Mind the Gap: Learning the Surface Forms of Movement Dependencies

Laurel Perkins[1], Naomi H. Feldman[2,3], Jeffrey Lidz[2]

[1]Department of Linguistics, University of California - Los Angeles

[2]Department of Linguistics, University of Maryland

[3]Institute for Advanced Computer Studies, University of Maryland

Author Note

Address for correspondence:

Laurel Perkins

3125 Campbell Hall

Los Angeles, CA 90025

perkinsl@ucla.edu

Abstract

In acquiring a syntax, children must detect evidence for abstract structural dependencies that can be realized in variable ways in the surface forms of sentences. In *What did David fix?*, learners must identify a non-local relation between a fronted object of the verb (*what*) and the phonologically null "gap" in canonical direct object position after the verb, where it is thematically interpreted. How do learners identify a non-adjacent dependency between an expression and something that has no overt phonological form?

We propose that identifying abstract syntactic dependencies requires statistical inference over both overt linguistic material and unsatisfied grammatical expectations: noticing when a predicted argument for a verb is unexpectedly missing may serve as evidence for the gap of an argument movement dependency. We provide computational support for this hypothesis. We develop a learner that uses predicted but unexpectedly missing objects of verbs to identify possible gaps of object movement, and identifies which surface morphosyntactic properties of sentences are correlated with these possible movement gaps. We find that it is in principle possible for a learner using this mechanism to identify the majority of sentences with object movement in child-directed English, and that prior knowledge of which verbs require objects provides an important guide for identifying which surface distributions characterize object movement. This provides a computational account for why verb argument structure knowledge developmentally precedes the acquisition of movement in a language like English. More broadly, these findings illustrate how statistical learning and learning from violated expectations can be combined to novel effect in the domain of language acquisition.

Keywords: language acquisition, computational modeling, statistical learning, expectation violation, non-adjacent dependencies, movement, argument structure

1 Introduction

In acquiring a syntax for their native language, children infer a system that specifies ways of combining expressions in hierarchical structures, and defines dependencies over those structures. These dependencies encode abstract grammatical relations, determined not by the specific form of any particular expression, but rather by the syntactic properties of expressions and their structural positions relative to each other.

For instance, the predicate-argument dependency between a verb and its direct object is established through a particular structural configuration (1a), and is the same regardless of the particular verb or the particular object noun phrase (underlined). And whereas in English this dependency is often established locally, between two adjacent expressions, the same abstract dependency can also be established non-locally, across potentially large amounts of linguistic material. In each of the sentences in (1b-1d), a fronted phrase bears the same object relation to the verb *fix* as does the corresponding phrase (*a toy*) in (1a), despite appearing in a non-adjacent position.

(1) a. David is fixing <u>a toy</u>. Amy is buying <u>a plane ticket</u>.

b. <u>What</u> did David fix?

c. <u>What</u> did the girl who we saw at the park say that David fixed?

d. I found <u>the toy</u> that David fixed.

These examples show us that syntactic dependencies are highly abstract in relation to the specific forms that express them. The same verb-object dependency can be satisfied by phrases with very different surface forms, appearing in very different positions in a sentence. And these dependencies take still different forms in other languages. This tension between the abstract nature of syntactic dependencies, and the variability of surface forms that realize them, presents a challenge for theories of how this central domain of syntax is acquired (Chomsky, 1965, 1980; Fodor, 1998; Lidz & Gagliardi, 2015; Pinker, 1984; Valian, 1990). How do language learners come to identify abstract structural relations in the face of such great variety in surface expression?

Prior accounts of dependency acquisition have largely focused on dependencies that are morphologically marked, such as the relation between the auxiliary verb *is* and the *-ing* form of the verb in (1a). Young children show awareness of the co-occurrence patterns of non-adjacent sounds and morphemes their input, statistical sensitivities that may allow them to discover morphosyntactic dependencies at early ages (Gómez, 2002; Gómez & Maye, 2005; Höhle, Schmitz, Santelmann, & Weissenborn, 2006; Nazzi, Barrière, Goyet, Kresh, & Legendre, 2011; Santelmann & Jusczyk, 1998; Tincoff, Santelmann, & Jusczyk, 2000; Van Heugten & Shi, 2010). But this represents only a narrow corner of the dependencies that learners must acquire. Here, we turn our attention to the sorts of dependencies illustrated in (1b-1d), in which an object is moved from its canonical position after the verb.[1] The abstract nature of movement dependencies poses a challenging learning problem. Identifying that the same verb-object dependency is present in (1a) and (1b-1d) requires tracking the co-occurrences not only of specific surface forms, but also of abstract syntactic categories and positions. Learners must become aware that a fronted noun phrase is standing in a non-local relation to something that has no overt phonological form: the "gap" associated with the verb, in canonical direct object position, where it is thematically interpreted.

In this paper, we argue that identifying abstract syntactic dependencies requires statistical inference over both overt and hidden grammatical structure. We pursue the hypothesis, consistent with a broader literature on the role of expectation violation in development (Denison & Xu, 2012; Kouider et al., 2015; Stahl & Feigenson, 2017, 2015; Téglás et al., 2011), that children learn from unsatisfied grammatical predictions. Our case study is the role of verb argument structure knowledge in the acquisition of argument movement. In their second year of life, children begin to identify subjects and objects in

[1] Here, "move" simply means that the relation between the object and the verb is established non-locally. Any syntactic theory needs to account for the fact that the same dependency can be satisfied both locally and non-locally. We use "move" as a theory-neutral term for this phenomenon.

their canonical positions, and to learn which verbs require objects (Lidz, White, & Baier, 2017; White & Lidz, 2022; Fisher, Jin, & Scott, 2019; Jin & Fisher, 2014; Yuan, Fisher, & Snedeker, 2012). Movement dependencies are acquired only after local argument structure knowledge has emerged (Gagliardi, Mease, & Lidz, 2016; Perkins & Lidz, 2020, 2021). This developmental trajectory points towards a particular learning mechanism: knowledge of local argument dependencies may help learners identify when arguments have been moved. If children notice when a predicted argument for a verb is missing in its expected position, this may compel them to search for that argument non-locally, and thereby learn the morphosyntactic footprints of particular movement dependencies in their language (Gagliardi et al., 2016; Perkins & Lidz, 2020; Perkins, 2019; Stromswold, 1995).

We provide computational support for this proposal. We develop a learner that identifies which surface morphosyntactic properties of sentences are correlated with expected but missing direct objects of verbs. In simulations on child-directed English, our model successfully identifies the majority of sentences with object movement in its input. Moreover, we show that prior argument structure knowledge plays a substantial role in the success of this distributional learning mechanism: knowledge of which verbs require objects provides an important guide for identifying which surface distributions characterize object movement. These findings provide insight into how learning from expected grammatical structure can work in concert with statistical learning to enable syntactic dependency acquisition in early development.

2 Acquiring Non-Local Syntactic Dependencies

A large body of literature finds that sensitivity to dependencies between non-adjacent sounds and morphemes develops in an infant's second year of life (Gómez, 2002; Gómez & Maye, 2005; Höhle et al., 2006; Nazzi et al., 2011; Santelmann & Jusczyk, 1998; Tincoff et al., 2000; Van Heugten & Shi, 2010). For instance, Santelmann and Jusczyk (1998) showed that 18-month-old English learners are aware of the dependency between *is* and *-ing* in sentences like *Everybody is baking bread.* Because these types of non-adjacent dependencies are morphologically marked, they leave detectable evidence on the surface forms of sentences that learners hear. That is, to identify that there is a dependency between *is* and *-ing*, learners need only notice that these sounds co-occur in their input with unusual regularity—although this still leaves open the question of how learners identify that this surface-level co-occurrence is marking a particular grammatical dependency, namely, the relation between the auxiliary *be* and a verb in the progressive aspect (Höhle et al., 2006; Nazzi et al., 2011; Santelmann & Jusczyk, 1998; Tincoff et al., 2000).

Other types of non-local syntactic dependencies, such as the argument movement dependencies in *wh*-questions, have received much less attention in prior work. These also pose a more substantial learning challenge. English *wh*-phrases have different surface forms than clause arguments in their canonical positions, and have different distributions: they overwhelmingly occur clause-initially. Therefore, recognizing that the same verb-object dependency is present in the *wh*-question in (1b) and in the basic transitive clause in (1a) requires abstracting away from these surface properties. Infants cannot merely track the co-occurrences of specific sounds or lexical items; they must represent the dependency abstractly, as an instance of the same dependency that is typically established locally between a verb and its direct object.

Prior experimental work has found that infants as young as 15 months sometimes respond appropriately to *wh*-questions (Seidl, Hollich, & Jusczyk, 2003; Gagliardi et al., 2016; Perkins & Lidz, 2020). But Gagliardi et al. (2016) and Perkins and Lidz (2020) argue that infants' success on these tasks may reflect an interpretive heuristic based on knowledge of local argument dependencies in combination with pragmatic reasoning, rather than syntactic representations of the non-local dependencies in these questions. This argument is motivated by earlier findings that children at 15 to 16 months show sensitivity to lexical and clause transitivity (Jin & Fisher, 2014; Lidz et al., 2017). Learners at this age are beginning to identify which verbs require direct objects (Lidz et al., 2017), and in the following months they gain facility in using this knowledge to predict upcoming direct objects during online sentence processing (Hirzel, Perkins, & Lidz, 2020; Lidz et al., 2017; White & Lidz, 2022). Infants in this age range also use subjects and objects to draw inferences about verb meaning, interpreting verbs with both subjects and objects as labels for causal events (Jin & Fisher, 2014). This early knowledge of local subject and object dependencies may lead to the appearance of *wh*-question comprehension in prior preferential looking tasks, even without representing *wh*-dependencies syntactically. Such tasks typically presented infants with *wh*-questions with transitive verbs, such as *Which dog did the cat bump?*, in the context of events in which e.g. a dog bumps a cat, and the cat bumps a different dog. A 15-month-old who can identify that *the cat* is the subject this question, and who knows that *bump* typically requires a direct object, may be inclined on the basis of that knowledge to look at an individual who got bumped by a cat— appearing to understand the question without necessarily representing *which dog* as a non-local object of the verb. In support of this account, Perkins and Lidz (2020) found that 15-month-olds' performance on this task depended on their vocabulary, a likely index of their verb knowledge.

Perkins and Lidz (2021) provided a more rigorous test of *wh*-dependency representations by asking when infants register the complementarity between between a local direct object and an object *wh*-phrase. If infants represent the *wh*-phrase in a sentence like (1b) as expressing the same grammatical relation as the local direct object in (1a), then they should be aware that the *wh*-phrase cannot co-occur with a local object: **What did David fix a toy* is ungrammatical. In a listening preference task, infants were presented with both *wh*-questions and basic declarative clauses with transitive verbs, with and without local direct objects. 18-month-olds listened longer to basic declarative sentences with local objects vs. without (e.g. *A dog! The cat should bump him!* > **A dog! The cat should bump!*), but displayed the opposite pattern of preference for *wh*-questions (e.g. *Which dog should the cat bump?* > **Which dog should the cat bump him?*). That is, 18-month-olds showed a consistent preference for grammatical sentences of each type. However, 14- and 15-month-olds did not differentiate between these sentence types. These results suggest that infants represent the *wh*-phrase as a non-local object of the verb at 18 months, but not before.

2.1 Learning Mechanisms

The experimental results surveyed above point towards the following developmental trajectory. Basic verb argument structure knowledge appears to develop early, at 15-16 months for English learners, and emerges before infants identify moved arguments, such as those in *wh*-questions. What learning mechanisms might allow learners to identify these non-local argument dependencies in their input? This is not a trivial task. Movement dependencies are not always marked with consistent morphology: for instance, English *wh*-phrases take a variety of different forms. The class of *wh*-elements in any language will distribute in specific ways in the surface forms of sentences: for instance, English *wh*-words are clause-initial and frequently occur in questions. However, even if a learner can identify a word class with these particular surface distributional properties, it does not necessarily follow that these are *wh*-elements. Many languages have question particles that can appear at sentence boundaries in both *wh*- and polar questions. An example is the particle *la* in Tz'utujil Mayan (2). A Tz'utujil learner needs a way to tell that *la* is a question particle and not a *wh*-word, and conversely an English learner needs a way to tell that *what* is a *wh*-word and not a question particle.

(2) *Tz'utujil Mayan* (Dayley, 1981)

La	xwari	ja	ch'uuch'?
Q	slept	the	baby

'Did the baby sleep?'

Moreover, in many languages, *wh*-phrases do not appear clause-initially. In *wh*-in-situ languages like Chinese, Japanese and Korean, *wh*-phrases are pronounced in their thematic position local to the verb, although on many accounts they still take scope in a higher clausal position through covert movement (e.g. Aoun, Hornstein, & Sportiche, 1981; Huang, 1982)[2]:

(3) *Mandarin Chinese* (Cheng, 2003)

Hufei	mai-le	shenme
Hufei	buy-PERF	what

'What did Hufei buy?'

Thus, in order to identify *wh*-dependencies in their language, children must solve multiple problems. They need to learn whether their language fronts *wh*-phrases, and if so, which surface forms signal that this movement has occurred. They also must identify the thematic position where the *wh*-phrase should be interpreted in relation to the verb. In English, surface signals for *wh*-movement include not only *wh*-words, but also a variety of other reflexes of movement, such as subject-auxiliary inversion and *do*-support in questions where the moved constituent is not a subject. Mature speakers of a language make efficient use of these signals in sentence processing to identify moved arguments and predict upcoming "gaps" where they should be interpreted (Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Frazier & d'Arcais, 1989; Frazier & Clifton, 1989; Sussman & Sedivy, 2003; Traxler & Pickering, 1996). But children must first learn these signals in order to use them in parsing *wh*-dependencies. In languages like English, identifying the tails of these dependencies is particularly challenging, because the thematic positions of moved elements are phonologically null. How do learners identify a non-adjacent dependency where only one element appears overtly?

One possible piece of the puzzle comes from the literature on "expectation violation" or "error-driven learning" in other areas of cognitive development. A large body of work has found that infants are capable of using knowledge about the physical and social properties of objects and agents, alone or in combination with learned statistical contingencies, to make

[2] Other non-movement accounts of *wh*-in-situ include binding by a covert operator (Reinhart, 1998), with some proposing different *wh*-in-situ representations across different languages (Cole & Hermon, 1994). See Cheng (2003) for an overview.

predictions about upcoming events (Denison & Xu, 2012; Kouider et al., 2015; Stahl & Feigenson, 2017, 2015; Téglás et al., 2011). Violations of these predictions may provide valuable opportunities for learning (Stahl & Feigenson, 2017, 2015). For instance, an experiment in Stahl and Feigenson (2015) presented 11-month-olds with events that either conformed with or violated object solidity. In one such event, a ball rolled down a ramp towards a solid wall, stopping behind an occluder. When the occluder was lifted, one group of infants saw that the ball had been stopped by the wall, while a second group of infants saw that the ball had apparently passed through the wall, violating their predictions about object solidity. After this event, both groups of infants were tested on their ability to map a novel property (e.g., squeaking) to the previously observed toy. Infants who had observed the prediction-violating event showed significantly greater learning than infants who had not. In a further experiment, infants who viewed these events were then given a choice to explore the ball or a novel object. Infants who had viewed the prediction-violating event chose to explore the ball more than infants who had not. Moreover, their exploration was consistent with testing the object's solidity properties: they banged the ball against the table to a greater extent than infants who had seen a different event type. These results suggest that even very young learners are sensitive to inconsistency between their own predictions and observed events, and when they observe a situation where their predictions are violated, they exploit this opportunity to learn, explore, and test hypotheses about the potential cause of that violation.

We pursue the hypothesis that a similar form of expectation-violation may underlie infants' discovery of argument movement dependencies in languages like English. Here, it is not predictions about physical events that drive learning, but rather predictions about grammatical structure. On this hypothesis, verb argument structure knowledge developmentally precedes argument movement acquisition because the former provides the basis for generating structural predictions— specifically, predictions about upcoming arguments of verbs. When infants encounter a case where an expected argument does not appear in its local position, they exploit this expectation violation to learn about the cause of the locally missing argument, scaffolding their identification of movement dependencies (Stromswold, 1995; Perkins, 2019; Perkins & Lidz, 2020; Gagliardi et al., 2016). For example, learners who know that a verb like *fix* requires a direct object might register that it is unexpectedly missing after the verb in a question like *What did David fix?*. This unsatisfied structural prediction may provide the basis of inferring the tail of a non-local argument dependency— a “gap” of argument movement— even though it is silent. And it may compel learners to search the rest of the sentence for the cause of the missing argument, eventually identifying that another expression in the sentence (*what*) is satisfying the verb’s transitivity requirement non-locally. This would allow them both to assign an appropriate parse to the sentence, and to begin to learn how various types of non-local dependencies are realized: i.e. that this question contains a *wh*-dependency, which is marked in English by various surface signals, such as *what*, *do*-support, and subject-auxiliary inversion.

In sum, we propose that the process of acquiring non-local dependencies follows three logically independent steps, which we will together call ***Gap-Driven Learning*** (Perkins & Lidz, 2020; Perkins, 2019):

(i) using knowledge of verb argument structure to detect argument gaps: predicted arguments that are unexpectedly missing in their local positions;

(ii) identifying what surface forms are correlated with these argument gaps; and

(iii) inferring what types of syntactic dependencies are responsible for those correlations.

Here, we investigate the Gap-Driven Learning hypothesis specifically in the domain of direct object gaps. This decision is motivated by empirical evidence for early knowledge of verb transitivity (Lidz et al., 2017; Jin & Fisher, 2014), making it plausible that direct object gaps are the type of argument gap that learners may be able to detect readily at the relevant stage of development. In support of this hypothesis, Perkins, Feldman, and Lidz (2022) show that it is computationally feasible for children to learn which verbs require objects even at stages of learning when they cannot yet identify objects that have been moved. The model in Perkins et al. (2022) assumes that it occasionally represents sentences erroneously, and it learns what portion of its input representations to treat as signal vs. noise for the purpose of learning verb transitivity. When tested on the distributions of direct objects in child-directed English, the model learned how to filter its data in order to correctly assign transitivity properties to the majority of the most frequent verbs in its input. This result tells us that it in principle possible for children to identify verb transitivity without accurately parsing argument movement, thereby providing a way for Gap-Driven Learning to get started.

In this paper, we present a computational model that instantiates the first two steps of learning under this hypothesis. The learner builds off of the model in Perkins et al. (2022), using the approximate verb transitivity knowledge that their learner identified. It tracks statistical regularities in the surface morphosyntactic features of sentences in order to identify clusters of sentences that share distributional properties. At the same time, it tracks when its expectations of upcoming direct objects are violated, in order to infer which clusters of properties are correlated with potential direct object gaps. When tested on child-directed speech, we find that the model identifies the large majority of sentences with object movement. Furthermore, we show that prior knowledge of verb transitivity, even if rough and approximate, is important for this distributional learning process to be successful. The learner performs better if it uses transitivity knowledge to infer likely object gaps, rather than clustering sentences on the basis of their overt surface features alone. These findings demonstrate that a learner could in principle identify object movement dependencies in English by using unsatisfied structural predictions to guide distributional learning. As verb transitivity knowledge forms the basis for generating these structural predictions, this provides an account for the empirically-attested order of argument structure and argument movement acquisition in early development.

3 Model

We present a Bayesian model that simultaneously tracks the statistical distributions of surface morphosyntactic features in sentences, and applies its knowledge of verb transitivity in order to infer which distributional properties are correlated with locally missing direct objects. This distributional learning takes the form of categorization: the learner infers "categories" of sentences according to their feature distributions, and infers which sentence categories likely contain direct object gaps. When the learner sees a sentence that violates its expectations about verb transitivity, the learner infers that that sentence contains a direct object gap, and that all other sentences in the distributionally-defined category do so as well. This allows the learner to generalize across sentences that share similar surface features, and to infer which of those shared features signal object movement dependencies.

This distributional learning mechanism follows prior computational work that has proposed similar mechanisms for the acquisition of phonetic categories in infancy, and for category learning domain-generally (Anderson & Matessa, 1990; Feldman, Griffiths, Goldwater, & Morgan, 2013; Maye, Werker, & Gerken, 2002; McMurray, Aslin, & Toscano, 2009; Sanborn, Griffiths, & Shiffrin, 2010). Similar to these previous models, the current account envisions the learning task as requiring two simultaneous inferences: discovering the underlying system of categories that give rise to distributions of surface features that a learner observes, and identifying which observations belong to which category. However, it departs from previous literature by envisioning this categorization process as merely a means to an end. Whereas the phonetic learning literature has traditionally assumed that there is a set of phonetic categories to be acquired (but see Feldman, Goldwater, Dupoux, & Schatz, 2021), here we do not assume that adult grammars necessarily represent "categories" of sentences in any meaningful way. Instead, the categories inferred by this learner are an intermediate step of learning: they enable further inference about the underlying properties of sentences that are formally similar. When the learner infers that one sentence in a category likely contains an object gap, it then infers that this property holds of other sentences in the category as well. In doing so, it identifies which surface features are correlated with object gaps and therefore may be the footprints of movement.

Following a rich tradition in the language acquisition literature (e.g., Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017; Alishahi & Stevenson, 2008; Berwick, 1985; Dillon, Dunbar, & Idsardi, 2013; Elman, 1990; Frank, Goodman, & Tenenbaum, 2009; Goldwater, Griffiths, & Johnson, 2009; Pearl & Sprouse, 2019; Perfors, Tenenbaum, & Wonnacott, 2010; Perfors, Tenenbaum, & Regier, 2011; Perkins et al., 2022; Sakas & Fodor, 2001, 2012; Vallabha, McClelland, Pons, Werker, & Amano, 2007; Wexler & Culicover, 1980; C. Yang, 2002), our model is framed at Marr's (1982) computational level. We aim to characterize a particular type of mental computation that could give rise to successful learning given the information available in children's data and a set of hypotheses about their knowledge at the relevant developmental stage. This model therefore represents an idealization of learners' actual inference processes, but an idealization that is nonetheless grounded in empirical data about their grammatical knowledge and representational abilities in development, described in more detail below. It also provides a measure of how much information is available in the child's representation of the input (at a particular stage of development) to support the hypothesized inferences. The results of our simulations open the door for further algorithmic questions concerning learners' abilities to access and use the information available in their environment, and whether their learning processes resemble this idealized mechanism.

In this section, we (i) specify the generative model, encoding the learner's assumptions about how its observations of sentence features are generated, and (ii) specify how the learner jointly infers sentence categories and object gaps, given its data and its knowledge of verb transitivity. The following sections present simulations demonstrating that this joint inference allows the learner to successfully identify features that characterize object movement dependencies in English, when tested on child-directed speech.

3.1 Generative Model

The data that our learner observes consists of the morphosyntactic features of sentences containing transitive, intransitive, or alternating verbs. The learner has approximate knowledge of these transitivity properties, as identified by the verb transitivity learner in Perkins et al. (2022). Intuitively, the learner assumes that there are two reasons why it might observe canonical direct objects or no direct objects after the verbs in these sentences. On the one hand, the transitivity of that verb determines whether it should always, never, or sometimes occur with a direct object. On the other hand, there may be a separate grammatical process, such as argument movement, that results in an apparent transitivity violation. The learner assumes that these transitivity violations are governed by latent "categories" of sentences with shared grammatical properties. Each category has a particular parameter governing whether it produces object gaps: if it does, then observations of canonical direct objects in that category may no longer reflect the transitivity properties of these verbs, but may instead be due to other grammatical properties that produce "non-basic" word orders. These properties also give rise to the distributions of other morphosyntactic features of sentences in a particular category.

For instance, the learner might identify that a sentence like *What did David fix?* belongs to a category of other sentences that have object gaps, and also tend to be questions with subject-auxiliary inversion, a form of *do*, and an unknown functional element sentence-initially (e.g., *what*). On the other hand, the learner might identify that a sentence like *Your toy got broken* belongs to another category of sentences that also have object gaps, but different morphosyntactic features: here, a form of *get* and the verbal suffix -*en*. The distributional features of the first sentence category are the footprints of object *wh*-questions in English; the features of the second category are the footprints of *get*-passives.

The learner does not know ahead of time how many sentence categories there will be, or what the properties of those categories are. Using the distributions of direct objects and the other observed sentence features in its data, the learner infers what categories of

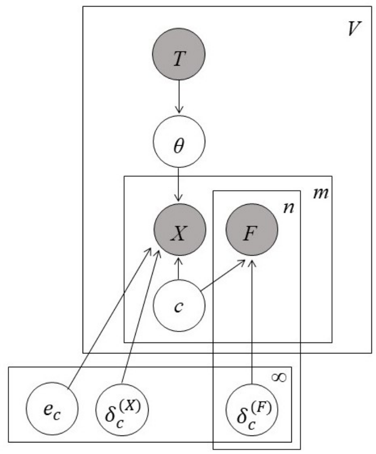

Figure 1. Graphical Model

sentences are present, what their distributional properties are, and which categories produce object gaps. This allows the learner to identify specific clusters of morphosyntactic features that are correlated with object gaps in different clause types, which may be candidates for entering into non-local movement dependencies.

More formally, we provide the graphical model for the learner in Figure 1. Observations of direct objects are formalized as the Bernoulli random variable X. Each $X^{(v)}$ encodes direct object data from a sentence containing verb v in the model's input, with a value of 1 if the sentence contains a direct object following the verb, and 0 if it does not. The model's observations of other relevant morphosyntactic features of the sentence are represented by the vector of Bernoulli random variables $\vec{F}$. Specific details of this feature set are discussed in the next section.

The direct object observations $X^{(v)}$ can be generated by two processes: the transitivity of verb v, represented by the variables T and θ in the upper half of the model, or the other grammatical properties of the category that the sentence belongs to, represented by the variables c, e, and $\delta^{(X)}$ in the lower half of the model. We describe each of these generative processes in turn.

In the upper part of the model, each observation $X^{(v)}$ of a direct object for a particular verb is conditioned on the parameter $\theta^{(v)}$, a continuous random variable that controls the probability that verb v will be used with a direct object. $\theta^{(v)}$ is conditioned on the variable $T^{(v)}$, a discrete random variable that can take on three values corresponding to transitive, intransitive, or alternating verbs. In order to model the hypothesis that learners are using prior knowledge of verb transitivity properties, we assume that the learner has approximate knowledge of these values of T for the set of verbs in the learner's data. This means that the learner knows some of the values of θ as well. If verb v is fully transitive, then the learner assumes that $\theta^{(v)} = 1$: the verb should always occur with a direct object. If the verb is fully intransitive, then $\theta^{(v)} = 0$: the verb should never occur with a direct object. If the verb belongs to the alternating category of T, then $\theta^{(v)}$ takes an unknown value between 0 and 1 inclusive. The prior probability over θ in this case is a $Beta(\alpha, \beta)$ distribution, where the parameters α and β are counts of direct objects and no direct objects for verb v in sentence categories without transitivity violations, excluding the current category.

In the lower part of the model, each $X^{(v)}$ is conditioned on the discrete random variable c, defined for all positive integers, which represents the category that the sentence belongs to. These sentence categories c also condition the other morphosyntactic features in the sentence, encoded in the vector $\vec{F}$. Each category c is assumed to reflect a particular set of underlying grammatical properties that give rise to the distributions of direct objects and other features of a sentence. The number and properties of these categories are *a priori* unknown, and the learner infers the properties that will allow it to explain the distributions of features and direct objects that it observes. Returning our earlier examples, the learner might infer a value of c that encodes English *wh*-object questions, giving high probability to sentence-initial function words (i.e., *wh*-words), subject-auxiliary inversion, forms of *do*, and direct object gaps. Another inferred value of c might encode English *get*-passives, giving high probability to direct object gaps, forms of *get*, and the *-en* verbal suffix. The prior probability over c is a Dirichlet process (Ferguson, 1973), which gives a particular category prior probability proportional to the number of sentence observations already assigned to that category. This process also reserves a small non-zero probability for new categories, allowing the model to flexibly converge on the number of sentence categories that best explains the distributions in its data. See Appendix A for details.

The random variables e, $\delta^{(X)}$, and $\delta^{(F)}$ represent the parameters of each sentence category. The Bernoulli random variable e_c encodes whether a given category c produces transitivity violations. If $e_c = 0$, then the category does not produce transitivity violations, and all observations of a direct object in $X^{(v)}$ were generated by the transitivity properties of verb v. But if $e_c = 1$, then the category does produce transitivity violations, and the observations of direct objects $X^{(v)}$ were generated by a particular grammatical property of category c. We build off of the knowledge gained by the learner in Perkins et al. (2022), which inferred that transitivity violations occurred approximately 19% of the time in sentences containing this same set of verbs in child-directed speech. We use this value as the prior probability that $e_c = 1$ in the current model.[3]

The random variable $\delta_c^{(X)}$ represents the probability of observing a direct object in a category with transitivity violations— that is, whether the particular violation in that category produces object gaps, or whether it adds an apparent extra object that isn't licensed by the verb. Intuitively, we can think of the probability that a sentence contains a direct object as depending on one of two biased coins. If $e_c = 0$ and the observation was generated by the verb's transitivity properties, then one biased coin is flipped and the sentence contains a direct object with probability $\theta^{(v)}$. But if $e_c = 1$ and the observation was generated by the grammatical properties of category c, then a different biased coin is flipped and the sentence contains a direct object with probability $\delta_c^{(X)}$. The parameter $\delta_c^{(X)}$ is assumed to have a uniform $Beta(1, 1)$ prior distribution. This uniform prior means that it is equally likely *a priori* for a sentence category to create object gaps as it is to add extra

[3] The model in Perkins et al. (2022) differs from the current model in that it did not group sentences into categories. In the previous model, this parameter represented the probability of transitivity violations across sentences in the corpus. In the current model, this parameter represents the probability of transitivity violations across categories of sentences. These two parameters are not necessarily equivalent; they will only be equivalent if sentences are equally distributed among sentence categories. Although this assumption may not be borne out, it is adopted here as a simplifying assumption of the learner's prior, which can be overridden as the learner updates its hypotheses upon seeing data.

objects, an assumption that is typologically odd but simplifies our model's inference. Analogous to $\delta_c^{(X)}$, the random variables in $\vec{\delta}_c^{(F)}$ represent the probabilities of observing the other morphosyntactic features in a given sentence category. Each $\delta_c^{(F)}$ is also assumed to have a uniform $Beta(1, 1)$ prior distribution, meaning that all features are equally likely *a priori* to be present as they are to be absent.

3.2 Inference

The learner uses component-wise Gibbs sampling (Geman & Geman, 1984) to jointly infer the category of each observed sentence (c) and whether or not each category contains transitivity violations (e). We first initialize values of c and e for each sentence. Then, for each sentence, we calculate a posterior probability distribution over new category assignments given the observed data in X and F, the known verb transitivity properties T, and the other sentence category assignments and properties. We re-sample new values of c for each sentence sequentially from this posterior probability distribution. Finally, we use the new category values to re-sample values of e for each category from its posterior probability distribution, given the other model parameters. This cycle is repeated over many iterations until the model converges to a stable distribution over c and e. Details of the initialization and sampling procedure are provided in Appendix A.

4 Simulations

We tested our learner on a dataset of child-directed English. We evaluated its performance by comparing it to two baseline models. Our model performs two steps of inference: it jointly categorizes sentences according to their surface feature distributions, and infers which sentence categories have direct object gaps. In order to assess the importance of each inference step, we constructed a baseline model that lacks one of these steps. The first baseline model uses verb transitivity knowledge to identify object gaps, but does not categorize sentences based on their feature distributions. The second baseline model categorizes sentences based on their feature distributions, but lacks verb transitivity

Table 1
Corpora of Child-Directed Speech

Corpus	**# Children**	**Ages**	**# Words**	**# Utterances**
Brown- Adam, Eve, & Sarah (Brown, 1973)	3	1;6-5;1	391,848	87,473
Soderstrom (Soderstrom et al., (2008))	2	0;6-1;0	90,608	24,130
Suppes (Suppes, 1974)	1	1;11-3;11	197,620	35,904
Valian (Valian, 1991)	21	1;9-2;8	123,112	25,551

knowledge and the ability to identify object gaps. We ask two primary questions: (i) how well can our learner identify instances of object movement in English, in comparison to these baselines? and (ii) how informative are the specific features of the model's categories for isolating movement dependencies from other grammatical processes?

4.1 Data

We prepared a dataset from four parsed corpora in the CHILDES Treebank (Pearl & Sprouse, 2013), which contains parse trees for child-directed English corpora on CHILDES (MacWhinney, 2000). Details of these corpora are provided in Table 1. From these corpora, we selected sentences containing the verbs whose transitivity properties are known by our learner. Because a child's knowledge of verb transitivity is likely to be highly imperfect before 18 months of age, we base our learner's knowledge on the transitivity classes inferred by the model in Perkins et al. (2022). We selected 18,503 sentences containing the verbs whose transitivity properties were inferred by the previous learner: these are the 50 most frequent transitive, intransitive, and alternating action verbs in these corpora. Because the previous learner assigned only 66% of these verbs to the correct transitivity category as specified in Perkins et al. (2022), this provides a noisy and imperfect source of knowledge for the current learner. Table 2 provides the frequencies of these verbs along with the transitivity categories assumed by our model.

We conducted an automated search over the Treebank trees for overt direct objects following each verb, as well as the morphosyntactic features of each sentence that our model

Table 2
Known Verbs and Transitivity Categories Assumed by Learner (T)

Verb	**Total**	**% Direct Objects**
Transitive		
feed	220	93%
fix	337	91%
pick	331	90%
bring	605	89%
drop	169	88%
throw	312	88%
hit	214	87%
lose	185	86%
close	166	85%
buy	358	84%
touch	183	84%
leave	356	83%
wash	195	83%
Alternating		
pull	331	81%
push	352	78%
open	342	77%
catch	185	76%
cut	263	75%
bite	191	73%
turn	485	72%
build	299	72%
knock	160	72%
hold	579	70%
read	509	69%
break	550	63%
Alternating, cont.		
drink	366	60%
wear	477	60%
eat	1318	59%
sing	306	53%
blow	255	52%
draw	375	51%
move	238	47%
ride	281	41%
hang	151	35%
stick	192	29%
write	583	27%
fit	227	22%
play	1568	19%
wait	383	15%
stand	294	7%
Intransitive		
run	228	6%
walk	253	4%
jump	197	4%
swim	180	4%
work	256	4%
cry	275	3%
sleep	451	3%
sit	859	1%
stay	308	1%
fall	605	0%

observes. We assume that the learner's inference is driven by information relevant to the predicate-argument structure of a sentence: morphosyntactic features pertaining to subjects, objects, and verbs. These features are listed in Table 3.

In selecting these features, we model a learner with the representational abilities of an infant between the ages of 15 and 18 months. Prior behavioral evidence finds that infants at these ages can use the word order properties of their language to identify clause subjects and objects in their canonical positions (Jin & Fisher, 2014; Hirsh-Pasek & Golinkoff, 1996; Seidl et al., 2003; Gagliardi et al., 2016; Perkins & Lidz, 2020; Lidz et al., 2017). They attend to auxiliaries, and can detect when the order of a subject and auxiliary is inverted (Geffen &

Type	**Features**
Object	Direct object of known verb is overt in canonical object position (right NP sister of V)
Subject	Subject of known verb is overt in canonical subject position (left NP sister of VP); sentence-initial; preceded by an auxiliary; preceded by another noun
Verb	Known verb is first verb in sentence; followed by a preposition or particle; has -*ed*, -*en*, -*ing*, -*s*, or irregular morphology
Tense & Auxiliaries	Verb is preceded by *to*, *be*, *have*, *get*, or *do*
Other	Question; unknown function word sentence-initially, sentence-medially before verb, sentence-medially after verb, or sentence-finally

Table 3
Direct Objects and Morphosyntactic Features Observed by Learner (X and F)

Mintz, 2015). They are able to segment a variety of verbal suffixes in English and other languages (Kim & Sundara, 2021; Mintz, 2013; Figueroa & Gerken, 2019; Santelmann & Jusczyk, 1998; Soderstrom, Wexler, & Jusczyk, 2002; Soderstrom, White, Conwell, & Morgan, 2007; Höhle et al., 2006; Nazzi et al., 2011; Van Heugten & Shi, 2010). In addition to auxiliaries and verbal affixes, infants at these ages are sensitive to the syntactic properties of a handful of other functional categories: determiners (Hicks, Maye, & Lidz, 2007; Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004; Shi & Melançon, 2010; Cauvet et al., 2014), pronouns (Cauvet et al., 2014), prepositions (Lidz et al., 2017), and negators (de Carvalho, Crimon, Barrault, Trueswell, & Christophe, 2021). Although they may not know the categories of other functional elements, they are able to recognize them as functional as opposed to lexical on the basis of their phonetic and prosodic properties (Monaghan, Chater, & Christiansen, 2005; Shi, Morgan, & Allopenna, 1998; Shi, Werker, & Morgan, 1999).

In coding for the features in in Table 3, we model an infant who can identify objects locally after verbs, but cannot yet identify non-local objects, such as fronted *wh*-phrases in *wh*-questions (Perkins & Lidz, 2021). This means that sentences like *You're eating* or *What are you eating?* were both coded as not having a direct object from our learner's perspective, even though the *wh*-word *what* acts as a non-local object in the second sentence of this pair.

Instead, *wh*-words are coded as “unknown function words,” a hyper-category that includes all functional elements assumed to be unknown at this age: *wh*-words, complementizers, quantifiers, focus particles, and conjunctions other than *and*.

We also code for the pragmatic feature “question,” which represents whether an utterance has interrogative force. Empirical evidence suggests that infants in their second year of life understand when a speaker is seeking information (Casillas & Frank, 2017; Goodhue, Hacquard, & Lidz, 2023; Luchkina, Sobel, & Morgan, 2018); see Carruthers (2018) on “questioning attitudes” as a basic component of human minds. They do so likely on the basis of distributional, prosodic, and socio-pragmatic cues (such as pauses and eye gaze) which differentiate questions from assertions in child-directed speech (Y. Yang, 2022). Young infants are sensitive to the prosodic and distributional differences between declaratives and polar questions (Frota, Butler, & Vigário, 2014; Geffen & Mintz, 2015; Soderstrom, Ko, & Nevzorova, 2011). Although *wh*-questions differ from polar questions in their prosody (Geffen & Mintz, 2017), it is possible that infants may know that these sentences are interrogatives, even before they are aware that they contain *wh*-dependencies (Seidl et al., 2003; Gagliardi et al., 2016; Perkins & Lidz, 2020). Questions were identified by the presence of a question mark in the transcription; this does not distinguish constituent questions from polar questions.

To verify the accuracy of our automated coding, a random sample of 500 sentences from the dataset were separately hand-coded by trained researcher. Percentage agreement between the hand-coding and automated coding was above 89% for all features.

The sentences in the dataset were also coded for their underlying clause types, listed in Table 4. These annotations were used as a gold standard to evaluate our model, and were not part of the model’s dataset. These clause types included three with movement: *wh*-questions, passives, and relative clauses. A given clause might be coded as multiple types, e.g. as both a question and a passive. For sentences with multiple clauses, coding was conducted for the clause containing the verb of interest. Accuracy of clause-type coded was

Clause Type	# Clauses	Description
Basic transitive	2855 (15%)	Matrix, finite, declarative clause with overt direct object following known verb
Basic intransitive	2704 (15%)	Matrix, finite, declarative clause without overt direct object following known verb
Wh-question	2336 (13%)	Clause has canonical syntactic form of a wh-question, with wh-element in a dependency with the known verb
Polar question	3641 (20%)	Clause has canonical syntactic form of a polar question
Other question	1922 (10%)	Clause was transcribed with a question mark, but does not have canonical syntactic form of a wh-question or polar question: includes tag, fragment, and echo questions, and rising intonation declaratives
Passive	268 (1%)	Known verb has been passivized, excluding forms that are clearly adjectival
Relative clause	298 (2%)	Known verb is in a full or reduced relative clause
Other embedded clause	4905 (27%)	Known verb is in a finite or non-finite embedded, non-relative clause
Imperative	2176 (12%)	Clause has canonical syntactic form of an imperative

Table 4
Distribution of Underlying Clause Types in Dataset

again evaluated by comparing against a hand-coded sample of 500 sentences. Percentage agreement between the hand-coding and automated coding was above 84% for all clause types. Additional hand-coding was conducted for *wh*-questions and relative clauses in order to annotate the gap site in these sentences, which could not be reliably identified automatically for the entire dataset.

4.2 Results

4.2.1 Sentence Category Distributions. The joint inference model inferred 38 total sentence categories, 15 with transitivity violations and 23 without. Of the 15 transitivity violating-categories, 13 had significantly lower odds of producing direct objects; we call these “object gap” categories. For each of the model’s categories, Figure 2 displays the proportion of the category made up of each underlying clause type. For example, the sentences in the model’s Category 1 are predominantly (97%) *wh*-questions; the sentences in Category 2 are predominantly both *wh*-questions and embedded clauses. Note that these

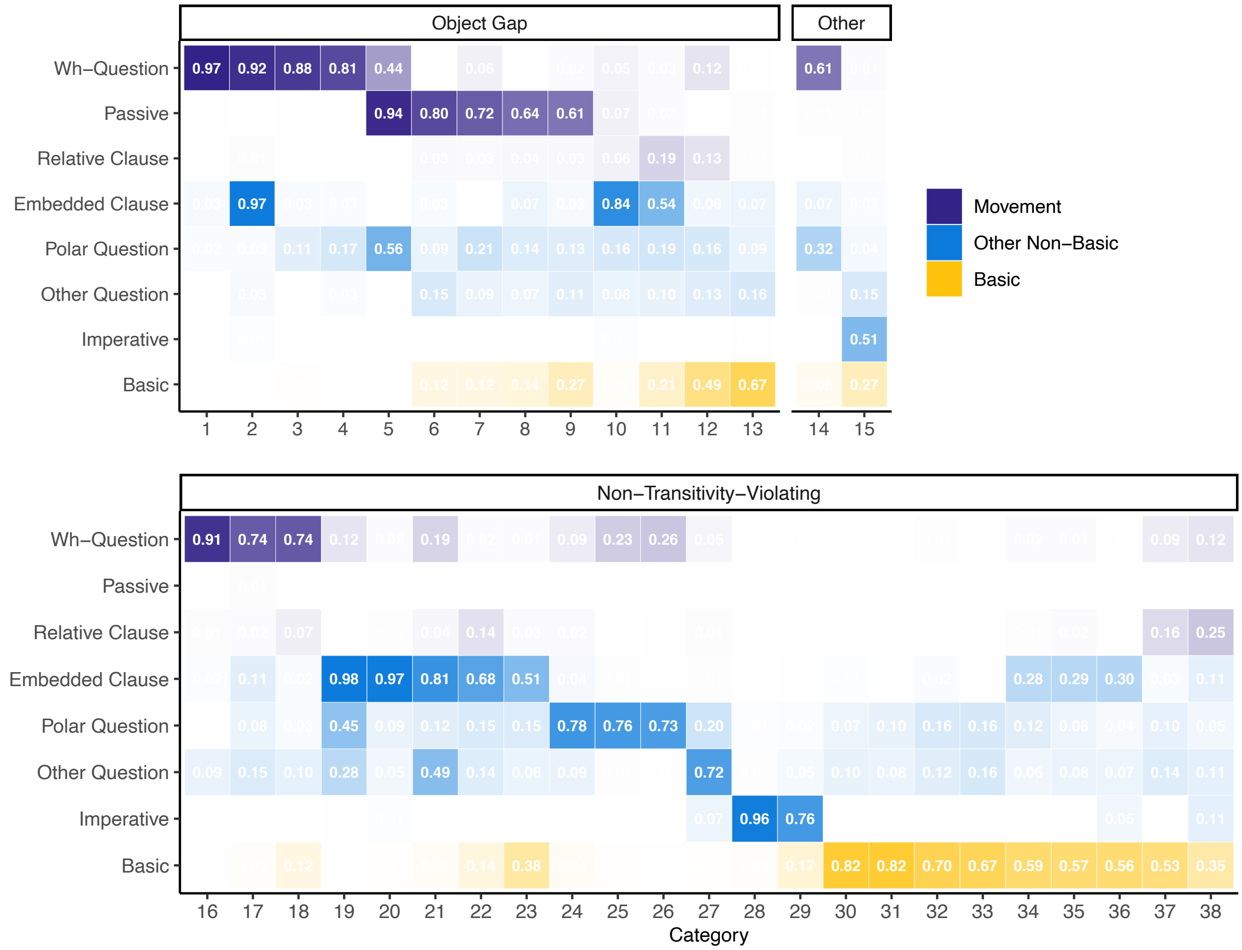

Figure 2. Proportions of Clause Types in Inferred Sentence Categories, Joint Inference Model

proportions do not necessarily sum to 1 because a single clause might be of multiple types.

In order to evaluate our model's categories, we first calculated their purity when compared to the true underlying clause types in the corpora. This measure calculates whether a particular category is predominantly made up of one single clause type, versus a mixture of different clause types. We calculated the overall purity of the model's categories by calculating the total number of sentences that belong to the predominant clause type in each category, and dividing by the total number of sentences in the dataset (Manning, Raghavan, & Schütze, 2008). When compared against a gold standard, this measure has a minimum value of 0 for poor clustering and a maximum value of 1 for perfect clustering. Our model's overall cluster purity is 0.76, which tells us that the model's categories were more

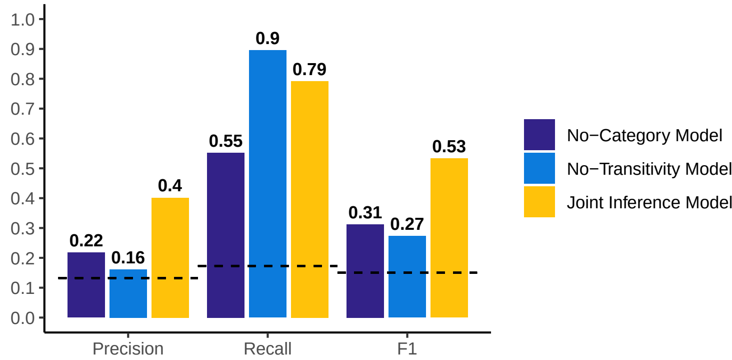

Figure 3. Accuracy on Identifying Sentences with Object Movement

likely to track one underlying clause type rather than a mixture.

The model inferred many more categories than necessary to identify the set of underlying clause types that it is being evaluated against. This is unsurprising: the learner was not given any information about how many clause type categories were present, nor the grain size at which to perform its categorization. Instead, it was given leeway to posit as many categories as needed to explain the distributions of features and transitivity violations in its data. The model divided *wh*-questions among eight different categories: five transitivity-violating categories and three with no transitivity violations. These categories differentiate monoclausal from biclausal questions (e.g., *What does he eat?* vs. *What would you like to read?*), questions in the progressive aspect (e.g., *What are you bringing?*) from those in other aspects, and questions where the *wh*-word is sentence-initial from those where it is not (e.g., *And what is he wearing?*). The model also categorized subject questions separately from object and adjunct questions, and correctly identified subject questions as non-transitivity-violating. These distinctions may have implications for the learner's ability to generalize about the surface forms that are distinctive of different types of movement dependencies, a point we return to in the following sections.

4.2.2 Accuracy on Identifying Object Movement. Here, we ask how well our learner can identify instances of object movement in its data. Visually, we can see from Figure 2 that clause types with movement were more likely to be categorized in object-gap

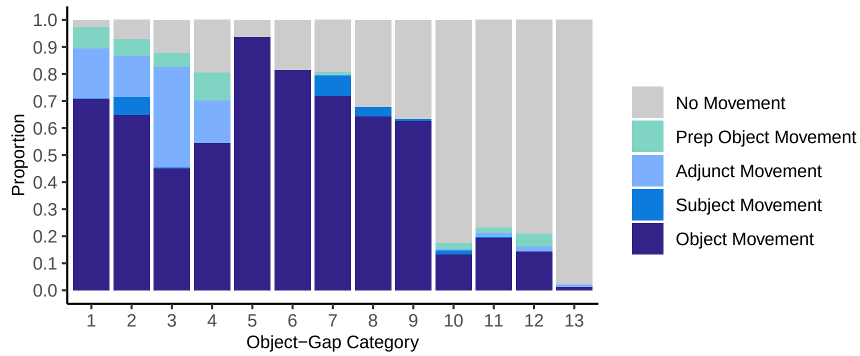

Figure 4. Distribution of Movement Types in Model's Object-Gap Categories

categories than in non-object-gap categories. To ask how well the model identified cases of object movement specifically, we compared its object-gap categories against the sentences that were coded as actually having object gaps in the corpus. The model's accuracy is displayed in Figure 3 using three metrics. Precision measures the proportion of sentences in the model's object-gap categories that contained object movement according to our gold standard— that is, the proportion of these categories made up of object *wh*-questions, object relative clauses, or passives. Recall measures the proportion of sentences with object movement in the corpus overall that were identified as belonging to one of the model's object-gap categories. These metrics are not always aligned: it would be possible to achieve perfect recall by identifying all sentences as having object movement, but this would result in very poor precision. The F1 score, the harmonic mean of precision and recall, reflects the model's overall accuracy by taking into account both of these metrics. For each of these metrics, we compare the model's performance to a chance baseline, indicated by the dashed horizontal line. This represents the expected performance of a learner that randomly categorizes sentences as having transitivity violations that cause direct objects gaps, by flipping a coin with weight 0.19, which is the probability of transitivity violations encoded in our learner's prior.

The model achieved an F1 score of 0.53. Its recall was 0.79, indicating that it identified 79% of sentences with object movement in its data. This accuracy rate is substantially above chance performance. Its precision was 0.40, indicating that on average, 40% of the sentences within its object-gap categories had instances of object movement. This precision rate is also above chance, but shows us that the model did not always manage to isolate object movement from other clause types in its data. To examine this further, we plotted the distribution of movement and non-movement types in the model's object-gap categories in Figure 4. Object movement was the predominant clause type in 69% of these categories, but occurred alongside other movement types as well, particularly adjunct movement. The other 31% of the model's object-gap categories were predominantly comprised of sentences without movement. Thus, while the learner achieved high accuracy in identifying sentences with object movement as such, in certain cases it categorized sentences with object movement together with other clause types.

The model achieves this performance despite several factors that limit its accuracy. First, the model does not receive credit for identifying cases of movement other than *wh*-questions, passives, and relative clauses; other rarer cases of movement were more difficult to code automatically, and thus were not annotated in the gold standard labels.[4] Second, the model only infers object movement from sentences that it believes violate verb transitivity: sentences with missing direct objects for verbs that it considers fully transitive. This means that the current evaluation measures how well the model was able to generalize from fully-transitive verbs to verbs that also allow intransitive uses. Table 5 displays the proportions of sentences with object movement that the model correctly identified as having object gaps, broken down by the verb classes that comprised the model's prior transitivity knowledge. The model achieved high recall even though the majority of sentences with object movement occurred with verbs that it believed to be alternating, rather than obligatorily transitive. Of the 1369 sentences coded as having object movement in the corpus, only 299 contained known transitive verbs, compared to 1055 containing known alternating verbs.[5] Nonetheless, the model achieved high accuracy across both the transitive and

[4] These rarer movement types included *tough*-movement, movement out of purposive clauses, clefting, pseudo-clefting, topicalization, and comparative movement.

[5] The few cases of object movement with intransitive verbs were uses of the verb in a rare or ungrammatical

Verb Class	# Object-Movement Sentences	% Identified
Transitives	299	0.72
Intransitives	15	0.33
Alternators	1055	0.82
Total	**1369**	**0.79**

Table 5
Proportion of Object-Movement Sentences Identified, by Verb Type

alternating verb classes. This tells us that it was able to generalize effectively: it used the presence of object gaps with known transitive verbs to identify the forms that object movement takes in its data, even with verbs that do not obligatorily require objects.

In summary, our joint inference model performed significantly higher than chance in categorizing sentences with object movement in its data. It achieved a high recall rate, indicating that it was correctly able to identify the large majority of sentences with object movement that it encountered. Its accuracy was high for both transitive and alternating verbs, indicating that it was able to use the presence of transitivity violations with fully-transitive verbs to identify direct object gaps with verbs that do not require objects. However, this object-gap inference produced a mixture of signal and noise: the sentences that the model categorized together with object movement also contained a variety of other movement and non-movement clause types. This has potential implications for how informative the learner's categories are for isolating object movement from other syntactic dependencies, a question we turn to next.

4.2.3 Identifying Distinctive Features of Object Movement. Under our hypothesis, the sentence categories inferred by the joint inference model are an intermediate step of learning. Jointly inferring how to categorize sentences according to their surface features, and which sentence categories contain object gaps, helps a learner identify the particular forms that characterize different types of object movement in the target language. Here, we ask how well the model identified which specific surface features are the footprints of object movement. To do this, we assessed which surface features are most distinctive in

transitive frame (e.g. *What did you run?*).

Category	Primary Clause Type	Distinctive Features
1	Wh-question	Subject is overt, preceded by an aux; verb is first in sentence, has *-ing*, preceded by *be*; sentence-initial function word; question
2	Wh-question	Verb is preceded by *to*; sentence-initial function word; question
3	Wh-question	Subject is overt, preceded by an aux; verb is first in sentence, preceded by *do*; sentence-initial function word; question
4	Wh-question	Subject is overt, preceded by an aux; verb is first in sentence, has *-ing*, preceded by *be*; sentence-medial function word before verb; question
5	Passive	Subject is overt, preceded by an aux; verb has *-en*, preceded by *get*; question
6	Passive	Subject (when overt) is sentence-initial; verb has *-en* or irregular form, preceded by *get*
7	Passive	Subject is overt, sentence-initial; verb is first in sentence, has *-en*, preceded by *be* or *have*
8	Passive	Subject is overt, preceded by an NP; verb has *-en* or irregular form, preceded by *be* or *have*, sentence-medial function word before verb
9	Passive	Subject (when overt) is sentence-initial; verb is first in sentence, has irregular form, preceded by *be* or *have*
10	Embedded	Verb preceded by *to* or *get*; sentence-medial function word before verb
11	Embedded	Subject is overt, preceded by an NP; verb has *-s* or irregular form; sentence-medial function word before or after verb
12	Basic	Subject is overt, preceded by an NP; verb has *-ing*, preceded by *be*; sentence-medial function word before verb
13	Basic	Subject is overt, sentence-initial; verb is first in sentence, has *-ed* or *-s*; sentence-medial function word after verb

Table 6
Features with Significantly Higher Odds in Object-Gap Categories

the categories that the model inferred to have object gaps. If these include the characteristic forms of English object movement dependencies, then the model's sentence categories contain helpful information for identifying the ways that object movement can be realized in English.

To assess feature distinctiveness, we calculated the odds ratio of each surface feature in the model's argument-gap categories. This measure divides the odds of observing a feature in a given category by the odds of observing that feature outside of that category; an odds ratio significantly greater than 1 indicates that a feature is more likely to be observed within than outside of the category. Significance was calculated using a Fisher's exact test with a Bonferroni correction for multiple comparisons.

Table 6 reports the features with odds ratios significantly greater than 1 for each of the model's object-gap categories (all *p*s < 0.001). Among these features are the characteristic forms of object movement dependencies in English. The categories that are predominantly *wh*-questions have greater odds of including subject-auxiliary inversion, *do*, and unknown function words sentence-initially or medially before the verb: these are *wh*-words. The categories predominantly made of passives have greater odds of including *get* or *be*, and -*en* or irregular verbal morphology.

However, the distinctive features of object-gap categories also include forms that are irrelevant to movement dependencies. These include many positional characteristics of subjects and verbs, but also some specific morphemes. For instance, *be* and *-ing* are distinctive of several of the model's *wh*-question categories, and *have* is distinctive of several of the model's passive categories. These features mark the realization of aspectual dependencies: *be* and *-ing* mark the progressive aspect, and the presence of *have* together with -*en* marks the perfect aspect. Thus, the model's categories contain both signal and noise for learning which surface features are the footprints of movement rather than other syntactic dependencies.

In summary, the current learner successfully identified the forms that characterize the most frequent types of movement in English, but it also identified some irrelevant features that are accidentally correlated with these forms. This invites the question of how a learner could effectively use this information for further steps of learning— how a learner could separate signal from noise by explaining some correlations as movement, and others as different dependencies. It is possible that a more sophisticated distributional learning mechanism might perform better. Further investigation is needed to determine whether the signal-to-noise ratio in the model's categories improves if it infers argument gaps using not only missing direct objects, but also other required but missing arguments (subjects and prepositional objects). This would give the learner the opportunity to identify non-object movement; it is an open question whether this could make its inference about categories with argument gaps more precise.

4.3 Model Comparisons

Our model achieves above-chance performance on identifying sentences with object movement by jointly inferring two properties: how sentences should be categorized together according to their surface feature distributions, and which sentence categories violate expectations about verb transitivity. To evaluate how important this joint inference is, we compare our model to baseline learners that only perform one step of inference at a time.

4.3.1 No-Category Baseline. If it didn't matter that our learner categorized sentences according to their surface features, then a learner should do just as well at identifying object movement on a sentence-by-sentence basis, by noting when objects are unexpectedly missing for known transitive verbs. To test whether the model's categorization process matters, we compared our model against a baseline learner that only used the presence or absence of direct objects in individual sentences, together with its knowledge of the transitivity properties of verbs in these sentences, to infer which sentences likely contain object gaps. This baseline learner has the architecture of the filtering model in Perkins et al. (2022), but we fix the the transitivity properties of each verb and the noise filter parameters to the values inferred by that learner. We then sampled transitivity violations for each sentence in the corpus from the posterior probability distribution over violations, given the observed direct objects and known model parameters.

To determine how well this "No-Category Baseline" identified movement, we compared the sentences without direct objects that it inferred to have transitivity violations against the actual cases of object movement in the corpus. Its precision, recall, and F1 score are reported in Figure 3. The model achieved above-chance accuracy overall, but scored substantially lower than the joint inference model on all three metrics. This because the baseline model's only source of reliable information for object gaps comes from the small percentage of verbs that it believes to be obligatorily transitive; it uses no other features in the sentences to inform this inference. If we examine its identification of object movement across verb classes, we find that it achieved high accuracy (74%) on identifying object movement with fully-transitive verbs. But for the much larger percentage of verbs that are alternating, it guesses at random which sentences contain gaps, identifying only 50% of object movement with these verbs. Thus, our joint inference model's ability to categorize sentences using a wide range of surface morphosyntactic features, and to generalize across sentences in a category, results in substantially better performance than inferring movement on a sentence-by-sentence basis from transitivity violations alone.

4.3.2 No-Transitivity Baseline. Our second baseline comparison investigates how much prior verb transitivity knowledge constrains the learner's identification of movement— specifically, how important it is that our learner uses transitivity violations in the process of categorizing sentences by their surface morphosyntactic features. We compare our model against a learner that performs this categorization without knowing which verbs require direct objects. It treats direct object observations identically to other surface features: for this learner, all direct objects are governed by the grammatical properties of a sentence category, not by the transitivity classes of verbs in the sentences. This learner therefore runs the risk of inferring categories that mix together sentences with movement and sentences without.

The architecture of this "No-Transitivity Baseline" assumes the lowest portion of the generative model in Figure 1, omitting the variables T, θ, and e. When the variables T and θ are omitted, the learner now assumes that all direct object observations X are generated by $\delta^{(X)}$, the grammatical properties of each sentence category, rather than by any properties of the verbs in these sentences. When the variable e is omitted, the learner no longer assumes that certain sentence categories contain transitivity violations. This means that its inference procedure consists of learning which sentence categories are present and which sentences belong to those categories, but no joint learning about transitivity violations in these categories. We sample category values for each sentence in the corpus from the posterior probability distribution over c given X and F, integrating over $\delta^{(X)}$ and $\delta^{(F)}$.

The No-Transitivity Baseline inferred 36 total categories. Of these, 19 had significantly

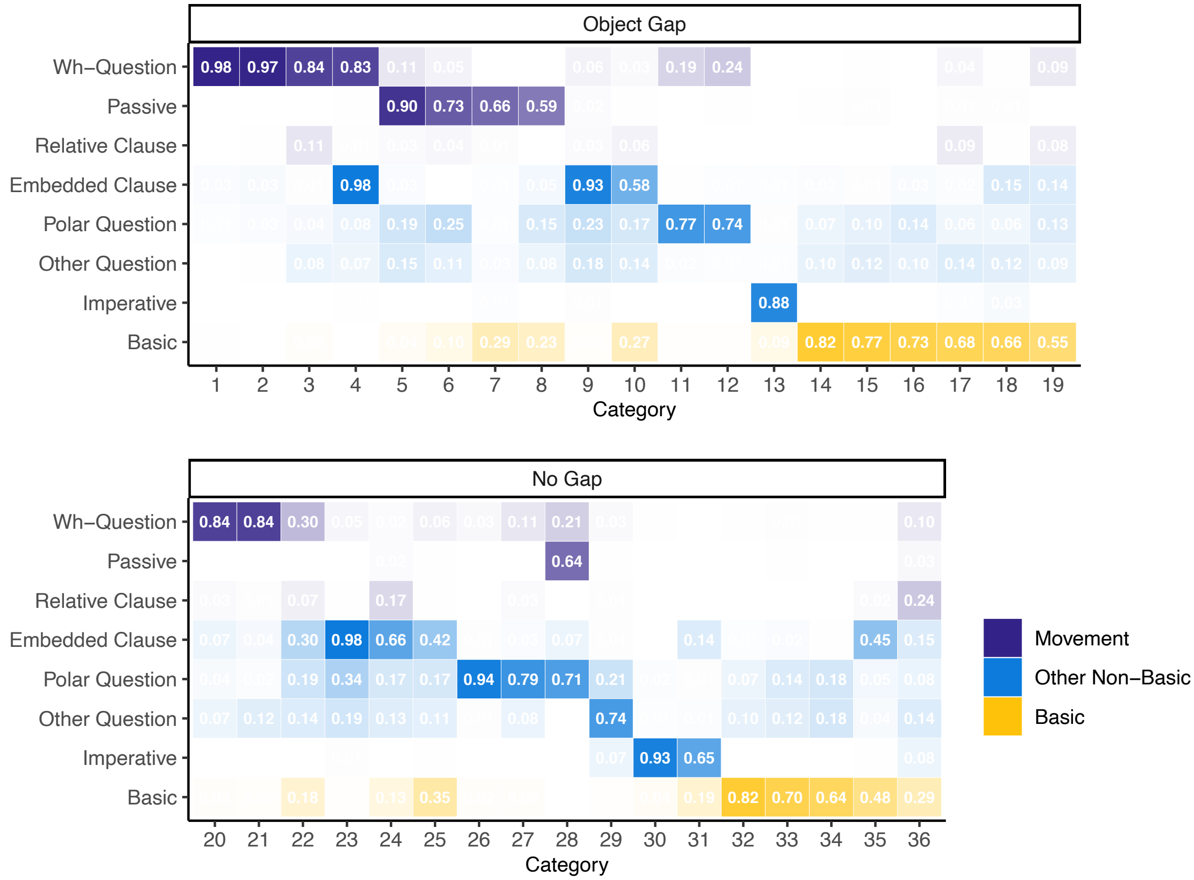

Figure 5. Proportions of Clause Types in Sentence Categories, No-Transitivity Baseline

lower odds of producing direct objects; we call these "object-gap" categories, under the assumption that these are the learner's candidate categories for object movement. The proportions of underlying clause types in the learner's categories are reported in Figure 5. These categories have similarly high purity to those inferred by the joint inference model: the baseline model's overall cluster purity is 0.77, compared to 0.76 for the joint inference model. This shows that the morphosyntactic features being tracked by both learners are informative for differentiating the different underlying clause types in the corpus, even without knowledge of which verbs require objects.

However, the baseline model's categories did not successfully differentiate sentences with movement from sentences without. The learner inferred many more sentence categories that were candidates for object movement, leading to slightly higher recall than our joint

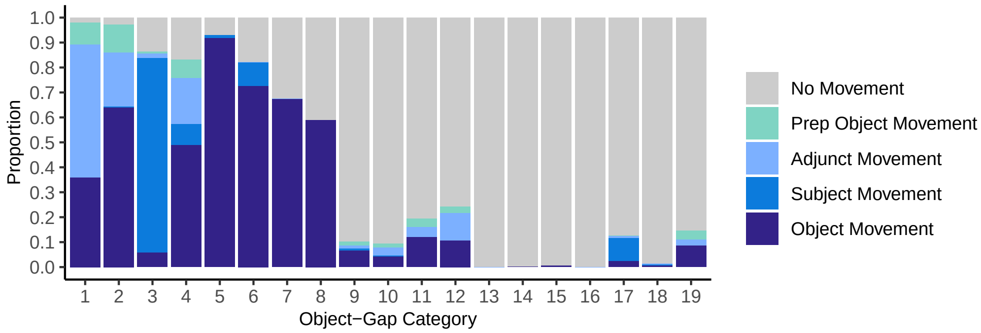

Figure 6. Distribution of Movement Types in Object-Gap Categories, No-Transitivity Baseline

inference learner (Figure 3). But its precision was quite poor, leading to a substantially worse F1 score. To examine the source of this worse precision, we plotted the distribution of movement and non-movement types in the model's object-gap categories in Figure 6. We find that object movement is the predominant clause type in only 32% of the learner's object-gap categories, compared to 66% in our joint inference learner. This tells us that our learner's ability to track transitivity violations is important for identifying categories of sentences with and without movement. While the distributions of morphosyntactic surface features of sentences convey a certain amount of information about the distinctions among different clause types, learning which of these distinctions signal movement, and which do not, requires the use of verb transitivity knowledge during distributional analysis.

4.4 Summary

In summary, our model identified approximately 80% of sentences with object movement in child-directed speech, by tracking the surface morphosyntactic features of sentences that violate its expectations of verb transitivity. The model jointly infers how to categorize sentences according to their surface feature distributions, and which of these sentence categories contain object gaps: unexpectedly missing objects of known verbs. This allowed the learner to generalize across sentences that share the same form and posit object gaps even for verbs that it does not know to be transitive. The learner performed substantially better than a baseline that relies only on known verb transitivity knowledge and does not categorize sentences on the basis of their surface feature distribution. This shows that the model's categorization process is important. It also out-performed a baseline that categorizes sentences using their surface features alone, without knowing which verbs require objects. The baseline learner performed substantially worse at differentiating sentences with and without object movement, showing that verb knowledge is an important guide for learning movement.

5 General Discussion

In order to acquire the system of syntactic dependencies in their language, children must detect evidence for abstract structure that is realized in highly variable ways within and across languages. Prior work has focused on how learners leverage statistical sensitivities to identify dependencies that are morphologically marked in their language (Gómez, 2002; Gómez & Maye, 2005; Höhle et al., 2006; Nazzi et al., 2011; Santelmann & Jusczyk, 1998; Tincoff et al., 2000; Van Heugten & Shi, 2010). But these statistical learning mechanisms face challenges when encountering the fuller range of syntactic dependency types that learners must acquire. Movement dependencies provide an extreme example, both in their degree of abstraction and the degree of overt evidence available on the surface forms of sentences. How do learners identify a non-adjacent dependency between a fronted expression and the "gap" of movement, which has no overt phonological form?

Here, we argue that solving this problem requires statistical learning not just over overt linguistic material, but also over hidden grammatical structure. Consistent with the literature on expectation-violation in other domains of cognition (Denison & Xu, 2012; Kouider et al., 2015; Stahl & Feigenson, 2017, 2015; Téglás et al., 2011), we pursue the hypothesis that statistical learning is informed by unsatisfied grammatical predictions. When a learner encounters an unexpectedly missing predicted argument of a verb, this may serve as evidence for a gap of an argument movement dependency. By tracking the surface forms that co-occur with these posited gap sites, learners may come to identify the distributional signatures of argument movement in the target language, enabling further inference about which specific syntactic dependencies underlie these surface forms. This hypothesis is motivated by prior empirical findings that knowledge of verb transitivity emerges before the identification of movement dependencies in infancy (Lidz et al., 2017; Jin & Fisher, 2014; Gagliardi et al., 2016; Perkins & Lidz, 2020, 2021).

Our findings demonstrate that this hypothesis is computationally feasible for the identification of object movement. Our learner jointly categorizes sentences according to similarities in their surface forms, and infers which of these sentence categories violate its expectations about verb transitivity. This joint inference allows it to accurately identify the majority of object movement in child-directed speech, and in doing so, to identify the formal properties that are the footprints of object movement in English. It performs substantially better than baseline learners that rely on only one of these two sources of information: either learning from verb transitivity violations without using surface morphosyntactic features of sentences, or learning from distributions of surface features with no knowledge of verb transitivity. This shows that the learner's expectations about hidden grammatical structure, coming from prior verb argument structure knowledge, place important constraints on its distributional learning mechanism. It thereby provides a computational account for why verb argument structure knowledge developmentally precedes the acquisition of movement in a language like English.

These findings raise two sorts of questions for future research. First, how does a learner take information about the formal correlates of object gaps in the language, and identify whether a particular form is participating in a movement dependency, versus another syntactic dependency? Our learner's inference yields both signal and noise for this next step of learning: the distinctive features of its object-gap categories include forms that characterize object movement in English, but also include forms that realize other non-movement dependencies. Separating signal from noise may require supplementing information from formal distributions with additional information about the likely dependencies in a given sentence and the ways that those dependencies can be realized, so that a learner can successfully factor out the features that realize other dependencies from those that realize movement.

Two relevant sources of information that are likely available to a young infant are prosody and pragmatics. Infants are sensitive to prosodic patterns from their first weeks of life (e.g. Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe, Mehler, & Sebastián-Gallés, 2001; Gerken, Jusczyk, & Mandel, 1994; Jusczyk et al., 1992; Nazzi, Bertoncini, & Mehler, 1998). Because prosodic breaks tend to fall at the edges of syntactic phrases, infants may be able to use this information to help identify some of the constituency structure of an utterance (Christophe, Millotte, Bernal, & Lidz, 2008; de Carvalho, He, Lidz, & Christophe, 2019; Gleitman, Gleitman, Landau, & Wanner, 1988; Gout, Christophe, & Morgan, 2004; Morgan, 1986; Morgan & Demuth, 1996). Infants also show early abilities to track the communicative intent of speakers (Csibra, 2010; Meltzoff, 1995; Woodward, 2009) and to identify the speech act of an utterance, at least at a coarse level of granularity (Casillas & Frank, 2017; Goodhue et al., 2023; Grosse, Behne, Carpenter, & Tomasello, 2010; Liszkowski, 2005; Luchkina et al., 2018). This speech act information might also provide useful information about the syntactic dependencies in a given sentence. However, as argued by Y. Yang (2022), it is likely that these other sources of information would need to work in tandem with the type of syntactically-guided distributional analysis proposed in the current work. Even a small amount of information about a speaker's communicative intent in using a particular sentence, along with the speaker's prosody, may help constrain the structure and interpretation that a learner assigns to that sentence. But this information is not by itself constraining enough to provide a complete parse; a learner must also have available a partial syntactic representation for which this top-down information could be useful. This invites further investigation into how statistical learning might be supplemented both by a child's developing knowledge of possible syntactic dependencies, and knowledge of how those dependencies relate to speakers' goals in discourse.

A second important question for future research is how the proposed learning mechanism might generalize cross-linguistically. Our learner uses expectations about the word order of English to detect when direct objects are missing in their canonical positions. This hinges on the assumption that learners at this stage of development have already acquired some knowledge of how their language marks canonical predicate-argument relations. Perkins & Hunter (2023) provide computational support for this assumption, but further empirical investigation is needed. In languages with a freer word order, other information, such as case morphology, may need to be recruited; see Fisher et al. (2019) and Suzuki and Kobayashi (2017) for evidence that Korean- and Japanese-learning 2-year-olds are sensitive to this information in verb learning.

Moreover, using argument gaps as evidence for movement dependencies requires at least a reasonable correlation between empty arguments and movement in a given language. This may be true for English, but will be complicated in languages that allow syntactic null arguments or *wh*-in-situ. In languages like Korean and Japanese, learners must come to identify that many of the argument gaps that they observe are null pronominals rather than the gaps of movement; conversely, English learners must rule out a null pronominal analysis in favor of movement. And learners of *wh*-in-situ languages will not be able to rely on argument gaps in order to identify *wh*-dependencies; instead, they must come to recognize such dependencies even when the *wh*-element has not overtly moved to the clause position where it takes scope (Aoun et al., 1981; Huang, 1982). It is possible that learners can more readily recognize when an in-situ *wh*-element bears a particular grammatical relation, but would need to use other formal, prosodic, or pragmatic information to recognize that this element is in a non-local dependency with a higher node in the clause, corresponding to the scope of the interrogative.

We suggest that the mechanism proposed here for English is one instance of a more general learning strategy that might be tailored to fit the evidence provided by a learner's data. Cross-linguistically, identifying canonical argument dependencies may be a necessary precursor to identifying non-local dependencies such as movement. An English learner may identify that word order provides a strong signal for canonical argument relations, and disruptions to this expected canonical word order signal that movement may be present. A Japanese learner may identify that case morphology is a better signal for these argument relations, that argument "gaps" occur with frequency that is more easily attributed to null pronominals rather than movement, and that overt and covert movement dependencies may be instead signalled by additional formal, prosodic, or pragmatic properties. In both cases, it is plausible that a learner's initial knowledge of the core predicate-argument structure of a clause provides an important grammatical scaffold for guiding future learning from the surface distributions in the data. This invites further empirical and computational work studying the developmental trajectory of argument structure and argument movement cross-linguistically.

More broadly, the current findings illustrate how two learning mechanisms with analogues in other areas of cognition— statistical learning and learning from expectation-violation— can be combined to novel effect in the domain of language acquisition. On this proposal, prior grammatical knowledge creates expectations that, when violated, form the basis for inferring hidden grammatical structure. Statistical learning may then be conducted over this hidden structure as well as more observable forms in the data. Here, we suggest that this combination provides a powerful foothold into syntactic dependency learning in early language development. This may also provide new avenues for understanding how incremental learning proceeds not only in language acquisition but also other domains of cognition, where predictions generated from knowledge acquired earlier in development form part of the data that learners use to draw new generalizations.

6 References

Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, *164*, 116–143.

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

Anderson, J. R., & Matessa, M. (1990). A rational analysis of categorization. In *Machine Learning Proceedings 1990* (pp. 76–84). Elsevier.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of memory and language*, *51*(1), 23–54.

Aoun, J., Hornstein, N., & Sportiche, D. (1981). Some aspects of wide scope quantification. *Journal of Linguistic Research*, *1*(3), 69–95.

Berwick, R. C. (1985). *The Acquisition of Syntactic Knowledge*. MIT Press.

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Carruthers, P. (2018). Basic questions. *Mind & Language*, *33*(2), 130–147.

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of memory and language*, *92*, 234–253. (Publisher: Elsevier)

Cauvet, E., Limissuri, R., Millotte, S., Skoruppa, K., Cabrol, D., & Christophe, A. (2014). Function words constrain on-line recognition of verbs and nouns in French 18-month-olds. *Language Learning and Development*, *10*(1), 1–18.

Cheng, L. L.-S. (2003). Wh-in-situ. *Glot International*, *7*(4), 103–109.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). On cognitive structures and their development: A reply to Piaget. In M. Piatelli-Palmarini (Ed.), *Language and Learning: The debate between Jean Piaget and Noam Chomsky* (pp. 35–54). Cambridge, MA: Harvard University Press.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994, March). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America*, *95*(3), 1570–1580.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001, July). Perception of Prosodic Boundary Correlates by Newborn Infants. *Infancy*, *2*(3), 385–394.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008, March). Bootstrapping Lexical and Syntactic Acquisition. *Language and Speech*, *51*(1-2), 61–75.

Cole, P., & Hermon, G. (1994). Is there LF wh-movement? *Linguistic inquiry*, *25*(2), 239–262.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers. In D. Dowty, D. Kartunnen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 94–128). Cambridge University Press.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168. (Publisher: Wiley Online Library)

Dayley, J. P. (1981). *Tzutujil grammar* (Doctoral dissertation). University of California, Berkeley.

de Carvalho, A., Crimon, C., Barrault, A., Trueswell, J., & Christophe, A. (2021). “Look! It is not a bamoule!”: 18-and 24-month-olds can use negative sentences to constrain their interpretation of novel word meanings. *Developmental Science*, *24*(4), e13085. (Publisher: Wiley Online Library)

de Carvalho, A., He, A. X., Lidz, J., & Christophe, A. (2019). Prosody and function words cue the acquisition of word meanings in 18-month-old infants. *Psychological Science*, *30*(3), 319–332. (Publisher: Sage Publications Sage CA: Los Angeles, CA)

Denison, S., & Xu, F. (2012). Probabilistic Inference in Human Infants. In F. Xu & T. Kushnir (Eds.), *Advances in Child Development and Behavior* (Vol. 43, pp. 27–58). JAI. doi: 10.1016/B978-0-12-397919-3.00002-2

Dillon, B., Dunbar, E., & Idsardi, W. (2013). A single-stage approach to learning phonological categories: Insights from Inuktitut. *Cognitive science*, *37*(2), 344–377.

(Publisher: Wiley Online Library)

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211. (Publisher: Wiley Online Library)

Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2021, November). Do Infants Really Learn Phonetic Categories? *Open Mind*, *5*, 113–131.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, *120*(4), 751-778.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209–230.

Figueroa, M., & Gerken, L. (2019). Experience with morphosyntactic paradigms allows toddlers to tacitly anticipate overregularized verb forms months before they produce them. *Cognition*, *191*, 103977.

Fisher, C., Jin, K.-S., & Scott, R. M. (2019). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77.

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, *27*(3), 339–374.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological science*, *20*(5), 578–585. (Publisher: SAGE Publications Sage CA: Los Angeles, CA)

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and cognitive processes*, *4*(2), 93–126.

Frazier, L., & d'Arcais, G. B. F. (1989). Filler driven parsing: A study of gap filling in Dutch. *Journal of memory and language*, *28*(3), 331–344.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy*, *19*(2), 194–213. (Publisher: Wiley Online Library)

Gagliardi, A., Mease, T. M., & Lidz, J. (2016). Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15-and 20-month-olds. *Language Acquisition*, *23*(3), 1–27.

Geffen, S., & Mintz, T. H. (2015). Can You Believe It? 12-Month-Olds Use Word Order to Distinguish Between Declaratives and Polar Interrogatives. *Language Learning and Development*, *11*(3), 270–284.

Geffen, S., & Mintz, T. H. (2017). Prosodic differences between declaratives and interrogatives in infant-directed speech. *Journal of child language*, *44*(4), 968–994. (Publisher: Cambridge University Press)

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*(6), 721–741.

Gerken, L., Jusczyk, P. W., & Mandel, D. R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*(3), 237–265.

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. *Linguistics: the Cambridge survey*, *3*, 150–193.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*(1), 21–54. (Publisher: Elsevier)

Goodhue, D., Hacquard, V., & Lidz, J. (2023). 18-month-olds understand the links between declaratives and assertions, and interrogatives and questions. In G. Paris & K. Jackson (Eds.), *Proceedings of the Boston University Conference on Language Development* (Vol. 47).

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, *51*(4), 548–567.

Grosse, G., Behne, T., Carpenter, M., & Tomasello, M. (2010). Infants communicate in order to be understood. *Developmental Psychology*, *46*(6), 1710. (Publisher: American

Psychological Association)

Gómez, R. L. (2002, September). Variability and Detection of Invariant Structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

Hicks, J., Maye, J., & Lidz, J. (2007, January). The role of function words in infants' syntactic categorization of novel words. Anaheim, CA.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 105–124). Cambridge, MA: The MIT Press.

Hirzel, M., Perkins, L., & Lidz, J. (2020). *19 Month-Olds Represent and Incrementally Parse Filler-Gap Dependencies* [Poster]. Amherst/Online. Retrieved from osf.io/v3k27

Huang, C.-T. J. (1982). *Logical relations in Chinese and the theory of grammar* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Höhle, B., Schmitz, M., Santelmann, L. M., & Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development*, *2*(4), 277–300.

Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, *5*(3), 341–353.

Jin, K.-S., & Fisher, C. (2014). Early evidence for syntactic bootstrapping: 15-month-olds use sentence structure in verb learning. In *Proceedings of the 38th Boston University Conference on Language Development.* Boston, MA: Cascadilla Press.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992, April). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*(2), 252–293.

Kim, Y. J., & Sundara, M. (2021, July). 6–month–olds are sensitive to English morphology. *Developmental Science*, *24*(4), e13089. doi: 10.1111/desc.13089

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015, October). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*(1), 8537. (Number: 1 Publisher: Nature Publishing Group) doi: 10.1038/ncomms9537

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist.*, *1*(1), 333–353.

Lidz, J., White, A. S., & Baier, R. (2017). The role of incremental parsing in syntactically conditioned word learning. *Cognitive Psychology*, *97*, 62–78.

Liszkowski, U. (2005). Human twelve-month-olds point cooperatively to share interest with and helpfully provide information for a communicative partner. *Gesture*, *5*(1-2), 135–154. (Publisher: John Benjamins)

Luchkina, E., Sobel, D. M., & Morgan, J. L. (2018, November). Eighteen-month-olds selectively generalize words from accurate speakers to novel contexts. *Developmental Science*, *21*(6), e12663.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval.* Cambridge University Press.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: W. H. Freeman and Company.

Maye, J., Werker, J. F., & Gerken, L. (2002, January). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), 101–111.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental science*, *12*(3), 369–378.

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology*, *31*(5), 838.

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in English-learning 15-month-olds. *Frontiers in Psychology*, *4*(24).

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition*, *96*(2), 143–182.

Morgan, J. L. (1986). *From simple input to complex grammar.* The MIT Press.

Morgan, J. L., & Demuth, K. (Eds.). (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Lawrence Erlbaum.

Nazzi, T., Barrière, I., Goyet, L., Kresh, S., & Legendre, G. (2011). Tracking irregular morphophonological dependencies in natural language: Evidence from the acquisition of subject-verb agreement in French. *Cognition*, *120*(1), 119–135. (Publisher: Elsevier)

Nazzi, T., Bertoncini, J., & Mehler, J. (1998, June). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(3), 756–766.

Pearl, L., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*(1), 23–68.

Pearl, L., & Sprouse, J. (2019). Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition. *Language*, *95*(4). (Publisher: Linguistic Society of America)

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338. (Publisher: Elsevier)

Perfors, A., Tenenbaum, J. B., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of child language*, *37*(3), 607–642.

Perkins, L. (2019). *How grammars grow: Argument structure and the acquisition of* *non-basic syntax* (Doctoral dissertation). University of Maryland, College Park, MD.

Perkins, L., Feldman, N. H., & Lidz, J. (2022). The power of ignoring: filtering input for argument structure acquisition. *Cognitive Science*, *46*(1).

Perkins, L., & Lidz, J. (2020). Filler-gap dependency comprehension at 15 months: The role of vocabulary. *Language Acquisition*, *27*(1), 98–115.

Perkins, L., & Lidz, J. (2021). 18-month-old infants represent non-local syntactic dependencies. *Proceedings of the National Academy of Sciences*, *118*(41), e2026469118.

Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge, MA: Harvard University Press.

Reinhart, T. (1998). Wh-in-situ in the framework of the Minimalist Program. *Natural language semantics*, *6*(1), 29–56.

Sakas, W., & Fodor, J. D. (2001). The structural triggers learner. In S. Bertolo (Ed.), *Language acquisition and learnability* (pp. 172–233). Cambridge: Cambridge University Press. (Publisher:)

Sakas, W., & Fodor, J. D. (2012). Disambiguating syntactic triggers. *Language Acquisition*, *19*(2), 83–143. (Publisher: Taylor & Francis)

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with markov chain monte carlo. *Cognitive Psychology,*, *60*(2).

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition*, *69*(2), 105–134.

Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early Understanding of Subject and Object Wh-Questions. *Infancy*, *4*(3), 423–436.

Shi, R., & Melançon, A. (2010). Syntactic Categorization in French-Learning Infants. *Infancy*, *15*(5), 517–533.

Shi, R., Morgan, J. L., & Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of child* *language*, *25*(01), 169–201.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11–B21.

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, *35*(4), 869–902.

Soderstrom, M., Ko, E.-S., & Nevzorova, U. (2011). It's a question? Infants attend differently to yes/no questions and declaratives. *Infant behavior and development*, *34*(1), 107–110. (Publisher: Elsevier)

Soderstrom, M., Wexler, K., & Jusczyk, P. W. (2002). English-learning toddlers' sensitivity to agreement morphology in receptive grammar. In *Proceedings of the 26th Annual Boston University conference on language development* (Vol. 2, pp. 643–652). Somerville, MA: Cascadilla Press.

Soderstrom, M., White, K. S., Conwell, E., & Morgan, J. L. (2007). Receptive grammatical knowledge of familiar content words and inflection in 16-month-olds. *Infancy*, *12*(1), 1–29.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science (New York, N.Y.)*, *348*(6230), 91–94. doi: 10.1126/science.aaa3799

Stahl, A. E., & Feigenson, L. (2017, June). Expectancy violations promote learning in young children. *Cognition*, *163*, 1–14. doi: 10.1016/j.cognition.2017.02.008

Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition*, *4*(1-2), 5–48.

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

Sussman, R. S., & Sedivy, J. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes*, *18*(2), 143–163.

Suzuki, T., & Kobayashi, T. (2017). Syntactic Cues for Inferences about Causality in

Language Acquisition: Evidence from an Argument-Drop Language. *Language Learning and Development*, *13*(1), 24–37.

Tincoff, R., Santelmann, L. M., & Jusczyk, P. W. (2000). Auxiliary verb learning and 18-month-olds' acquisition of morphological relationships. In *Proceedings of the 24th annual Boston University conference on language development* (Vol. 2, pp. 726–737). Cascadilla Somerville, MA.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011, May). Pure reasoning in 12-month-old infants as probabilistic inference. *Science (New York, N.Y.)*, *332*(6033), 1054–1059. doi: 10.1126/science.1196404

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. G. De Villiers (Eds.), *Language Processing and Language Acquisition.* Dordrecht: Kluwer.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1-2), 21–81.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences*, *104*(33), 13273–13278. (Publisher: National Acad Sciences)

Van Heugten, M., & Shi, R. (2010). Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries. *The Journal of the Acoustical Society of America*, *128*(5), EL223–EL228. (Publisher: Acoustical Society of America)

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

White, A. S., & Lidz, J. (2022). Lexicalization in the developing parser. *Glossa*

Psycholinguistics, *1*(1).

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current directions in psychological science*, *18*(1), 53–57.

Yang, C. (2002). *Knowledge and learning in natural language*. Oxford: Oxford University Press.

Yang, Y. (2022). *Are you asking me or telling me? Learning clause types and speech acts in English and Mandarin* (Doctoral dissertation). University of Maryland, College Park, MD.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child development*, *83*(4), 1382–1399.

Appendix

Details of Gibbs Sampling

We use Gibbs sampling (Geman & Geman, 1984) to jointly infer c and e, integrating over θ, $\delta^{(X)}$, and $\delta^{(F)}$.

6.1 Sampling c

To begin, values of c for each sentence are initialized to one of three initial sentence categories: one category with transitivity violations and two without. These initial categories are sampled from the posterior probability distribution that a given sentence contains a transitivity violation under the model in Perkins et al. (2022). If a sentence is sampled as containing a transitivity violation under that model, it is initialized to the transitivity-violating category; if not, it is randomly initialized to one of the two non-violating categories.

After initializing c, new values of c for each sentence are re-sampled sequentially. From observations of direct objects and other features in a sentence, and across other sentences in the model's data, the model determines which previously seen or new value of c was most likely to have generated those observations. For direct object observation $X_i^{(v)}$ and other feature observations $\vec{F}_i^{(v)}$ in sentence i, together with all other direct object observations $\mathbf{X}$, feature observations $\vec{\mathbf{F}}$, and sentence category assignments $\mathbf{c}$ for other sentences in the dataset, we use Bayes' Rule to compute the posterior probability of each value for c,

$$p(c_i|X_i^{(v)}, \vec{F}_i^{(v)}, T^{(v)}, e_c, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c}) = \frac{p(X_i^{(v)}, \vec{F}_i^{(v)}|c_i, e_c, T^{(v)}, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c})p(c_i|\mathbf{c})}{\sum_{c_i'} p(X_i^{(v)}, \vec{F}_i^{(v)}|c_i', e_c, T^{(v)}, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c})p(c_i'|\mathbf{c})} \tag{1}$$

The posterior probability of a particular value of c given the observed data, known transitivity categories, and other sentence category values is proportional to the likelihood, the probability of $X_i^{(v)}$ and $\vec{F}_i^{(v)}$ given that value of c, other observed data and category values, and the prior probability of c. We assume that c is independent of all other model parameters. The prior probability of c is a Dirichlet process (Ferguson, 1973) with parameter α. In this process, each category value c_i has prior probability proportional to the number of sentence observations already assigned to that category, n_{c_i}. This process also reserves a small non-zero probability for new categories of c, determined by the parameter α, which we set equal to 1. The proportion of this extra probability that is reserved for new transitivity-violating categories is 0.19, the mean rate of transitivity violations inferred by the model in Perkins et al. (2022), and the proportion reserved for new categories without violations is set to 0.81. For n total observations of sentences across all categories, we define the prior on c,

$$p(c_i|\mathbf{c}) = \begin{cases} \dfrac{n_{c_i}}{n+\alpha} & \text{for previously seen values of } c \\ \dfrac{0.19\alpha}{n+\alpha} & \text{for new values where } e_c = 1 \\ \dfrac{0.81\alpha}{n+\alpha} & \text{for new values where } e_c = 0 \end{cases} \quad (2)$$

Assuming independence between X and F, we calculate the likelihood as the product of the probabilities of observing $X_i^{(v)}$ and $\vec{F}_i^{(v)}$, given the other observations and model parameters,

$$p(X_i^{(v)}, \vec{F}_i^{(v)}|c_i, e_c, T^{(v)}, \mathbf{X}, \vec{\mathbf{F}}, \mathbf{c}) = p(X_i^{(v)}|c_i, e_c, T^{(v)}, \mathbf{X}, \mathbf{c})p(\vec{F}_i^{(v)}|c_i, e_c, \vec{\mathbf{F}}, \mathbf{c}) \quad (3)$$

The first term in this likelihood function is calculated differently depending on the value of e_c for the current category c_i. If c_i is a transitivity-violating category ($e_{c_i} = 1$), then direct objects are generated by the grammatical property of that category $\delta_{c_i}^{(X)}$. We calculate the probability of a direct object by integrating over all possible values of $\delta_{c_i}^{(X)}$, conditioning on other observations of sentences in this category,

$$p(X_i^{(v)}|c_i, e_i = 1, T^{(v)}, \mathbf{X}, \mathbf{c}) = \int p(X_i^{(v)}|\delta_{c_i}^{(X)})p(\delta_{c_i}^{(X)}|c_i, \mathbf{X})d\delta_{c_i}^{(X)} \quad (4)$$

The first term inside the integral is equal to $\delta_{c_i}^{(X)}$ if $X_i^{(v)} = 1$, or $1 - \delta_{c_i}^{(X)}$ if $X_i^{(v)} = 0$.

We can use Bayes' Rule to compute the second term inside the integral, the probability of $\delta_{c_i}^{(X)}$ given all other observations within the category,

$$p(\delta_{c_i}^{(X)}|c_i, \mathbf{X}) = \frac{p(\mathbf{X}|\delta_{c_i}^{(X)}, c_i)p(\delta_{c_i}^{(X)}|c_i)}{\int p(\mathbf{X}|\delta_{c_i}^{(X)}, c_i)p(\delta_{c_i}^{(X)}|c_i)d\delta_{c_i}^{(X)}} \tag{5}$$

The prior probability $p(\delta_{c_i}^{(X)}|c_i)$ is assumed to follow a uniform $Beta(1, 1)$ distribution. Let n_{c_i} be the total observations in category c_i and k_{c_i} be the total direct object observations in this category. The likelihood term, $p(\mathbf{X}|\delta_{c_i}^{(X)}, c_i)$, is the probability of observing k_{c_i} direct objects in n_{c_i} total observations. This follows a binomial distribution with parameter $\delta_{c_i}^{(X)}$,

$$p(\mathbf{X}|\delta_{c_i}^{(X)}, c_i) = \binom{n_{c_i}}{k_{c_i}}(\delta_{c_i}^{(X)})^{k_{c_i}}(1 - \delta_{c_i}^{(X)})^{n_{c_i}-k_{c_i}} \tag{6}$$

Solving the integral in equation (4), we calculate that $X_i^{(v)}$ takes a value of 1 with probability $\frac{k_{c_i}+1}{n_{c_i}+2}$, and 0 with probability $\frac{n_{c_i}-k_{c_i}+1}{n_{c_i}+2}$.

If c_i is not a transitivity-violating category ($e_{c_i} = 0$), then direct objects in this category are generated by the transitivity properties of each verb. The first term in the likelihood function in (3) thus depends on the known transitivity category $T^{(v)}$ and $\theta^{(v)}$, the rate of direct objects under that transitivity category. If verb v is transitive or intransitive, then θ is known, and $X_i^{(v)}$ takes a value of 1 with probability θ, and 0 with probability $1 - \theta$. If verb v is alternating, we again integrate over all possible values of $\theta^{(v)}$, conditioning on observations of this verb in other categories without argument gaps. This integral is analogous to the integral in equation (4). Here, let $n_1^{(v)}$ be the total observations for verb v in categories where $e_c = 0$, and $k_1^{(v)}$ be the total direct object observations for verb v in these categories. Following equations analogous to (4)-(6), we calculate that $X_i^{(v)}$ takes a value of 1 with probability $\frac{k_1^{(v)}+1}{n_1^{(v)}+2}$, and 0 with probability $\frac{n_1^{(v)}-k_1^{(v)}+1}{n_1^{(v)}+2}$.

The second term in (3) is the probability of the other observed features occurring in the given category. Assuming independence among features, this is equivalent to the product over the probabilities of observing each feature in this category,

$$p(\vec{F}_i^{(v)}|c_i, e_c, \vec{\mathbf{F}}, \mathbf{c}) = \prod_{F_i'^{(v)}} p(F_i'^{(v)}|c_i, e_c, \mathbf{F}, \mathbf{c}) \tag{7}$$

The probability of observing a particular feature F in a category c_i is given by $\delta_{c_i}^{(F)}$ for that feature and that category. We integrate over all possible values of $\delta_{c_i}^{(F)}$, conditioning on other observations of feature F. Let n_{c_i} be the total observations in category c_i and $k_{c_i}^F$ be the total observations of feature F in this category. Following equations analogous to (4)-(6), we calculate that $F_i^{(v)}$ takes a value of 1 with probability $\frac{k_{c_i}^F+1}{n_{c_i}+2}$, and 0 with probability $\frac{n_{c_i}-k_{c_i}^F+1}{n_{c_i}+2}$.

6.2 Sampling e

After sampling values for c for each sentence in the dataset, we then sample new values of e for each category. We calculate the posterior probability of each value of e_c for a category c given all of the direct object observations in that category $\mathbf{X}_c$ and known verb transitivity properties T,

$$p(e_c|c, \mathbf{X}_c, T) = \frac{p(\mathbf{X}_c|e_c, c, T)p(e_c)}{\sum_{e_c'} p(\mathbf{X}_c|e_c', c, T)p(e_c')} \tag{8}$$

We assume that e_c is independent of T and c, and that the prior probability $p(e_c) = 1$ is again set to 0.19, the mean rate of transitivity violations inferred by the model in Perkins et al. (2022). In other words, the learner assumes that the prior probability of a transitivity-violating category is equivalent to the probability that any single sentence contains a transitivity violation, as inferred by the previous learner. This will only be the case if sentences are equally distributed among categories, a simplifying assumption of the learner's prior that may be overridden if not supported by the data.

The likelihood term, $p(\mathbf{X}_c|e_c, c, T)$, is the probability of seeing particular observations of direct objects for verbs in this category. If $e_{c_i} = 1$ and c_i is a transitivity-violating category, this probability is determined by $\delta_{c_i}^{(X)}$. We calculate the joint probability of the direct object observations for each verb in that category given $\delta_{c_i}^{(X)}$, integrating across all possible values of $\delta_{c_i}^{(X)}$,

$$p(\mathbf{X}_c|e_c = 1, c, T) = \int \prod_{v'} \left(p(\mathbf{X}_c^{(v')}|\delta_{c_i}^{(X)}) \right) p(\delta_{c_i}^{(X)}|c_i) d\delta_{c_i}^{(X)} \tag{9}$$

The first term inside the integral is the product across all verbs of probability of the direct observations for that verb $\mathbf{X}_c^{(v)}$ in the category, given $\delta_{c_i}^{(X)}$. This probability is given in equation (6). We again assume that the prior probability $p(\delta_{c_i}^{(X)}|c_i)$ follows a uniform $Beta(1,1)$ distribution. Let n_c be the total observations in a particular category and k_c be the total direct object observations in that category. Solving the integral in equation (9), we find that

$$p(\mathbf{X}_c|e_c = 1, c, T) = \frac{\Gamma(k_c+1)\Gamma(n_c-k_c+1)}{\Gamma(n_c+2)} \left(\prod_{v'} \frac{\Gamma(n_c^{(v')}+1)}{\Gamma(k_c^{(v')}+1)\Gamma(n_c^{(v')}-k_c^{(v')}+1)} \right) \tag{10}$$

If $e_{c_i} = 0$ and c_i is not a transitivity-violating category, the likelihood term in equation (8) is determined by the known transitivity $T^{(v)}$ of each verb in the category. The probability of the particular direct object observations $\mathbf{X}_c$ in the category is the joint probability of seeing those direct object observations for each verb, given the transitivity of that verb,

$$p(\mathbf{X}_c|e_c = 0, c, T) = \prod_{v'} \left(p(\mathbf{X}_c^{(v')}|T^{(v')}) \right) \tag{11}$$

We can again re-write $\mathbf{X}_c^{(v)}$ as k_c^v direct object observations out of n_c^v total observations for a given verb in a given category. The probability of observing k_c^v direct objects out of n_c^v total observations of a verb follows a binomial distribution with parameter $\theta^{(v)}$. Recall that $\theta^{(v)} = 1$ for transitive verbs and $\theta^{(v)} = 0$ for intransitive verbs. For alternating verbs, we must integrate across all possible values of $\theta^{(v)}$,

$$p(\mathbf{X}_c|e_c = 0, c, T) = p(k_c^{(v)}|n_c^{(v)}, T^{(v)}) = \int p(k_c^{(v)}|n_c^{(v)}, \theta^{(v)})p(\theta^{(v)}|T^{(v)})d\theta^{(v)} \tag{12}$$

We assume that $p(\theta^{(v)}|T^{(v)})$ follows a $Beta(\alpha, \beta)$ distribution, where the parameters α and β are counts of direct object observations and no direct object observations for verb v in other categories without argument gaps. Solving the integral in equation (12), we find that

$$p(k_c^{(v)}|n_c^{(v)}, T^{(v)}) = \left(\frac{\Gamma(n_c^{(v)} + 1)}{\Gamma(k_c^{(v)} + 1)\Gamma(n_c^{(v)} - k_c^{(v)} + 1}\right)\left(\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)\left(\frac{\Gamma(k_c^{(v)} + \alpha)\Gamma(n_c^{(v)} - k_c^{(v)} + \beta)}{\Gamma(n_c^{(v)} + \alpha + \beta)}\right) \tag{13}$$

6.3 Sampling with Annealing

The simulations reported here used 5,000 total iterations of Gibbs sampling. To aid in the model's search process, simulated annealing was used during the first 1,000 iterations. In this process, we raise the posterior probabilities of c and e to the power of an annealing constant defined as $1/t$, where t is the current temperature. Then, we slowly lower the temperature (reduce t) until the annealing constant reaches 1. While the temperature is warm, the posterior probability distributions are flattened so the learner is able to explore more of its hypothesis space. After 1,000 iterations of Gibbs sampling with annealing, another 4,000 iterations were run without annealing. The final iteration was taken as a sample from the posterior distribution over c and e.