COGNITIVE SCIENCE
A Multidisciplinary Journal

Check for updates

Cognitive Science 47 (2023) e13328

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13328

When Children's Production Deviates From Observed Input: Modeling the Variable Production of the English Past Tense

Libby Barak,[a,b] Zara Harmon,[c,d] Naomi H. Feldman,[c,d] Jan Edwards,[e] Patrick Shafto[f,g]

[a]*Department of Linguistics, Montclair State University*
[b]*Department of Computer Science, Montclair State University*
[c]*Institute for Advanced Computer Studies (UMIACS), University of Maryland*
[d]*Department of Linguistics, University of Maryland*
[e]*Department of Hearing and Speech Sciences, University of Maryland*
[f]*Department of Mathematics and Computer Science, Rutgers University*
[g]*School of Mathematics, Institute for Advanced Studies*

Received 1 June 2022; received in revised form 14 June 2023; accepted 17 July 2023

Abstract

As children gradually master grammatical rules, they often go through a period of producing form-meaning associations that were not observed in the input. For example, 2- to 3-year-old English-learning children use the bare form of verbs in settings that require obligatory past tense meaning while already starting to produce the grammatical *–ed* inflection. While many studies have focused on overgeneralization errors, fewer studies have attempted to explain the root of this earlier stage of rule acquisition. In this work, we use computational modeling to replicate children's production behavior prior to the generalization of past tense production in English. We illustrate how seemingly erroneous productions emerge in a model, without being licensed in the grammar and despite the model aiming at conforming to grammatical forms. Our results show that bare form productions stem from a tension between two factors: (1) trying to produce a less frequent meaning (the past tense) and (2) being unable to restrict the production of frequent forms (the bare form) as learning progresses. Like children, our model goes through a stage of bare form production and then converges on adult-like

Correspondence should be sent to Libby Barak, Department of Linguistics, Schmitt Hall 222, Montclair State University, Montclair, NJ 07043, USA. E-mail: libby.berk@gmail.com

production of the regular past tense, showing that these different stages can be accounted for through a single learning mechanism.

Keywords: Language acquisition; Non-parametric Bayesian models; Morphology learning; Fragment grammars

1. The acquisition of past tense regular form

When learning grammatical rules in their native language, children apply those rules intermittently at first and only later become consistent in using the correct form. They produce the auxiliary verbs *be* and *have* only some of the time, and they also produce morphemes such as the plural *–s* and past tense *–ed* only some of the time in obligatory contexts (Marcus et al., 1992; Matthews & Theakston, 2006; Nicoladis, Palmer, & Marentette, 2007; Theakston, Lieven, Pine, & Rowland, 2005). In all of these cases, children sometimes produce incorrect forms, despite producing the target form correctly in other cases. For example, children may say "I *drop* the ball" to refer to a past event while using *dropped* correctly in other instances. Importantly, the acquisition of *–ed* as past-tense inflection serves as a key observation that characterizes the developmental trajectory of children with typical language development, children with language disorders, and bilinguals (Leonard, 2014). Yet, previous models of morphology acquisition have largely ignored the variability in children's productions between past tense *–ed* and the bare form. Here we focus on the acquisition of the past tense *–ed* morpheme as a case study of this broader phenomenon of early production variability.

Most previous models of morphology acquisition have focused on capturing (1) the cognitive mechanism that allows for limited generalization, that is, applying the past tense *–ed* to regular forms and not for irregular forms and/or (2) a possible stage of overgeneralization in which irregular verbs are occasionally produced with the regular inflection, for example, *goed* (Corkery, Matusevych, & Goldwater, 2019; Cottrell & Plunkett, 1994; Hare & Elman, 1995; Hoeffner, 1992; Kirov & Cotterell, 2018; Legate & Yang, 2007; MacWhinney & Leinbach, 1991; O'Donnell, 2015; Pinker & Prince, 1988; Plunkett & Juola, 1999; Rumelhart & McClelland, 1986; Yang, 2016). Those that have tried to capture the variable production of the regular *–ed* suffix and the bare form have either assumed that children's grammars contain incorrect form-meaning mappings, that is that they erroneously associate past tense to bare forms (Legate & Yang, 2007) or have made predictions about form production without distinguishing whether or not children are trying to specifically produce past tense meanings in these cases (Freudenthal, Gobet, & Pine, 2023; Freudenthal, Pine, & Gobet, 2010, 2006; Freudenthal, Pine, Jones, & Gobet, 2015; Freudenthal, Ramscar, Leonard, & Pine, 2021).

Psycholinguistic theories of this acquisition period also fail to fully explain the period of variable productions of inflectional morphemes that children go through. Two main theories have been proposed to explain this trajectory, inflection omission, and form licensing. Based on correct production of the inflected form in some obligatory cases, inflection omission argues that children have acquired the target form, for example, know that *–ed* should be used for regular past tense (Wexler, 1994, 1998). Thus, this theory interprets bare forms as an omission of the *–ed*. While this explanation aligns with observations from English, production variability alternates between different inflected forms in languages with more complex morphological systems, for example, using third-person singular inflection rather than first- or second-person inflection in Spanish (Aguado-Orea & Pine, 2015). An alternative explanation suggests that bare forms are licensed by children as grammatical for past tense (Legate & Yang, 2007). However, children do not observe evidence for this association in the input (e.g., child-directed speech), which means that this account necessitates a mechanism that aligns bare forms with unobserved meanings and a second mechanism that learns the correct mappings based on observations. We aim to replicate the production variability stage and subsequent grammatical convergence using a single learning mechanism.

In this paper, we propose a new account of the phenomenon of bare form production for obligatory past tense context. We build a computational model that acquires inflectional morphology and show that the variable production of bare form and *–ed* in past tense contexts arises naturally as a stage in the acquisition process in the model. In our model, rather than arising from incorrect form-meaning mappings, the variable productions arise because there is a protracted learning process for acquiring the strong correlation between the *–ed* form and the past tense meaning and for learning to restrict bare form production to grammatical contexts. This gradual acquisition of the form-meaning mapping leads the model to sometimes produce bare forms even after it has acquired the *–ed* form, and even when it is specifically trying to express a past tense meaning. We show that specific representational assumptions in the model are necessary for generating this variable behavior. Moreover, we show that, like children, during this variable stage of acquisition, our model demonstrates a comprehension–production asymmetry, that is, it is much better at using the mapping between *–ed* and the past tense meaning in comprehension than in production (Carr & Johnston, 2001; Wagner, Swensen, & Naigles, 2009). Together, these modeling results provide a novel explanation of why mixed productivity should occur as a stage in the acquisition process.

In the following sections, we present our formulation of morphology learning in a computational model followed by a review of the computational model. We then test our model on a production task and a comprehension task. Finally, we explain our results through an inspection of the model's learned grammar and its components. We show that bare form productivity can emerge in a learning model that aims to converge with naturalistic grammatical input, without requiring that we analyze this behavior as omission of *–ed* or as grammatical licensing of the bare form. Rather, we replicate child behavior by requiring the model to learn form-meaning association in the full complexity of naturalistic child-directed speech. Both stages of the incremental learning are achieved over a single learning mechanism and based exclusively on grammatical production, which leads to an interplay between accessibility of the target form and accessibility of other forms and their associated meanings as full grammatical knowledge develops.

2. Comparing models with different representational assumptions

We aim to capture children's behavior of bare form production for past tense meaning using a computational model. This presents a challenge because computational models are designed to adapt to observed input. How can a model produce a pattern that has not been experienced in the input? We argue that the interplay between frequency patterns of forms and the expression of meaning can lead a computational model to use the bare form to express the past tense, even when the bare form and the past tense meaning do not co-occur frequently in its input. The high frequency of bare forms can make it more accessible, while the lower frequency of past tense would delay its association with grammatical forms.

In support of this hypothesis, psycholinguistic evidence shows that production errors of bare form production correlate with the frequency of bare form use per verb (Räsänen, Ambridge, & Pine, 2014). This likely arises because of competing influences of the high frequency of bare form in the input and the desired expression of past tense. While the high frequency of bare forms aids its acquisition by increasing its accessibility (Harmon & Kapatsinski, 2017), the past tense meaning needs to be associated with a significantly less frequent form. Learning the correct inflection for past tense—the past-ed association—poses a dual challenge because, first, the frequency of *–ed* inflections is much lower compared with the bare form, and second, past tense meaning may be coupled with the *–ed* or with irregular forms.

We further argue that a computational model can only reach the stage of mixed productions observed in children if it needs to learn the correct association of tense and inflections. Learning semantic–syntactic associations from distributional input has been shown to be instrumental in replicating evidence from learning word meanings and avoiding overgeneralizations (Ambridge, Pine, Rowland, Jones, & Clark, 2009; Andrews, Vigliocco, & Vinson, 2009; del Prado Martın, Kostić, & Baayen, 2004; Goldberg, 2011; Ramscar, Dye, & McCauley, 2013). Surprisingly, previous computational models of bare form production have not been required to learn this association. Those previous models have either built-in knowledge that tenses and inflections were associated or have been purely form-based learners that were not required to learn associations between tenses and inflections at all (Freudenthal et al., 2010; Legate & Yang, 2007; O'Donnell, 2015). Models that required form-meaning learning from the model have focused on avoiding overgeneralization errors for irregular forms, overlooking this important prior stage of mixed production (MacWhinney & Leinbach, 1991; Ramscar et al., 2013).

Children hear *–ed* used primarily for past tense meaning, for example, *walked*. If the model already assumes that the *–ed* form in every use such as *walked* maps onto past tense meaning and bare form, for example, *walk*, implies the present tense, the model cannot form unobserved associations without directly implementing an underlying hypothesized cause. By requiring the model to learn form-meaning association over naturalistic input, we show that bare form may arise as a possible form for past tense. For example, the model needs to learn to associate the meaning of *walk* observations with the present tense but may infer any possible tense meaning for this form until this restriction is learned. More generally, the model infers a grammar that represents any of the possible English form-meaning associations while

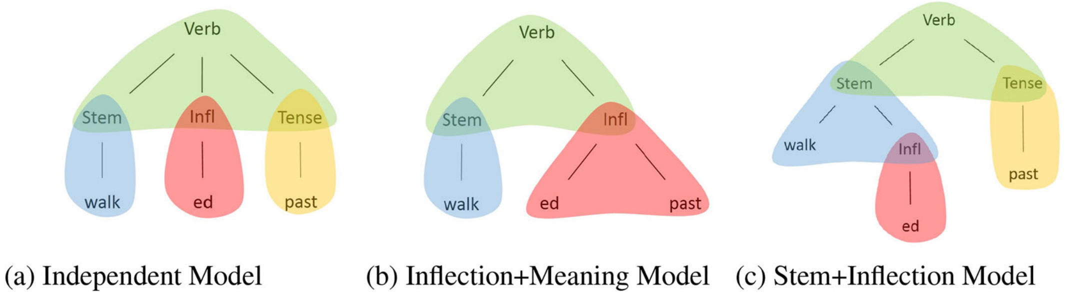

Fig. 1. Atomic context-free grammar rules of each model: (a) Independent, (b) Inflection+Meaning, and (c) Stem+Inflection. The rules include non-terminal verb rule (green), stem rule, terminal rule for the Independent and Inflection+Meaning, and non-terminal for the Stem+Inflection model (blue), terminal inflection rule (red), and terminal tense rule (yellow).

observing the high frequency of the bare form and use of *–ed* for multiple tense meanings, that is, past tense and past participle. The model needs to tie *–ed* to past tense specifically to make it the preferred form for past tense, while learning to limit the bare form to grammatical contexts despite its high frequency.

Fig. 1 illustrates three different possible representational assumptions about the associations between tense and inflection. Our model, the *Independent model*, represents the building blocks of inflected verbs as independent terminals: stem, inflection, and meaning (Fig. 1a). This represents how a child observes verbs used in sentences over every possible inflection and tense and needs to infer the correct grammar of the language, that is, what is the preferred way to generate these observations. We compare this with two baseline models. The first is the model proposed by O'Donnell (2015), hereafter the *Inflection+Meaning model*, which assumes an association of inflection and meaning, for example, "ed-past", as part of its rule structure (Fig. 1b). This model represents a child who learns for every stem, for example, *walk*, which meaningful inflections can be used with it, for example, "ed-past" or "ing-progressive". A second baseline model, the *Stem+Inflection model* learns the meaning for inflected stems (Fig. 1c). This model represents a child who observes the stems as associated with their parsed inflections, for example, "walk*–ed*" and learns the meaning for the inflected verb, for example, "past". As we explain in the next section, our Independent model can learn to mimic the structural assumptions of either the Inflection+Meaning model or the Stem+Inflection model if the input supports such associations, but it is required to learn those associations from the input rather than assuming one or the other a priori.

3. The fragment grammars model

We simulate morphology learning using the fragment grammars (FGs) model (O'Donnell, 2015). FGs has been previously shown to match human behavior in multiple aspects of morphology learning such as rule generalization and a limited period of overgeneralization. Yet, this model has not been applied to or tested on the mixed productivity period of *–ed* and

Table 1
Partial example of possible learned grammar over the FGs structure for each of the three models

	Independent	Inflection+Meaning	Stem+Inflection
Non-terminal rules	$Verb \rightarrow$ *Stem Infl tense*	$Verb \rightarrow$ *Stem Infl*	$Verb \rightarrow$ *Stem tense*
			$Stem \rightarrow$ walk *Infl*
			$Stem \rightarrow$ jump *Infl*
			$Stem \rightarrow$ think *Infl*
Terminal rules	$Stem \rightarrow$ walk	$Stem \rightarrow$ walk	
	$Stem \rightarrow$ jump	$Stem \rightarrow$ jump	
	$Stem \rightarrow$ think	$Stem \rightarrow$ think	
	Infl $\rightarrow$ ed	*Infl* $\rightarrow$ ed-past	*Infl* $\rightarrow$ ed
	Infl $\rightarrow$ null	*Infl* $\rightarrow$ null-present	*Infl* $\rightarrow$ null
	tense $\rightarrow$ present		*tense* $\rightarrow$ present
Memoized fragments	$Verb \rightarrow$ walk *Infl*	$Verb \rightarrow$ walk *Infl*	$Verb \rightarrow$ walk –ed *tense*
	$Verb \rightarrow Stem$ null-present	$Verb \rightarrow Stem$ null-present	$Verb \rightarrow Stem$ present
	$Verb \rightarrow Stem$ ed-past	$Verb \rightarrow Stem$ ed-past	$Verb \rightarrow Stem$ past

bare forms. We fill this gap by extending the model to capture our representational assumptions of the need to learn form-meaning associations over time based on the co-occurrence information in the input. We demonstrate that with those assumptions, the model captures the developmental trajectory of bare form use. We further show that models that lack such assumptions—that is, the two baseline models of Inflection+Meaning and Stem+Inflection—fail in modeling the phenomenon of bare form mixed productivity period.

FGs is a Bayesian non-parametric model that learns over a probabilistic context-free grammar (PCFG) structure. The context-free grammar (CFG), $\mathcal{G}$, consists of rules of the form $A \rightarrow \beta$, where β is the collection of terminal and non-terminal production rules (see Table 1 for possible rules in the CFG). The CFG rules encode into the model what the learner already knows, that is, the knowledge instantiated by the atomic rules. Table 1 illustrates the differences in the CFG rules across the three different models, which entail the encoded differences in each model's assumption on the learner's prior knowledge. For example, the Inflection+Meaning model associates every inflection with its tense meaning in a terminal rule, for example, *Infl* $\rightarrow$ ed-past. This rule implies that every occurrence of a verb inflected with *–ed* for the past tense is considered as an association of *–ed* with past tense. In contrast, the Independent model requires learning this association by representing the inflection and tense as two independent rules. While the CFG rules specify the assumed knowledge, the model requires the learner to infer the distribution of the rules from the input. The PCFG uses a Dirichlet multinomial distribution to draw over the finite set of rules in the CFG. The Dirichlet parameters balance the production probability of the rules included in the CFG.

FGs extend the PCFG framework by allowing the model to store fragments of the tree in the memory in a process termed *memoization*. By storing several rules as a single fragment, the model can encode the fact that certain rules often co-occur in the input. In this process, the Independent model, for example, can memoize ed-past as a model-inferred fragment. Creating this larger fragment allows the model to parse past tense *–ed* verbs in two ways: either by using the fragment or by using each rule independently. Fragments that are used frequently in the model's parses come to have a higher probability than the product of their component rule probabilities and are preferred because of this probability advantage. The memoization is thus a stochastic process that aims to balance a dual goal: (1) inferring a grammar that best accounts for the observed data while (2) moderating the creation and storing of new fragments to ensure efficiency.

The model implements the memoization process using the Pitman–Yor process, which is a generalization of a Dirichlet process (Pitman & Yor, 1997). As the model observes input items, the memoization function will choose whether to use an existing fragment or create a new one. The model considers non-terminal rules, terminal rules, and memoized fragments as fragments. The Pitman–Yor process relies on two parameters to moderate the memoization process $\{< a^A, b^A >\}_{A \in V_{\mathcal{G}}}$, where $V_{\mathcal{G}}$ is the set of non-terminal symbols in the grammar and a^A and b^A are Pitman–Yor parameters a and b associated with non-terminal A. The function either chooses to use an existing fragment f with the probability $\frac{n_f - a}{N+b}$, where n_f is the number of times fragment f has been used, a and b are the Pitman–Yor parameters, and N is the total number of fragments sampled so far. Alternatively, the function samples a new value with the probability $\frac{aK+b}{N+b}$, where K is the number of times a new fragment has been sampled from the underlying function (i.e., the number of different fragments that have been used to account for previously seen input items).

The Pitman–Yor process balances two principles in the number of fragments created: (1) a rich-get-richer bias—fragments used for a higher number of observed input items are more likely to be used to parse new items and (2) a novelty bias—higher chance of creating new fragment when more fragments have been created. The $\{< a^A, b^A >\}$ hyperparameters affect the two principles, respectively, by controlling the number of fragments created and how these are used in early stages of learning. For example, smaller values of a result in the model memoizing fewer fragments, for example, by giving less weight to K the number of fragments already created. As the number of observed input items increases, N increases, respectively, and becomes more significant in the memoization function compared with b. Thus, the model is more likely to memoize fragments for rule associations that are observed frequently. For example, since bare forms are often used in child-directed speech, each occurrence of bare form may lead the model to create the fragment $Verb \rightarrow$ null present. On the other hand, *–ed* is only used for only 25% of past tense use in the English portion of CHILDES, which results in the model more rarely observing this combination of rules.

Notably, the memoization step can lead to the creation of any fragment of the observed tree but cannot split terminal rules in the CFG. As such, the Stem+Inflection model can create the fragment $Verb \rightarrow$ –ed tense, which can only support the production of the verb *walk* with *–ed* but not associate *–ed* with past tense as a fragment. In addition to the examples provided in Table 1, other fragments may be created based on their usefulness for grammatical predictions. For example, the combined high frequency of the verb *go* with its irregular past tense form and meaning may lead to the memoization of a fragment that connects all three. We discuss the types of fragments generated for each model and their contribution to the observation in our analysis.

In our evaluations, we analyze the learning patterns by looking into differences in the grammar each model learns. The probability of the model predicting a value is calculated over all the fragments included in the grammar. For example, to predict the probability of the model generating *walked* versus *walk* for past tense, we consider all the possible fragment combinations that can generate the inflections *–ed* versus the bare form in the context of the stem *walk* and the tense "past." Similarly, we can ask whether the model is more likely to predict the tense meaning "past" or "present" given the context of a stem, for example, *walk*, and an inflection, for example, *–ed*.

More formally, the *inflection score*—the probability of predicting the inflection, I, given a tense, t, and a stem, s—is calculated as

$$P(\mathit{Inflection} = I | \mathit{Tense} = t, \mathit{Stem} = s) = \frac{\sum_f P(\mathit{Fragment} = f, \mathit{Inflection} = I, \mathit{Tense} = t, \mathit{Stem} = s)}{\sum_{I'} \sum_{f'} P(\mathit{Fragment} = f', \mathit{Inflection} = I', \mathit{Tense} = t, \mathit{Stem} = s)}. \quad (1)$$

Note that we compute $P(\mathit{Inflection} = -ed | \mathit{Tense} = past, \mathit{Stem} = s)$ by summing over all possible ways to generate the inflected verbs using the memoized fragments. For example, in the Independent model, one combination of fragments corresponds to the sum of the probabilities of each subpart of the tree (green, blue, red, and yellow) in Fig. 1a. Alternatively, the same form+meaning combination can be generated using the verb fragment with a fragment that associates the tense and rule, as shown in Fig. 8. The denominator sums over all possible inflection values, I', including bare form, *–ed*, –ing, –s, *and all irregular inflections*.

Similarly, the same approach can be used to predict the tense to evaluate the comprehension of an *–ed* inflected verb as past tense. That is, we will consider the possible values for the tense, t, rather than the inflection,

$$P(\mathit{Tense} = t | \mathit{Inflection} = I, \mathit{Stem} = s) = \frac{\sum_f P(\mathit{Fragment} = f, \mathit{Tense} = t, \mathit{Inflection} = I, \mathit{Stem} = s)}{\sum_{t'} \sum_{f'} P(\mathit{Fragment} = f', \mathit{Tense} = t', \mathit{Inflection} = I, \mathit{Stem} = s)}. \quad (2)$$

The denominator here sums over all tense values including past, present, etc. We refer to $p(\mathit{Tense} = past | \mathit{Inflection} = I, \mathit{Stem} = s)$ as the *past tense score*.

We calculate these probabilities for every stem included in the data that applies to the evaluation at hand and report the averaged probability for all stems.

3.1. Input and training

We extracted all verbs in the child-directed portion of the English CHILDES database for ages 18–60 months (MacWhinney, 2000). We modeled development using datasets of increasing size: each subset for ages 18–60 months is drawn from the respective month of child data and appended to the preceding months' data. Verbs were sampled according to their frequency in the data while preserving overall distribution of tense, inflection, and inflection regularity.

We created 10 independent samples representing 10 hypothetical children drawn from the same distributional properties but varying in observation of verb usages.[1] Thus, each of the

10 hypothetical children was modeled using 42 consecutive samples corresponding to data from 18 to 60 months. The samples range from 50 verbs for 18 months up to 5,000 verbs for 60 months. Note that, for all samples, the data hold naturalistic complexity, which implies *–ed* and past tense are not inherently mapped to each other. Past tense can be used with *–ed* but also with any of the grammatical irregular forms in the sample. In addition to being used for past tense, *–ed* can also be used for the past participle.

The hyperparameterization followed the settings proposed by O'Donnell (2015) (see the Appendix for the full set of parameters) with $a = 0.5$, $b = 100$, and the Dirichlet-multinomial parameter that controls the balance between memory use and fragment memoization set to 1 (the most balanced value). A random seed was used to initialize each simulation. Every simulation was run for a total of 50 sweeps. For each simulated task, we compute the predictions given to each modeled child and average across children similar to the within-subject design of experimental studies. We generate multiple samples, each representing child-directed speech (CDS) input to a theoretical child, to prevent the idiosyncracies of any specific sample from affecting our results. Similarly, since every run of a probabilistic model has some variation in the result, we run each sample multiple times. There were no systematic variations observable across samples and runs.

4. Simulations

We evaluate the models on their ability to replicate two aspects of the mixed productivity period. The first is children's *production* of regular and irregular verbs in past tense using a mixture of *–ed* and bare form of the verb. The second is children's preference for *–ed* inflected verbs for past tense meaning in their *comprehension*. Below, we describe our implementation and results for a production and a comprehension experiment.

4.1. Production simulations

We first look into the ability of each of the three models to replicate a production period with mixed use of *–ed* and bare forms for regular past tense verbs. We formalize this as expecting the model to produce similar probabilities for *–ed* and bare form inflections for the past tense meaning for a limited period in early stages of learning. Specifically, the model needs to only gradually increase the probability of producing *–ed* inflected verbs for the past tense.

We limit our analysis to contexts where the production goal is referring to a completed/past action. Therefore, for each regular verb stem, s, observed in the input, we measure the probability of the inflection, I, given the stem's meaning in past tense, normalized over all possible inflection choices. The probability of the form having inflection I given the context of past-tense meaning and a given stem, $P(Inflection = I | Tense = past, Stem = s)$, is calculated by summing over all the possible ways of generating this inflected form given the probability of the single and multi-rule fragments in the grammar. For example, the probability of *walk-ed*-past would sum over (1) the probability of the non-terminal fragment for the verb and the terminal rule-based fragments for the stem, inflection, and tense composing (Fig. 1a), (2) the

Fig. 2. The average score for the bare form and *–ed* inflection form for past tense productions of observed regular verbs for each of the three models. The Independent model predicts bare form as more or similarly likely to *–ed* for past tense before learning an *–ed* preference (left panel), the Inflection+Meaning model shows strong preference to *–ed* for past tense from the onset of learning (middle panel), and the Stem+Inflection model fails to learn *–ed* preference (right panel). The probabilities are marginalized over all possible inflections for the past tense. Irregular inflections are omitted from the figure for clarity.

two single-rule fragments for the stem and the tense and one multi-rule fragment associating the verb and inflection as shown in Fig. 6 (left), and (3) the single-rule fragment for the stem and the multi-rule fragment that associates the verb, inflection, and tense as shown in Fig. 8 (left), and so forth. The model may have different types of fragments that apply for each verb calculation. For example, if a verb is highly frequent in the past tense the model may memoize the entire tree as one fragment that can contribute to the overall probability in addition to every other combination of fragments. We normalize this probability as shown in Eq. 1 by summing over all possible inflection values for I' (bare form, *–ing*, *–ed*, *–s*, and every possible irregular inflection). We present the probabilities for I being either *–ed* or the bare form in the results section below. Similar to observations from child productions, the model's prediction of other inflections for the past tense (e.g., *–ing* or *–s*) is close to zero.

4.1.1. Regular verbs production experimental results

Our goal in these simulations is to replicate a period of mixed production of regular verbs with bare form and *–ed* inflected verbs for past tense. Importantly, we aim to replicate production of both forms by a single learning mechanism by exposing the models to CDS distributional properties exclusively (i.e., without encoding ungrammatical forms in the input). Fig. 2 shows the results for production of regular verbs with *–ed* or bare forms. We focus on production of observed regular verbs to align with psycholinguistic analysis. The Independent model is the only model to replicate both increased production of bare forms for past tense and the eventual memoization of a mapping between *–ed* and past tense. The model captures this through an early high probability of generating the bare form, which turns into a stage of nearly equal probability of generating verbs in the bare form or with an *–ed* inflection. The model then gradually learns the *–ed* inflection as the most likely to express past tense for regular verbs. This pattern matches evidence from children generating their first *–ed* inflected verbs soon after the comprehension of past tense and inflection meaning, but still alternating between the two forms. Our results are consistent with a higher probability of *–ed* inflection for verbs that are highly frequent in the past tense compared with other regular verbs, as discussed in more detail below. Notably, the model replicates the period while aiming to generate a grammar that best matches the input, which does not include bare forms used for past tense.

The Inflection+Meaning model fails to generate a heightened mixed productivity period in which bare form is produced for past tense meaning. This model knows the accurate form-meaning mapping from the start as a result of inherently representing inflections as tied to their meaning. Every co-occurrence of past tense and *–ed* can be tracked as opposed to the Independent model that tracks the inflection and tense independently until a multi-rule fragment is generated for them. Moreover, this representation prevents the production of verbs in the bare form for the past tense meaning. That is, bare forms are counted towards their grammatical use, for example, present tense, which prevents their probabilistic prediction for other tenses. We analyze the learned grammar and generated fragments below to further explain this process. This model highlights that the assumptions encoded in computational models may restrict their ability to replicate human behavior even when the encoded knowledge has not been hypothesized as central for the specific analysis. This result is consistent with previous studies that point to the importance of requiring form-meaning learning from a computational model in the acquisition of irregular forms (Arnon & Ramscar, 2012; Ramscar et al., 2013).

Finally, the Stem+Inflection model continuously prefers the bare form over the *–ed* inflected verbs. Thus, this model fails to learn an association of *–ed* form and past tense meaning independently of verb-specific observation. We hypothesize that the Stem+Inflection model predicts the most frequent inflection for a verb and fails to learn an overall pattern since it cannot generate a multi-rule fragment that contains the inflection and tense but not the verb.

For all three models, we predict that distributional patterns and the structure of the grammar lead to the observed result because of differences in the memoized fragments for each model. Although the bare form can be used for past tense meaning for verbs such as *cut*, *set*, and *put*, their extremely low frequency in past tense in CDS is not sufficient to lead to a memoization of this pattern or increased probability of its corresponding single-rule fragment. The results reflect an interaction of multiple distributional properties of forms, meanings, and their co-occurrences. Bare forms are the most commonly used form in child-directed speech. Our analysis of the North American English portion of all child-directed speech recorded in CHILDES shows that about 72% of verb tokens are with the bare form used for various tenses (MacWhinney, 2000). The Independent model does not simply reflect the frequency of the bare form apart from its grammatical context but simultaneously shows a delayed association of the past tense and the *–ed* form. Looking at the use of past tense in the same dataset of CDS, most verb types take the regular form in past tense using the *–ed* inflection but only about 25% of past tense tokens are of regular form. This observation is consistent with previous findings on regular past tense use in English (Legate & Yang, 2007; Marcus et al., 1992; Taatgen & Anderson, 2002). The rest of the verb tokens are spread across a wide array of irregular forms that change the form in respect to the stem. Thus, prediction of *–ed* for past tense is delayed compared with the overall productivity of bare forms. At the same time, we hypothesize that this association has the highest frequency among observed past tense inflections, which can make it strong enough for comprehension. Below, we evaluate this hypothesis by analyzing the model's comprehension behavior.

4.1.2. Irregular verbs production experimental results

Children often start by producing grammatical past tense forms for irregular verbs, such as *broke* and *fell* (Brown, 1973; Kuczaj, 1977). Computational models have mostly focused on analyzing the subsequent possible period of producing irregular verbs with *–ed* inflection, for example, *goed*. This finding offers an interesting example for U-shape learning, a period of erroneous production after observations of the correct form. However, none of the previously proposed computational models, to our knowledge, has looked into the production of irregular verbs with the bare form throughout the acquisition period. Importantly, the FGs model has been shown to replicate multiple properties of overgeneralization of irregular verbs for past tense (O'Donnell, 2015).[2] Moreover, many studies distinguish production of irregular forms from overgeneralizations as both forms are easily distinguishable in corpus analysis. However, fewer studies have taken the more laborious task of identifying erroneous productions of bare forms, which requires identifying whether this grammatical form is used for past tense context. Despite some early production of irregular forms, children produce bare forms for past tense meaning for *irregular verbs* for a longer period compared with regular verbs and more consistently across multiple children compare with overgeneralization period (Kuczaj, 1977; Marcus et al., 1992; Matthews & Theakston, 2006).

Thus, our goal in these simulations is to replicate a period of mixed production of irregular verbs with bare form and the grammatical irregular form for past tense. Results are shown in Fig. 3. For each of the irregular verbs, we consider the probability of the verb with the bare form and its irregular form. For example, we measure the probability of producing *go* for past tense meaning either as *go* or as *went*, while for *fall* we consider *fall* and *fell*. We omit the production of irregular forms with *–ed* as our model replicates the previous findings that are not relevant to the current discussion (O'Donnell, 2015).

The Independent model is the only model to replicate both increased production of bare forms for past tense for irregular verbs and the eventual production of the irregular forms for past tense. In comparison to the results for regular verbs, the model predicts a longer period of mixed production of bare forms and irregular forms for irregular verbs as discussed in psycholinguistic studies (Kuczaj, 1977; Marcus et al., 1992; Matthews & Theakston, 2006). We note that the model produces higher variability for irregular verb production with bare forms compared with regular verbs. This variability results from the interaction of the distributional properties of CDS and the memoization function of the model. Many irregular verbs are highly frequent compared with regular verbs, for example, *think*, *go*, and *fall*. Moreover, each irregular form is used primarily with one or a limited group of verbs. As a result, the model memoizes several irregular verbs with the full tree that generates their irregular

Fig. 3. The average score for the bare form and irregular inflection form for past tense productions of observed irregular verbs for each of the three models. The Independent model predicts bare form as more or similarly likely to the irregular form for past tense before learning an irregular form preference (left panel), the Inflection+Meaning model shows a strong preference for irregular forms for past tense from the onset of learning (middle panel), and the Stem+Inflection model fails to learn irregular form preference (right panel). The probabilities are marginalized over all possible inflections for the past tense. Regular inflections are omitted from the figure for clarity.

inflection, for example, *Verb* → *go went*-past. Such fully memoized trees increase the probability of producing the memoized verb-form-meaning combination over the bare form for the same verb. But, lower frequency irregular verbs remain more vulnerable to overgeneralization since (1) their irregular inflection form is either not memoized or of low probability due to its low frequency and (2) the alternative forms (bare form and *–ed*) are more likely given their higher productivity in the input. Overall, a few high-frequency irregular verbs are more likely to be produced with their irregular forms, but many low-frequency irregular verbs are more likely to be produced with the bare form for a longer period compared with regular verbs. This result replicates observations of irregular form and bare form production in child data.

The Inflection+Meaning model fails to generate a heightened mixed productivity period in which bare form is produced for past tense meaning for irregular verbs. Despite replicating many aspects of the overgeneralization period (O'Donnell, 2015), theInflection+Meaning model cannot replicate bare form production for irregular verbs since it inherently represents inflections as tied to their meaning. Even low-frequency irregular forms are learned with their respective meaning from the earliest stages. Finally, the Stem+Inflection model prefers the bare form over the irregular inflected verbs. This model improves on the prediction of higher frequency irregular verbs that are fully memoized with their form and meaning but fails to restrict errors for low-frequency irregular verbs more generally.

4.2. *Comprehension simulations*

In the first set of simulations, we showed that by requiring the model to learn the association between inflectional suffixes and their meaning, the Independent model goes through a period of predicting both bare forms and *–ed* to express the past tense meaning. Empirical evidence shows that even when children produce bare form for past tense, they still recognize *–ed* as the most likely suffix associated with past tense as measured in eye-gaze comprehension design (Wagner et al., 2009). In the second set of simulations, we aim to show that even though the model goes through a developmental stage where it sees *–ed* and bare forms as equally good in production, it shows a much stronger asymmetry in perception (as children do), inferring a past tense meaning only when it hears an *–ed* form. That is, *walked* would be more highly associated with past tense meaning compared with *walk* or *walking*, for example. For this purpose, we consider the situations where the child observes an inflected verb and needs to deduce what is the probability of it having a completed/past action meaning.

For each regular verb stem, s, observed in the input, we measured the probability that an inflection, I, would be interpreted as past tense compared with every other possible tense interpretation for these inflected forms of the verb. The probability $P(Tense = past|Inflection = I, Stem = s)$ is calculated by summing over all the possible ways of generating this inflected form given the probabilities recorded in the grammar. To illustrate, the probability of an inflected verb, for example, *walked*, *walk*, or *walking*, having a past tense meaning would sum over the probability of each possible way of generating it with past tense as the meaning. We normalize this probability as shown in Eq. 2 by summing over all possible values for t, for example, past tense, gerund or present participle, past participle, etc.

We exclude from this prediction verbs observed in irregular form or not observed in the past tense to directly quantify the model's behavior on familiar verbs. Following the experimental protocol of Wagner et al. (2009), we report the past tense probability of three inflections, *–ed*, bare form, and *–ing*, which was used as a competing inflection in their design.

4.2.1. Comprehension experimental results

Based on the preferential-looking paradigm, psycholinguistic findings show that children more strongly associate past tense events to *–ed* inflected verbs over other verb inflections (Wagner et al., 2009). We ask whether our model makes the same preference of interpreting *–ed* inflected verbs as past tense compared to the probability when using other inflections. The results for this experiment are shown in Fig. 4 including the probability of interpreting as past tense a verb inflected with either bare form, *–ed*, or *–ing*, over any other possible tense interpretations. As is evident from the left panel of Fig. 4, the Independent model shows a higher probability for the *–ed* inflection to be interpreted as past tense over the bare form and *–ing* inflections, even at stages where bare form and *–ed* form are competing for production. The Independent model learns this association over time and initially gives some non-zero probability for other inflections. These results are in line with the findings of Wagner et al. (2009) that children similarly master this mapping task over time.

The Inflection+Meaning model represents inflections and tenses as a single-rule fragment based on a single terminal rule, for example, "*Infl* → ed-past." Therefore, every occurrence of *–ed* with the past tense can be counted towards their joint probability even without creating a multi-rule fragment. Similarly, *–ing* and the bare forms are linked to their respective tenses in every use and cannot gain probability apart from their meaning. As a result, this baseline

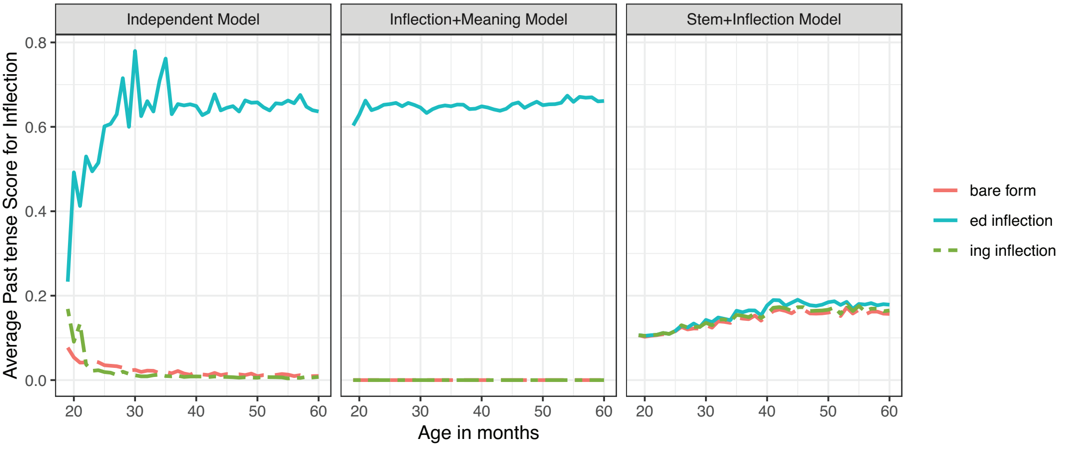

Fig. 4. The probability of each of the three inflections, bare form, *–ed*, and *–ing*, to be interpreted as past tense compared with all other possible tense interpretations. Even at stages that bare form and *–ed* forms are competing for production, the Independent model shows a higher probability for the *–ed* inflection to be interpreted as past tense compared with the bare form and *–ing* (left panel), the Inflection+Meaning model predicts only *–ed* from the onset of learning, failing to produce a gradual learning trajectory (middle panel), and the Stem+Inflection model fails to learn the *–ed* preference for the past tense (right panel).

predicts past tense as the correct interpretation for *–ed*, while eliminating any other inflection, even with extremely limited input (see the middle panel of Fig. 4).

Our additional baseline, the Stem+Inflection model, fails to learn *–ed* as the preferred form for past tense meaning (right panel of Fig. 4). This model learns only a marginally higher probability for the *–ed* inflection since the underlying structure cannot capture a generalized association of *–ed* to a specific tense regardless of the stem. Since this baseline represents the stems with their possible inflections, it relies on the memoization of full verb forms including stem, inflection, and tense to avoid misinterpretations of other inflections as past tense. The marginal advantage of *–ed* emerges from such memoization of highly frequent verbs but cannot be learned as a generalization of form-meaning association.

5. Grammar analysis

Our results above illustrate how each of the three simulated models yields a different trajectory of learning the preferred forms for past tense meaning. In this section, we look into the grammar that each of the models learns and how it can explain the above results. This analysis offers an understanding of what is considered grammatical by the simulated learner, whether bare forms are ever considered grammatical for past tense or rather predicted as a favorable production for other reasons and whether *–ed* is omitted or replaced by another inflection.

We distinguish between two types of fragments in the grammar of each model. First, the models learn the distributional properties of the single-rule fragments (see Fig. 1). Second, the model may memoize multi-rule fragments as part of its learning process. Such generalizations can only be achieved over time, as the model gathers information on what fragments of the tree can contribute to future computation if stored in the grammar. Below, we analyze the contribution of each type of fragment to the observations from each model. We show how each model reaches a form-meaning mapping at a different point in learning by relying on its assumed structure. We illustrate how the pathway to learning is guided by the type of fragments created by each model given the amount of input required to memoize each fragment.

5.1. Single-rule fragments

Our first analysis of the grammars looks at probabilities of individual rule fragments in each model, henceforth "single-rule fragments". The models rely on the rules included in the CFG to generate verbs until larger fragments (i.e., multi-rule fragments) are stored and become a preferable way of generating a form. Producing a form is analogous to a child producing the verb *walked* by deciding to express a "walking" meaning by adding the *–ed* inflection to the verb *walk* to represent its past meaning. The Independent and Stem+Inflection models can learn the probability of the inflection itself, independent of the stem or its tense meaning, as shown by the red-shaded rules in Fig. 1a and 1c. However, the most basic rule in the Inflection+Meaning model already associates the inflection to its past tense meaning (Fig. 1b). Thus, the Inflection+Meaning model does not allow for *–ed* rule storage without the corresponding meaning. On the one hand, this gives the Inflection+Meaning model an advantage in tracking the co-occurrence of inflections with grammatical meanings without having to store a multi-rule fragment. On the other hand, this rule structure limits this model to reproducing only observed form-meaning associations. It cannot represent the bare-form-past combination unless that specific combination occurs in its input. This could occur if the model observes the verb *put* in the past tense, for example. However, even if bare-form-past has been encountered, the Inflection+Meaning model cannot ascribe it higher probability as a result of encounters with other bare form usages or other past tense inflections, and, thus, the rule that combines bare form and past tense meaning remains low probability throughout the simulation.

We present the probability of single-rule fragments for each model in Fig. 5 for the fragments that reflect the use of *–ed* and bare form inflections. We refer to this as the *rule score*. The Independent model (left panel) learns a higher probability for the bare form fragment across the learning stages. This probability stems from the high frequency of the bare form in child-directed speech in English. Although the probabilities of the bare form and *–ed* inflection are initially close, the model learns an increasingly high probability for the bare form inflection while decreasing the probability of the *–ed* inflection for later stages. The high bare form probability provides support for competing prediction of bare form for past tense production in Section 2. The single-rule fragments alone do not offer an explanation for the eventual preference of *–ed* for production since the probability of *–ed* inflection decreases as more input is encountered. We will return to this aspect of the results in the next section.

The middle panel of Fig. 5 shows that the Inflection+Meaning model assigns a consistently low probability to the single-rule fragment associating *–ed* and past tense. This shows that the model does not increase the probability for this rule as more regular verbs inflected with *–ed*

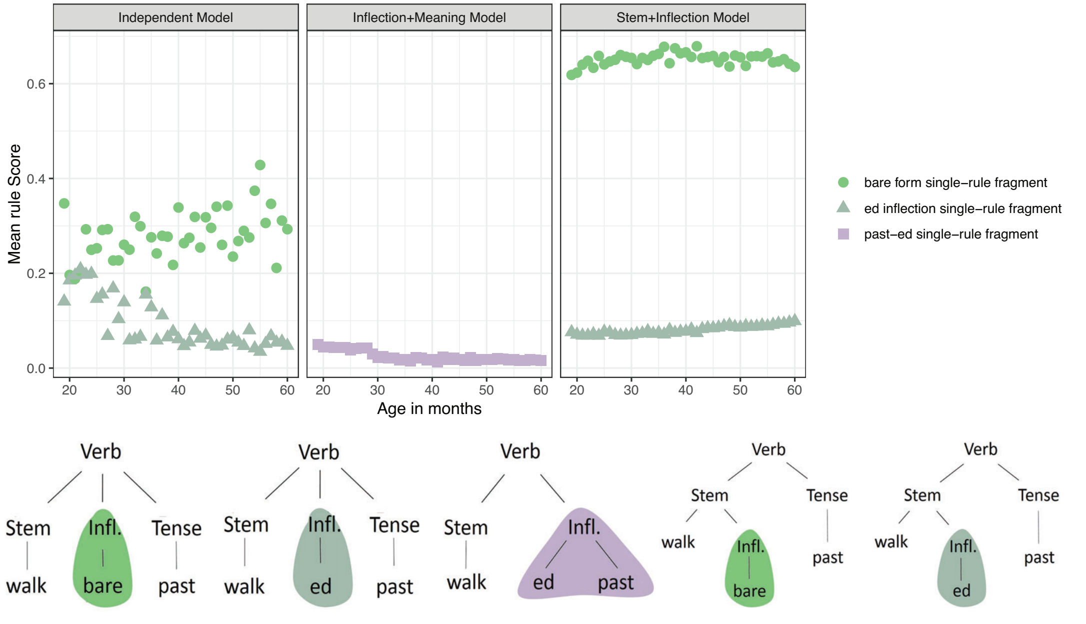

Fig. 5. Independent (left panel) bare form (green) and *–ed* (gray) fragments are initially close but gradually correspond to the frequency of the inflection in the input relative to the variability of inflections as the input size increases. Inflection+Meaning (middle panel) continuously denotes a low probability for the inflection fragment (purple), which implies the model relies on stored multi-rule fragments. Stem+Inflection (right panel) learns a probability that does not change much given more/less input and reflects an overall distribution of inflected verbs across tenses. Each type of fragment is illustrated below the panels (see Fig. 1 for details).

are observed. This model increases the probability of past-ed through a different mechanism despite having the *–ed* and past tense already associated with a single rule. However, this single-rule explains the near-zero probability of bare form and past tense for this model. In most cases, the model does not encounter a bare form used for the past tense and thus, does not have a single-rule fragment that enables the production of the bare form of verbs to express past tense.

On the right panel of Fig. 5, we see the probability of single-rule fragments for the Stem+Inflection model. This second baseline learns inflection probabilities independently as a single-rule fragment. Both *–ed* and bare forms have fairly consistent probability measures across input sizes. The probability of the inflections correlates with their frequency in the data regardless of tense meaning.

By requiring the model to learn the distributions of more fine-grained rules in the grammar, only the Independent model learns a probability distribution that supports the exhibition of the desired pattern. The two other models are limited in their ability to capture the variability in the language because of the strong assumptions these models make on what is already known. Though learning form-meaning associations is a well-attested part of language acquisition, this relation to the bare form production period has not previously been tested in a computational model. Because they assumed the form-meaning associations are already known, previous computational models required the implementation of a shift in the learning mechanism or in the input to replicate this period. Our proposed model produces this shift to more consistent use of *–ed* for past tense with a single learning mechanism over CDS input. Nevertheless, the replication of *–ed* production and comprehension in the Independent model cannot be explained solely by the single-rule fragment probability. To better understand the main results, we look into the multi-rule fragments stored by each model that can be used to parse a regular verb in the past tense with either *–ed* or in the bare form.

5.2. Multi-rule fragments

One powerful tool in children's language learning is the ability to recognize co-occurring patterns and track their behavior in the language (Saffran, Aslin, & Newport, 1996; Yu & Smith, 2007). The learning mechanism of FGs mimics this ability by allowing each model to memoize frequently occurring fragments of subparts of its CFG to create multi-rule fragments. Each model can only learn a combination of represented rules but cannot extract parts of a CFG rule into a novel fragment. Thus, the model tracks the distributional properties of rules as a whole and cannot trace subparts of a rule or combine parts of rules. For example, the Inflection+Meaning model observes the co-occurrence of bare forms and the present tense, but it cannot track the productivity of bare forms in a way that is dissociated from tense or associate the bare forms with other tenses it does not occur within the input. Although the Independent model needs to create a multi-rule fragment that associates bare forms and present tense through observations, it can reach the same level of association through experience while also maintaining an ability to learn more fine-grained distributional properties. Figs. 6–8 show examples of possible multi-rule fragments stored by the three models. The color scheme in the model and the following grammar analysis distinguish the type of information captured in each fragment: *–ed* fragment (yellow), bare form fragment (blue), past tense fragment (brown), and ed-past fragment (purple). This section looks into such multi-rule fragments, and how their probabilities affect the overall pattern of learning *–ed* versus bare form inflections for regular verbs in the past tense.

Fig. 9 shows the probability of stored fragments corresponding to multi-rule fragments. The figure only includes fragments that can be used to generate regular verbs in the past tense with either the *–ed* or the bare form inflections. We look at four types of multi-rule fragments: (1) *–ed* multi-rule (Fig. 6) fragment, (2) bare form multi-rule fragment (Fig. 6), (3) past tense multi-rule fragment (Fig. 7), and (4) *–ed*-past multi-rule fragment (Fig. 8). The first two fragments correspond to productive inflections, the third denotes a frequently used tense, while only the last multi-rule fragment type represents a full form-meaning association. Other inflections and verb-specific multi-rule fragments are omitted from this figure for clarity. We note that verb-specific fragments can lead to a higher production probability of *–ed* for past tense meaning if the fragment includes all parts of the generating tree, that is, stem, past tense, and the *–ed* inflection such as "*Verb* → *go went*-past". Although we observe this behavior to an extent in each of the three models, the parameterization of the Pitman–Yor function limits the memoization of such fragments to balance the use of memory versus the ability to

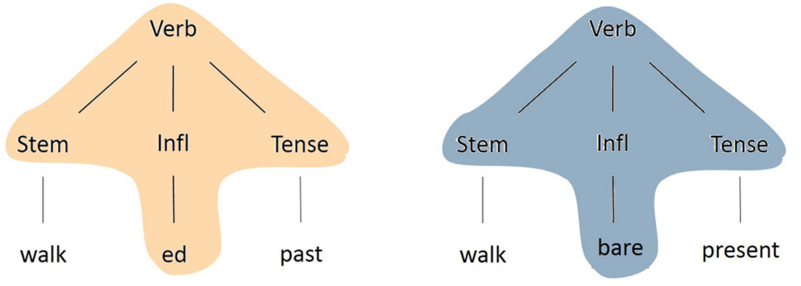

Fig. 6. Multi-rule fragments generated by the Independent mode recognizing an inflection as productive without associating it to a specific tense or verb for *–ed* (left), and bare form (right).

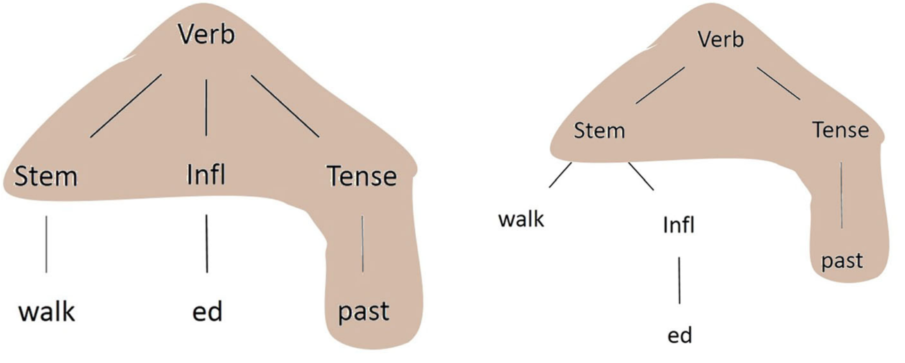

Fig. 7. Past tense multi-rule fragments recognizing the productivity of the past tense without associating it to a specific inflection, as memoized by the Independent model (left) and the Stem+Inflection model (right).

account for observed data. Such fragments are only stored for high-frequency verbs. As most high-frequency verbs are irregular, verb-specific fragments are mostly memoized for irregular verbs and thus have limited influence on the production and comprehension of *–ed* for past tense. As denoted by the equations used for evaluation (see Eq. 1), fragments that memoize inflections other than *–ed* and bare form are considered as part of the denominator. However, these fragments do not influence the difference in prediction of *–ed* versus bare forms as measured by the numerator of the equation.

The left panel of Fig. 9 shows that the Independent model initially stores two fragments that are more likely than others, a bare form fragment and a past tense fragment (Fig. 6 (right) and Fig. 7 (left), respectively). These fragments imply that the model recognizes certain productive patterns without fully breaking into the full form-meaning generalization. For example, the model could use the fragments to generate a past tense verb with "*Verb* → *Stem Infl* **past**" instead of "*Verb* → *Stem Infl Tense*." These fragments simplify the computation without offering generalization that associates form and meaning. First, the model learns that bare forms are frequent without knowing from which stem or for which tense. Second, the model

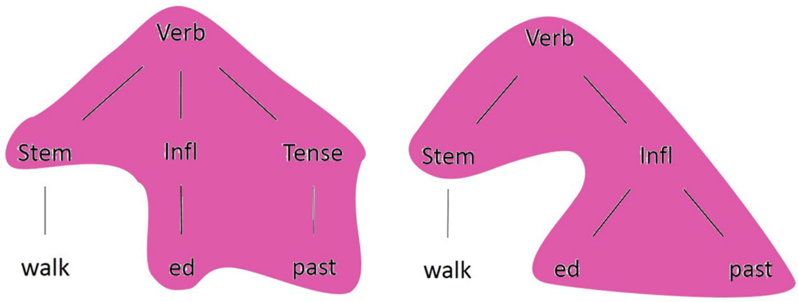

Fig. 8. Past-ed multi-rule fragments recognizing the productivity of *–ed* for past tense as generated by the Independent model (left) and the Inflection+Meaning model (right).

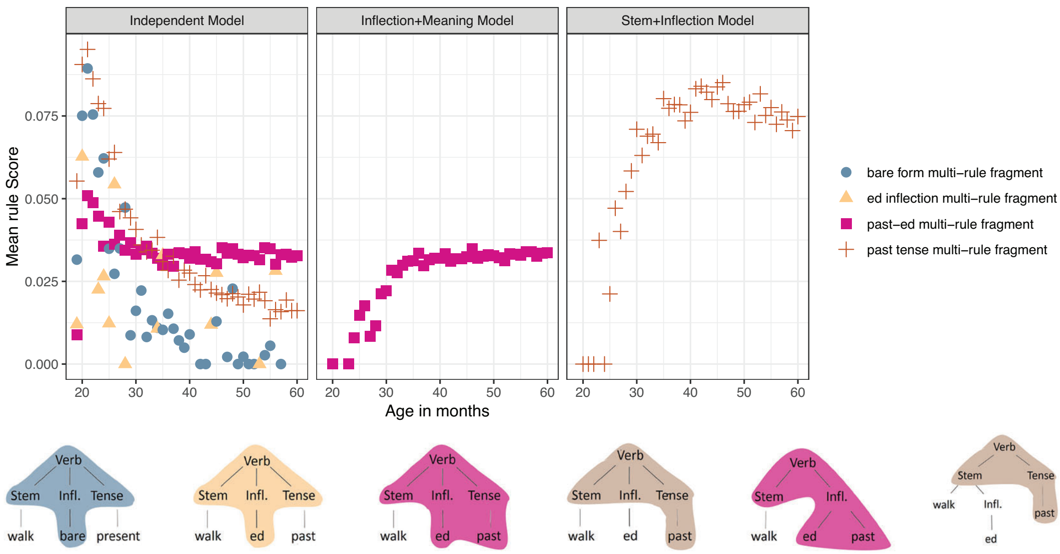

Fig. 9. Generalization rules. This figure shows for each model the probability of stored multi-rule fragments that are not verb-specific. While the models partially overlap in the type of multi-rule fragments they memoize, the fragments differ in probability (shown here) and structure (illustrated below the panels in the corresponding figures for reference). The Independent model stores inflection-only fragments illustrated in Fig. 6. The Independent model and the Stem+Inflection model memoize a past tense fragment as shown in Fig. 7. The Independent and Inflection+Meaning models memoize a past-ed fragment as shown in Fig. 8.

also learns that the past tense is a frequent meaning without associating it to a specific form of expression neither stems nor inflections.

The Independent model also stores an *–ed* fragment (Fig. 6 (left)), which corresponds to an *–ed* inflection not associated with any specific verb or tense. However, the probability of the *–ed* fragment is always lower than the probability of the other fragments for the bare form and the past tense and quickly becomes the least likely fragment. During the period when bare forms are preferred for past tense production, the bare form fragment suppresses the

–ed fragment as a possible inflection, while none of the fragments associates the inflection to a specific tense meaning. The model eventually ties the *–ed* inflection to the past tense by storing a fourth type of multi-rule fragment, the past-ed fragment, as shown in Fig. 8 (left). The past-ed fragment becomes more likely than the tense-only or inflection-only fragments over time. This shift marks the point where the model identifies the connection between the inflection and its common meaning, for example, *–ed* and the past tense. Despite the high frequency of the bare form, the model gradually reduces the probability for the bare form fragment until it approaches zero for bigger input sizes (later ages). We note that the model learns to associate the bare form with the grammatical tenses (other than past tense) through memoization of additional multi-rule fragments that are omitted from this figure for clarity. While these are not relevant for the production or comprehension of past tense directly, these bare form fragments result in limiting its probabilistic association for past tense because increasing the probability of fragments that associate the bare form with a particular meaning requires the model to correspondingly reduce the probability of the bare form fragment that occurs without an associated meaning.

The Inflection+Meaning model stores one type of multi-rule fragment that is relevant to the past tense acquisition trajectory, the *–ed* for a past tense fragment. Although the model assumes this association as part of its CFG rules, the model memoizes an additional fragment that reduces the number of required computations by putting together the root verb and form-tense rules as shown in Fig. 1c versus Fig. 8 (right). As noted in Section 4.1, the increasing probability of the *–ed* inflection for past tense for this model cannot be explained by the fairly constant probability of the single-rule fragment. Here, we see that the increased probability of the multi-rule fragment is a better explanation for the gradual increase in production probability. Importantly, for comprehension purposes, this model can always rely on either fragment, allowing it to correctly associate *–ed* and the past tense even from the earliest stages of learning.

The Stem+Inflection model also memoizes only one of the types of multi-rule fragments mentioned above. This model stores a tense-only fragment reflecting the observation of past tense productivity without learning its grammatical inflections. The model cannot store a fragment corresponding to a fragment of the tree that ties the tense meaning to the inflection independently of specific verb usages since its CFG structure represents stems as tied to their inflection. In other words, the model can observe that the inflected verb *walked* is composed of the verb *walk* and the inflection *–ed* and that together they mean past tense rather than learning the meaning for the inflection independently.

6. Discussion

We presented a computational model that accounts for a gradual learning process in the production and comprehension of past tense *–ed*. Our model replicates the developmental shift from bare form to *–ed* production for the past tense without encoding any corresponding change in the underlying learning mechanism. Like children, the model produces a form-meaning association that does not appear in the input while aiming to converge with the distributional properties of the observed language. This behavior provides a unique explanation of why children produce novel form-meaning associations that are seemingly erroneous and do not follow the observed grammar. The model replicates the erroneous production as a result of underspecifying the grammatical inflectional meanings for the bare form and not yet recognizing *–ed* as the most common inflection for past tense meaning. In later stages of learning, the model then conforms with grammatical productions by memoizing the relation of form (inflections) and meaning (tenses) for both *–ed* and the bare forms.

The proposed Independent model accounts for the mixed production period because of its assumption that children's productions stem from the need to learn the grammatical association of form and meaning from the linguistic input. Although the mathematical formulation of FGs does not encode specific linguistic information, tenses, or meanings, the chosen CFG rules used in O'Donnell (2015)—represented here as our Inflection+Meaning baseline model—imply an assumption that the learner already mapped inflections to tense. Our results show that requiring the model to learn what had previously been assumed as known information is the key to replicating developmental patterns. By representing the full complexity of naturalistic child-directed speech across various inflections and meanings, the model can capture how the learning patterns observed in children can arise as they learn to limit more accessible forms, such as the more frequent bare form, instead of considering the more productive form as competitive for the target meaning, that is, using bare form with past tense given the high probability of the bare form. Thus, our model explains the mixed bare form/*–ed* production period within a more general framework of language acquisition involving gradual learning and competition (Barak & Goldberg, 2017; Chater, McCauley, & Christiansen, 2016; Goldberg, 2019; Harmon & Kapatsinski, 2017; Hudson Kam & Newport, 2005; MacDonald, 2013). Our analysis explains why comprehension is not as compromised by the demands of form-meaning association as production is. The learner's comprehension leverages the form provided in the input, allowing the comprehender to ignore forms that are more accessible and focus on meanings associated with the given form, even if the association is weak.

Comparison with previous computational models. Our model makes a novel contribution with respect to previous computational models of this phenomenon by proposing a general learning framework that can be applied to additional morphology learning tasks and other languages. Several computational models have targeted the problem of bare form production. They cannot be directly compared with ours because they do not attack the problem of how the combination of bare form with past tense meaning arises in a learning model and also cannot straightforwardly be tested on the production and comprehension tasks used in this paper. Here we instead discuss the differences among these models in qualitative terms.

One model, the variational learning model, assumes that learning simply consists of collecting evidence for several competing grammars (Legate & Yang, 2007). The grammars themselves are innate to the learner, and the learner updates the relative strength of the different grammars over the course of learning using innate knowledge of the relevant parameters. A weakness of the variational learning model is that it is restricted to relying on a distinction between pre-defined grammars to match each developmental period. Our learning model is richer, in that it can build fragments that capture statistical co-occurrences in the input, and it is precisely this learning process that enables it to capture the behavior seen in children during development. Moreover, our simulations can compare production and comprehension abilities since the input consists of all possible forms and meanings, unlike the variational model that has simulated learning of past tense inflection independently of other tenses.

The MOSAIC model has also been used to study the mixed productivity period (Freudenthal et al., 2006, 2010, 2023). MOSAIC is a distributional learning model with strong utterance-final position bias. The model learns the probability of producing utterances as sequences of words while accounting for the inflection distribution in the input. Importantly, the model does not encode any semantic meaning of tense in the data or of the morphological inflections. Thus, the goal of the model is to reproduce the distribution across observed inflections in the data without evaluating whether bare forms have been produced to express a specific tense, for example, present or past tense (Freudenthal et al., 2006, 2010). Although this model offers a general framework that can and has been applied across linguistic tasks and languages, it lacks the ability to capture the learning of form and meaning that we argue forms the basis of children's behavior.

Finally, van Noord and Spenader (2015) propose a memory-based model that learns inflection classification for regular and irregular verbs for the English past tense. The model captures some gradual marking of regular forms, exhibiting variability in whether it correctly produces the *–ed* form or not. However, in cases where the model does not produce *–ed*, it does not specify which form it produces. Thus, the model is unable to specifically predict that children will use the bare form in English. Moreover, the model learns from past tense data exclusively, and errors are hence biased towards other past tense inflections rather than bare form production.

Relationship to psycholinguistic theories. Psycholinguistic studies have mainly interpreted unmarked verbs as either bare forms created by inflection omission or infinitival forms. Omission theories attribute such productions to dropping the inflection given the low saliency of *–ed* or given the higher frequency of the bare form, despite an overreaching knowledge of the grammatical rule (Bloom, Lifter, & Hafitz, 1980; Matthews & Theakston, 2006; Nicoladis et al., 2007; Owen Van Horne & Green Fager, 2015; Shirai & Andersen, 1995). Theories of infinitival production hypothesize that the child reaches a grammar that licenses the infinitival form for the target tense despite not observing such use in the input (Freudenthal et al., 2010; Legate & Yang, 2007; Räsänen et al., 2014; Wexler, 1998). Both theories have drawbacks that are resolved by our proposal. Inflection omission theories fit the English case of bare form production for past tense but fail to fit languages, such as French, where children produce verbs with infinitival marking rather than a bare form (Wexler, 1994). Theories of infinitival production relate to observations from a broader set of languages but assume an acquisition trajectory that is not supported by the language children observe.

Our results bridge these theories by allowing the grammar to gradually develop tense-form mapping representations, which enables an analysis of their validity over the learning trajectory. The high frequency of the bare form indeed guides our model in creating grammatical rules for efficient use of the form early on. Importantly, the model does not predict a default behavior but shows that comprehension still favors the grammatical *–ed*. Similarly, the model does not omit the *–ed* inflection intentionally. The bare form gains higher probability overall, regardless of intended meaning, which makes it more accessible rather than preferable for any tense. The model can only record and memorize fragments that are based on grammatical verb forms in the input and thus cannot license ungrammatical forms in the input. Rather than licensing ungrammatical form, our model produces bare forms for past tense as a result of not constraining the production sufficiently.

While bare form production characterizes a developmental stage in English morphology learning, the same is not necessarily true in other languages. Languages differ in their morphological structures, and many languages have more complex inflectional morphology than English. In those languages, production ranges over many possibilities, and bare forms are not necessarily the most frequent erroneous production. For example, children produce an infinitival form (with an infinitival marker) for past tense obligatory contexts in French and Dutch (Legate & Yang, 2007; Wexler, 1994) and affixed present tense in Hebrew (Ashkenazi, Ravid, & Gillis, 2016; Lustigman, 2015). Notably, our model predicts the production of bare forms not as an omission of a required inflection but as a higher probability of generating a more productive form. Moreover, the delay in producing the expected inflection stems from particular properties of the input: lower frequency of the form and variability of forms used for the same tense. This makes our framework suitable to analyze other languages because a variety of patterns of production may arise depending on the distribution of tenses, inflections, and form-meaning mappings of the given language. Recent work has already shown that it can be extended to capture the acquisition trajectory of Spanish verb inflection (Barak, Fernandez, Feldman, & Shafto, 2023).

Overall, our results highlight the importance of evaluating the learning of the production of a single form-meaning association within the context of naturalistic variable data. The association of past tense meaning to the *–ed* form is delayed by difficulty in restricting the most frequent and productive form in the input and the difficulty in identifying an association of less-frequent meaning to a less salient form. We show how the gradual learning of multiple form-meaning associations can explain the emerging ability of comprehension followed by production as the grammar develops given the properties of the observed language.

Acknowledgments

We thank Tim O’Donnell for sharing his code. This research was supported by NIH R21 DC017217.

Conflict of interest

We have no conflict of interest or financial relationships to disclose.

Notes

1 Data for training is deposited at the following link: https://github.com/CoDaS-Lab/BareProduction.
2 We have confirmed that the Independent model maintains this prediction ability.

References

Aguado-Orea, J., & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS ONE*, *10*(3), e0119613.

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.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292–305.

Ashkenazi, O., Ravid, D., & Gillis, S. (2016). Breaking into the Hebrew verb system: A learning problem. *First Language*, *36*(5), 505–524.

Barak, L., Fernandez, N., Feldman, N. H., & Shafto, P. (2023). Modeling substitution errors in Spanish morphology learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*. Advance online publication. https://escholarship.org/uc/item/740374zf

Barak, L., & Goldberg, A. (2017). Modeling the partial productivity of constructions. In *2017 AAAI spring symposium series* (pp. 131–138). Palo Alto, CA: AAAI Press.

Bloom, L., Lifter, K., & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language*, *56*, 386–412.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carr, L., & Johnston, J. (2001). Morphological cues to verb meaning. *Applied Psycholinguistics*, *22*(4), 601–618.

Chater, N., McCauley, S. M., & Christiansen, M. H. (2016). Language as skill: Intertwining comprehension and production. *Journal of Memory and Language*, *89*, 244–254.

Corkery, M., Matusevych, Y., & Goldwater, S. (2019). *Are we there yet? Encoder-decoder neural networks as cognitive models of English past tense inflection*. arXiv. https://doi.org/10.48550/arXiv.1906.01280

Cottrell, G. W., & Plunkett, K. (1994). Acquiring the mapping from meaning to sounds. *Connection Science*, *6*(4), 379–412.

del Prado Martın, F. M., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18.

Freudenthal, D., Gobet, F., & Pine, J. M. (2023). MOSAIC+: A crosslinguistic model of verb-marking errors in typically developing children and children with developmental language disorder. *Language Learning*. Advance online publication. https://doi.org/10.1111/lang.12580

Freudenthal, D., Pine, J., & Gobet, F. (2010). Explaining quantitative variation in the rate of optional infinitive errors across languages: a comparison of MOSAIC and the variational learning model. *Journal of Child Language*, *37*(3), 643.

Freudenthal, D., Pine, J. M., & Gobet, F. (2006). Modeling the development of children's use of optional infinitives in Dutch and English using MOSAIC. *Cognitive Science*, *30*(2), 277–310.

Freudenthal, D., Pine, J. M., Jones, G.,& Gobet, F. (2015). Defaulting effects contribute to the simulation of cross-linguistic differences in Optional Infinitive errors. In *Proceedings of the annual meeting of the Cognitive Science Society* (pp. 746–751) Austin, TX: Cognitive Science Society.

Freudenthal, D., Ramscar, M., Leonard, L. B., & Pine, J. M. (2021). Simulating the acquisition of verb inflection in typically developing children and children with developmental language disorder in English and Spanish. *Cognitive Science*, *45*(3), e12945.

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

Goldberg, A. E. (2019). *Explain me this*. Princeton, NJ: Princeton University Press.

Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, *56*(1), 61–98.

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

Hoeffner, J. (1992). Are rules a thing of the past? The acquisition of verbal morphology by an attractor network. In *Proceedings of the fourteenth annual conference of the Cognitive Science Society: July 29 to August 1, 1992, cognitive science program, Indiana University, Bloomington* (pp. 861–866). Red Hook, NY: Curran Associates Inc.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151–195.

Johnson, M., Griffiths, T. L., & Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models. In *Advances in neural information processing systems* (pp. 641–648). Cambridge, MA: MIT Press.

Kirov, C., & Cotterell, R. (2018). Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics*, *6*, 651–665.

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

Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, *14*(3), 315–344.

Leonard, L. B. (2014). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Lustigman, L. (2015). Non-finiteness in early Hebrew verbs. In *The acquisition of Hebrew phonology and morphology* (pp. 211–229). Leiden, the Netherlands: Brill.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. transcription format and programs* (Vol. 1). East Sussex, England: Psychology Press.

MacWhinney, B., & Leinbach, J. (1991). Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, *40*(1-2), 121–157.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*(4), 1–178.

Matthews, D. E., & Theakston, A. L. (2006). Errors of omission in English-speaking children's production of plurals and the past tense: The effects of frequency, phonology, and competition. *Cognitive Science*, *30*(6), 1027–1052.

Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science*, *10*(2), 237–254.

O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. Cambridge, MA: MIT Press.

Owen Van Horne, A. J., & Green Fager, M. (2015). Quantifying the relative contributions of lexical and phonological factors to regular past tense accuracy. *International Journal of Speech-Language Pathology*, *17*(6), 605–616.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, *28*(1-2), 73–193.

Pitman, J., & Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, *25*(2), 855–900.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, *23*(4), 463–490.

Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of "mouses" in adult speech. *Language*, *89*(4), 760–793.

Räsänen, S. H., Ambridge, B., & Pine, J. M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest-frequency form? *Journal of Child Language*, *41*(4), 756–779.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In *Parallel distributed processing: Explorations in the microstructure of cognition, vol. 2: Psychological and biological models* (pp. 216–271). Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Shirai, Y., & Andersen, R. W. (1995). The acquisition of tense-aspect morphology: A prototype account. *Language*, *71*(4), 743–762.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "broke"? A model of learning the past tense without feedback. *Cognition*, *86*(2), 123–155.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics*, *16*(1), 247–277.

van Noord, R., & Spenader, J. K. (2015). Modeling the learning of the English past tense with memory-based learning. *Computational Linguistics in the Netherlands Journal*, *5*, 65–80.

Wagner, L., Swensen, L. D., & Naigles, L. R. (2009). Children's early productivity with verbal morphology. *Cognitive Development*, *24*(3), 223–239.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot& N. Hornstein (Eds.), *Verb movement* (pp. 305–350). Cambridge, England: Cambridge University Press

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, *106*(1-4), 23–79.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

Appendix: Fragment grammars

The model underlying structure is a PCFG consisting of (N, T, P, S), where N is a finite set of nonterminal symbols, T is a set of terminal symbols, R is a set of production rules, and $S \in N$ is the start symbol. Each rule in R is of the form $A \rightarrow \beta$ where $A \in N$ and $\beta \in (N\&T)^*$ (see Table 1 for an example of possible terminal and non-terminal rules in the CFG). The CFG is restricted to the generation of input items from the rules in N and T assuming the rules' independence. However, this assumption has been found to be too strong to replicate language learning behavior. Instead, the FGs model uses a method of stochastically lazy evaluation, that is, memoization, to decide for each non-terminal rule whether to compute the rule or delay its computation. In other words, a CFG may generate the inflected verb *walked* using three rules, the non-terminal rule at the root of the tree and two terminal rules for the verb *walk* and the "ed-past" inflection (see Table 1). The FGs can make the computation more efficient by adding a last fragment shown in Table 1 by delaying the computation of the verb and storing a direct fragment that can inflect any verb with an *–ed* to express past tense. The stored fragments can be used in production to reduce the number of rules needed to generate an inflected verb, for example, *walked*.

Following Johnson, Griffiths, and Goldwater (2007), the model implements the stochastic memoization using the Pitman–Yor process, a generalization of the Dirichlet process (Pitman & Yor, 1997). The process can be understood similar to the Chinese restaurants process. At a first encounter of a value, the memoization function will choose a table. After this, the memoization function either chooses an observed value f with the probability $\frac{n_f - a}{N+b}$ or samples a new value with the probability $\frac{aK+b}{N+b}$, where n_f is the number of times value f has been used, N is the number of values sampled so far, and K is the number of times a new value has been sampled from the underlying function. a and b are the Pitman–Yor parameters governing the underlying distribution, where $0 \geq a \geq 1$ and $b > -a$. For a full description of the working of the process, please refer to Johnson et al. (2007), O'Donnell (2015), and Pitman and Yor (1997).

The FGs model learns the Pitman–Yor parameterization for each non-terminal and a beta-binomial distribution for the right side of each non-terminal. This allows the model to decide on which fragments to store given a specific input and rule structure. The FGs can be described as a tuple:

$$\mathcal{F} = < \mathcal{G}, \{\vec{\pi}^A\}_{A \in V_\mathcal{G}}, \{< a^A, b^A >\}_{A \in V_\mathcal{G}}, \{\vec{\psi}_B\}_{B \in \text{rhs}(r \in R_\mathcal{G})} > \tag{A.1}$$

where, $\mathcal{G}$ is a CFG, $\{\vec{\pi}^A\}_{A \in V_\mathcal{G}}$ are the vectors in the Diriclet-multinomial pseudocounts for each non-terminal, $\{< a^A, b^A >\}_{A \in V_\mathcal{G}}$ is the set of Pitman–Yor hyperparameters for each non-terminal, and $\{\vec{\psi}_B\}_{B \in \text{rhs}(r \in R_\mathcal{G})}$ is the set of pseudocounts for the beta-binomial distributions associated with the nonterminals on the right-hand side of each production rule in $\mathcal{G}$. Formally, the model can be described in the following stochastic equations:

$$G_A(t) = \begin{cases} \sum_{s \in prefix(d)} \text{mem}\{L_A\}(s) \prod_{i=1}^{n} G_{l(s'_i)}(s'_i), & l(t) = A \in N \\ 1 & l(t) = A \in T \end{cases}$$

$$L_A(t) = \sum_{r: A \rightarrow l(\hat{t}_i) \ldots l(\hat{t}_k)} \theta_r \prod_{i=1}^{k} \left[v_{r_i} G_{l(\hat{t}_i)}(\hat{t}_i) + (1 - v_{r_i})1 \right]$$

$$v_{r_i} \sim \text{BETA}(\psi_{\text{continue}}, \psi_{\text{delay}})$$

$$\vec{\theta_r} \sim \text{DIRICHLET}(\vec{\pi_A})$$

$$\text{mem}\{L_A\} \sim \text{PYP}(a_A, b_A, L_A).$$

That is, the model either uses the memoization function, for nonterminals $A \in N$ or stops at terminals $A \in T$. For non-terminals, the model uses the stochastically lazy recurrence, L_A , to generate fragments, as shown in Table 1, that partially compute the nonterminal rules. The recursion may be stopped with probability v_{r_i}. For a full description of the underlying mechanism of the FGs model, please refer to O'Donnell (2015).