find*. On the left, the results using the observed frequencies in CHILDES data, followed by 2 artificial settings: (i) increased token frequency for all verbs in the CA except *find* and (ii) increased token frequency of only the verb, *find*, in the CA.

novel meaning and therefore its likelihood would have been reduced. That is, the increase in likelihood due to higher frequency of the construction is due to the higher number of usage-events of the CA *with* the intended meaning. This is consistent with psycholinguistic results reported by Suttle and Goldberg (2011).

The model also provides support for the influence of the frequency of the verb *find*, in particular, in the CA, as suggested by Robenalt and Goldberg (2015). By increasing the number of occurrences of *find* with the CA, the likelihood of the CA reaches the highest rank among the three scenarios (see rightmost panel of Figure 4). In this case, the cluster recording the usage-events of *find* with the CA becomes more entrenched as it did when the frequencies of other verbs were increased. In addition, the fact that the additional usage-events share the same semantic properties (rather than, only partially overlapping meaning), increases the likelihood of the CA for *find* all the more, as it provides evidence that matches the novel formulation in use of verb (fully), syntactic (partially), and semantic properties (fully).

5 Discussion

This study extends the explanatory power of current psycholinguistic methods by using a Bayesian computational model to simulate the language acquisition process. The computational framework enables us to analyze the interaction of several distributional properties as the model is trained incrementally on a naturalistic stream of usage-events. The model successfully demonstrates the ability to automatically differentiate between novel usage-events that should be avoided and those that are more acceptable as creative extensions. The model correctly captures the following: a) constructions emerge as generalizations over usage-events that have the same form and similar meanings, b) familiar formulations of an intended message (CAs) are preferred over novel formulations, c) if there is no CA, a novel-usage-event is predicted, d) increasing the frequency of the CA construction (with related verbs) increases the likelihood that the CA will be preferred, and e) increasing the frequency of the more familiar CA construction with a particular verb results in an even higher likelihood that the alternative will be preferred for that verb.

The present model, incrementally trained on naturalistic data, learns generalizations about the relationships between form and meaning for specific verbs, for specific constructions, and for combinations of verbs and constructions. The model and analysis presented here has benefited from work in psycholinguistic research, and the model in turn can be a benefit to psycholinguistic research. The model supports the idea that the existence of competing alternatives plays a role in the likelihood of accepting a novel combination of verb and construction. The model also suggests that the frequency of the competing construction, especially with that same verb, plays a role. Future work is needed in order to add more nuanced semantic properties to the model and more usage-events. We aim to have demonstrated that this type of model provides a useful testbed for work on construction-learning and use of constructions.

Importantly, Barak, Fazly, and Stevenson (2014a) presents an extension of the model of Alishahi and Stevenson by simultaneously learning verb classes that capture high-level correlations among verbs and constructions. The extended model enables a closer replication of human behaviour on a range of different tasks relating to argument structure acquisition (Barak, Goldberg, and Stevenson 2016; Barak, Fazly, and Stevenson 2014a; 2014b). In this study, we choose to perform a preliminary analysis using only the original model of Alishahi and Stevenson (2008) to evaluate the role of the full set of properties available to a speaker who aims to convey a certain message with a particular verb and construction. We aim to extend this analysis in the future with a novel methodology, utilizing the extended model of Barak, Fazly, and Stevenson (2014a) to simulate grammatical error detection as analyzed by psycholinguistic studies. In addition, we hope to extend our current analysis to a more comprehensive set of verbs and constructions that would evaluate the language acquisition process over a broader range of case scenarios.

References

Alishahi, A., and Stevenson, S. 2008. A computational model of early argument structure acquisition. *Cognitive Science* 32(5):789–834.

Ambridge, B., and Blything, R. P. 2015. A connectionist model of the retreat from verb argument structure overgeneralization. *Journal of child language* 1–32.

Ambridge, B.; Barak, L.; and Wonnacott, E. 2018. Effects of both preemption and entrenchment in the retreat from verb overgeneralization errors.

Ambridge, B.; Pine, J. M.; Rowland, C. F.; and Chang, F. 2012. The roles of verb semantics, entrenchment, and morphophonology in the retreat from dative argument-structure overgeneralization errors. *Language* 88(1):45–81.

Barak, L.; Fazly, A.; and Stevenson, S. 2012. Modeling the acquisition of mental state verbs. In *Proceedings of the 3rd*

Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2012).

Barak, L.; Fazly, A.; and Stevenson, S. 2013. Acquisition of desires before beliefs: A computational investigation. In *Proceedings of CoNLL-2013*.

Barak, L.; Fazly, A.; and Stevenson, S. 2014a. Learning verb classes in an incremental model. In *Proceedings of the 5th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2014)*. Association for Computational Linguistics.

Barak, L.; Fazly, A.; and Stevenson, S. 2014b. Gradual acquisition of mental state meaning: A computational investigation. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*.

Barak, L.; Goldberg, A. E.; and Stevenson, S. 2016. Comparing computational cognitive models of generalization in a language acquisition task. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, volume to appear. Association for Computational Linguistics.

Bloom, L.; Hood, L.; and Lightbown, P. 1974. Imitation in language development: If, when, and why. *Cognitive Psychology* 6(3):380–420.

Bowerman, M. 1990. Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics* 28(6):1253–1290.

Boyd, J. K., and 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. 1971. On two types of models of the internalization of grammars. *The ontogenesis of grammar* 153–186.

Brown, R. 1973. *A first language: The early stages.* Harvard Univ. Press.

Connor, M.; Fisher, C.; and Roth, D. 2012. Starting from scratch in semantic role labeling: Early indirect supervision. *Cognitive Aspects of Computational Language Acquisition*.

Davies, M. 2008. *The corpus of contemporary American English*. BYE, Brigham Young University.

Dowty, D. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67(3):547–619.

Freudenthal, D.; Pine, J.; Aguado-Orea, J.; and Gobet, F. 2007. Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science* 31(2):311–341.

Goldberg, A. E. 1995. *Constructions, A Construction Grammar Approach to Argument Structure*. {Chicago University Press}.

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

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

Kipper, K.; Korhonen, A.; Ryant, N.; and Palmer, M. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation* 42(1):21–40.

Kuczaj, Stan, A. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16(5):589–600.

Lieven, E.; Salomo, D.; and Tomasello, M. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3):481–507.

MacWhinney, B. 2000. *The CHILDES project: Tools for analyzing talk*, volume 2. Psychology Press.

Parisien, C., and Stevenson, S. 2011. Generalizing between form and meaning using learned verb classes. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*.

Perfors, A.; Tenenbaum, J. B.; and Wonnacott, E. 2010. Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language* 37(03):607–642.

Pinker, S. 1989. *Learnability and cognition: The acquisition of argument structure.* The MIT Press.

Robenalt, C., and 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.

Sachs, J. 1983. Talking about the There and Then: The emergence of displaced reference in parent–child discourse. *Children's language* 4.

Stefanowitsch, A. 2008. Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19(3):513–531.

Stefanowitsch, A. 2011. Constructional preemption by contextual mismatch: A corpus-linguistic investigation. *Cognitive Linguistics* 22(1):107–129.

Suppes, P. 1974. The semantics of children's language. *American psychologist* 29(2):103.

Suttle, L., and Goldberg, A. E. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6):1237–1269.

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

Yang, C., and Montrul, S. 2017. Learning datives: The tolerance principle in monolingual and bilingual acquisition. *Second Language Research* 33(1):119–144.