Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition

Lisa Pearl & Jon Sprouse*

Abstract

To successfully learn language – and more specifically how to use verbs correctly – children must solve the linking problem: they must learn the mapping between the thematic roles specified by a verb's lexical semantics and the syntactic argument positions specified by a verb's syntactic frame. We use an empirically-grounded and integrated quantitative framework involving corpus analysis, experimental meta-analysis, and computational modeling to implement minimally distinct versions of mapping approaches that (i) are either specified *a priori* or learned during language acquisition, and (ii) rely on either an absolute or a relative thematic role system. Using successful verb class learning as an evaluation metric, we embed each approach within a concrete model of the acquisition process and see which learning assumptions are able to match children's verb learning behavior at three, four, and five years old. Our current results support a trajectory where children (i) may not have prior expectations about linking patterns between ages three and five, and (ii) begin with a relative thematic system, progressing towards optionality between a relative and an absolute system. We discuss implications of our results for both theories of syntactic representation and theories of how those representations are acquired. We also discuss the broader contribution of this study as a concrete modeling framework that can be updated with new linking theories, corpora, and experimental results.

1 Introduction

To successfully learn how to use a verb, children must learn (at least) three pieces of information: (i) the syntactic properties of the verb, such as the syntactic frames that it can appear in, (ii) the lexical semantics of the verb, including the thematic roles assigned by the verb, and (iii) a mapping between the thematic roles specified by the verb's lexical semantics and the syntactic argument

---

[*] We are very grateful to Megan Crowhurst, Christina Tortora, two anonymous reviewers, and the audiences at the Symposium on Advances in Distributional Models of Language and Meaning 2017, the UC Irvine Language Science Colloquium 2017, SynLinks 2016, and the McGill Linguistics Colloquium 2016 for very useful feedback. Anything we got wrong is our fault. This material is based upon work supported by the National Science Foundation under Grant No. BCS-1347028 to LP and BCS-1347115 to JS.

positions specified by the verb's syntactic frame(s). The learning of this third component is often called the *linking problem*.

At the level of individual verbs and individual syntactic frames, the linking problem doesn't appear to be much of a problem. We might imagine that children simply learn the mapping between thematic roles and syntactic positions for each combination of a verb and syntactic frame one at a time. However, this doesn't account for children's ability to generalize their knowledge to new verbs. That is, if the linking between thematic roles and syntactic positions is only ever learned on a verb-by-verb basis, how could children use a new verb appropriately without hearing all its possible uses? It seems children must be learning linking patterns at a more abstract level because they're capable of generalizing linking patterns from one verb to another (sometimes incorrectly during the course of development): see, for example, Gropen, Pinker, Hollander, Goldberg, and Wilson (1989); Naigles (1990); Naigles and Kako (1993); Gelman and Koenig (2001); Bunger and Lidz (2004); Huttenlocher, Vasilyeva, and Shimpi (2004); Kidd, Lieven, and Tomasello (2006); Conwell and Demuth (2007); Papafragou, Cassidy, and Gleitman (2007); Bunger and Lidz (2008); Thothathiri and Snedeker (2008); Scott and Fisher (2009); Yuan and Fisher (2009); Kidd, Lieven, and Tomasello (2010); Becker (2014) and Hartshorne, Pogue, and Snedeker (2015).

The additional complexity of the linking problem becomes apparent when we consider the broader linking patterns that we see cross-linguistically. Two core linking patterns emerge[1]:

(i) For the vast majority of verbs in accusative languages, AGENT-like thematic roles tend to appear in syntactic *subject* position, PATIENT-like thematic roles tend to appear in syntactic *object* position, and INSTRUMENT/SOURCE/GOAL-like roles tend to appear in oblique syntactic positions such as *indirect object* or *object of PP*.

(ii) Exceptions to this pattern tend to be contained within very specific semantic classes of verbs (see section 2 for examples).

How and why does this regularity in linking patterns emerge? There are currently two general approaches. The first is that the linking patterns could result from children possessing explicit innate knowledge of the linking patterns themselves, such that the linking pattern does not need to be learned during development. We'll call these *innate-mapping* approaches. We note that innate-mapping approaches may be coupled with either early maturation or late maturation of the innate linking knowledge, in terms of the predicted developmental trajectory. Early maturation predicts the knowledge is present as young as we can test, while late maturation predicts the knowledge to only be present in older children. The second possibility is that the linking patterns could derive from the interplay between the input that children receive and the learning mechanisms underlying verb learning. We'll call these *derived-mapping* approaches. Derived-mapping approaches would predict that the linking knowledge will take time to develop, and so it would be less likely to be present in younger children.

To empirically compare these approaches, we must create a framework that meets two criteria: it must be possible to (i) systematically manipulate the presence or absence of prior knowledge

---

[1] We'll be using SMALL CAPITALS to indicate thematic roles and *italics* to indicate syntactic positions. We'll also be abstracting away from the details of syntactic structure, which may vary cross-linguistically, by referring to positions using grammatical labels (e.g., *subject*) rather than phrase structure labels (e.g., spec-TP).

of linking patterns, and (ii) evaluate both approaches on a neutral metric of success. We note that achieving knowledge of the linking pattern itself can't be the metric of success because the innate-mapping approach builds that pattern into the learner explicitly, and thus would automatically "win" under such a metric. With this in mind, we propose to measure success by assessing one prominent type of acquired knowledge that relies on learning linking patterns: whether developmentally-attested verb classes can be learned from the data children encounter, given a computationally modeled child who either explicitly has or doesn't have prior linking knowledge. In other words, we'll use an argument from acquisition to evaluate theories of knowledge representation (Pearl, Ho, & Detrano, 2016; Pearl, 2017) for linking patterns.

Though the verb class learning literature and the linking pattern literature don't always intersect (presumably because the verb class learning literature focuses on development, and the linking pattern literature focuses on adult end states), we believe that verb class learning is a useful common denominator for evaluating the two major approaches. This is because the linking pattern is defined over verb classes (see section 2). More specifically, because innate-mapping approaches predict linking knowledge to also be operative during language learning, it's reasonable to expect modeled learners that incorporate that innate linking knowledge to better match the observed developmental trajectory of human children. Exactly how much prior linking knowledge aids a modeled learner in achieving children's observed verb class learning behavior is an empirical question – one that we investigate here using our integrated quantitative framework of the acquisition process.

For this study, we explore two of the most prominent innate-mapping approaches in the literature that are built on cognitively plausible assumptions about the thematic role systems available to children during development: (i) the Uniformity of Theta Assignment Hypothesis (**UTAH**: Baker 1988, building on Perlmutter and Postal 1984), which uses an absolute thematic system, and (ii) the relativized Uniformity of Theta Assignment Hypothesis (**rUTAH**: Larson 1990, Speas1990, Grimshaw 1990), which uses a relative thematic system. We contrast these with derived-mapping versions that use the same thematic systems, but which don't build in knowledge of how to map to syntactic positions. In this way, these two derived-mapping approaches are minimally different from UTAH and rUTAH, leveraging the cognitively plausible approaches to thematic systems, but without the added assumption of innate linking patterns. Our modeling framework can therefore contribute to two sets of debates: the debate between innate-mapping and derived-mapping approaches, and the debate about the details of the thematic role system.

Within our integrated quantitative framework, we create computationally modeled learners that rely on different combinations of assumptions (e.g., innate-mapping vs. derived-mapping, absolute vs. relative thematic systems). The framework uses existing child-directed corpus data to determine the input for each modeled learner and existing child behavioral evidence to determine the target output knowledge modeled learners should achieve. The modeled learners use hierarchical Bayesian inference to infer verb classes from realistic input distributions, and these inferred verb classes are compared against the target verb classes at different ages. The modeled learner whose output best matches the target verb classes (known by children) can be considered the modeled learner which is most likely to encode the learning assumptions children actually use.

The rest of this paper is organized as follows. We first discuss the linking problem in more detail, along with the theoretically-motivated solutions mentioned above: the innate-mapping UTAH and rUTAH, and their derived-mapping equivalents. We then discuss our use of verb class learning as a neutral evaluation for comparing different approaches. We also present the verb classes that children have acquired by ages 3, 4, and 5, as derived from a review of 38 studies from the experimental acquisition literature; we additionally review the *verb behaviors* examined in those studies, where verb behavior refers to which syntactic frames a verb can appear in, and the thematic role information of its arguments within each frame. We subsequently introduce our acquisition modeling framework, highlighting (i) the components necessary to implement a modeled learner that attempts to learn verb classes, and (ii) how different learning assumptions impact a modeled learner. This includes discussion of how a modeled learner interprets the syntactic and conceptual information available in the input, as well as the empirical data from the CHILDES Treebank (Pearl & Sprouse, 2013a) that the modeled learner's input is based on. We also discuss the hierarchical Bayesian inference process that allows the modeled learner to use the available input to infer verb classes.

Our first key finding is that there's always at least one modeled learner at every age who performs relatively well, which affirms that verb classes can be probabilistically learned from relatively sparse linguistic and conceptual information, as opposed to requiring richer information. Our second key finding is that the modeled learner (and therefore, the specific learning assumption combination) that best matches children's verb class knowledge can change over time. Here, we assume that the progression of learning assumptions that best matches children's verb class knowledge is a reasonable reflection of children's true learning assumptions. With this as a working hypothesis about children's underlying knowledge, our results support a developmental trajectory that begins at three years old with a relative thematic system; it then progresses towards optionality between absolute and relative thematic systems. Interestingly, our results also support either innate-mapping or derived-mapping approaches to linking, depending on the other learning assumptions active in three-, four-, and five-year-olds. We discuss the implications of our current results for syntactic theory, acquisition theory, and future experimental and computational studies of verb learning. Finally, at the broadest level, we discuss the value of explicit, integrated quantitative frameworks like the one here for exploring fundamental questions in syntactic theory and language acquisition, and how the framework we develop here can be extended with additional empirical data and additional theoretical proposals.

2 The linking problem and its potential solutions

In this section, we describe the specific linking theories that we evaluate in this study. Readers familiar with the linking problem, UTAH, and rUTAH should feel free to skim this section.

2.1 A brief introduction to the linking problem

As mentioned above, the linking problem is predicated on two observations. First, there seems to be a primary pattern robustly observed cross-linguistically (see Baker 1997 for a review), as shown in the English examples in (1). This pattern has AGENT-like roles in the syntactic *subject* position, PATIENT-like roles in syntactic *object* position, and INSTRUMENT/SOURCE/GOAL-like roles in oblique syntactic positions.

(1) The primary pattern
  a. Jack cut the pie with a knife.
    (*subject* = AGENT, *object* = PATIENT, *object of PP* = INSTRUMENT)
  b. Jack stole the jewels from the store.
    (*subject* = AGENT, *object* = PATIENT, *object of PP* = SOURCE)
  c. Lily sent the letter to her parents.
    (*subject* = AGENT, *object* = PATIENT, *object of PP* = GOAL)

Second, verbs that are exceptions to this primary pattern tend to form well-defined semantic classes (again, see Baker 1997 for a brief review). For instance, in English, one example is the semantic class known as *psych*-verbs, which involve one of the verb arguments experiencing a psychological or mental state (see Postal, 1971, Belletti & Rizzi, 1988, and Dowty, 1991, among many others). The psych-verb pair in (2) involves two verbs, *fear* and *frighten*, that have very similar lexical semantics but nonetheless yield two distinct linking patterns: the EXPERIENCER of the psychological state and the apparent CAUSER of the psychological state alternate syntactic positions. Interestingly, we don't tend to find this sort of alternation for verbs from other semantic classes.

(2) Psych-verb examples
  a. Lily fears spiders.
    (*subject* = EXPERIENCER, *object* = CAUSER)
  b. Spiders frighten Lily.
    (*subject* = CAUSER, *object* = EXPERIENCER)

A second example of exceptions connected to semantically-defined verb classes involves *split-intransitivity*, where intransitive verbs can be subdivided into two or more subclasses (sometimes called *unergative* and *unaccusative*) that are derived from the lexical semantics of the verbs (Perlmutter, 1978; Burzio, 1986; Levin & Rappaport Hovav, 1995; Sorace, 2000). In English, we can see this in the examples in (3): unergative *sneeze* maps an AGENT to the *subject* position while unaccusative *arrive* maps a PATIENT to the *subject* position.

(3) Split-intransitivity examples
  a. Jack sneezed during the meeting.
    (*subject* = AGENT)
  b. The package arrived during the meeting.
    (*subject* = PATIENT)

The regularity of the primary pattern cross-linguistically and the semantic coherence of the exceptions to it have spurred theories of representation (e.g., explicit linking patterns like UTAH and rUTAH) that compactly encode this regularity. From a representational standpoint, this compact representation would allow easier storage and use of the relevant knowledge that links thematic roles to syntactic positions. From a developmental standpoint, this compact representation would helpfully constrain children's hypotheses and so enable them to solve the linking problem more quickly (Pearl et al., 2016; Pearl, 2017).

As mentioned above, developmental approaches diverge on whether this linking pattern representation is available innately as explicit knowledge (innate-mapping) or is instead derived from language experience (derived-mapping). For innate-mapping approaches, the primary pattern comes for free and children learn exceptions (such as certain psych-verbs and split-intransitivity verbs) through language experience, drawing on learned knowledge of lexical semantics and specific grammatical mechanisms (e.g., the movement operation in Minimalism). In contrast, for derived-mapping approaches, all linking patterns (both the primary one and any exceptions) are inferred from experience with particular verbs. General mechanisms of abstraction allow children to generalize across verbs and learn any linking patterns that exist, based on the input available.

To be clear, there can be significant variability among the theories within each type: innate-mapping theories can vary substantially in how they capture the exceptions to the mapping (Fillmore, 1968; Perlmutter & Postal, 1984; Jackendoff, 1987; Larson, 1988; Grimshaw, 1990; Larson, 1990; Speas, 1990; Dowty, 1991; Baker, 1997), and derived-mapping theories can vary substantially in how they capture regularities (Bowerman, 1988; Tomasello, 1992; Braine & Brooks, 1995; Goldberg, 1995; Tomasello, 2003; Goldberg, 2006; Boyd & Goldberg, 2011; Goldberg, 2013). Given this, we intend to compare modeled learners instantiating (i) approaches that assume or don't assume prior explicit linking knowledge, and (ii) more fine-grained differences within each approach. To that end, we have focused on two prominent innate-mapping solutions, UTAH and rUTAH, and their derived-mapping counterparts; importantly from a cognitive standpoint, these approaches rest on plausible assumptions about the complexity of the thematic system available during development (discussed in section 4.2.2) but differ in the thematic system details.

2.2 The Uniformity of Theta Assignment Hypothesis (UTAH)

UTAH (Fillmore, 1968; Perlmutter & Postal, 1984; Jackendoff, 1987; Baker, 1988; Grimshaw, 1990; Speas, 1990; Dowty, 1991; Baker, 1997) has two components: (i) an inventory of thematic roles that will be used for the calculation of syntactic position, and (ii) an expected mapping between each of the thematic roles and syntactic positions. Here, we assume the implementation from Baker (1997), which posits an inventory of three thematic macro- or proto-roles (Dowty, 1991): proto-AGENT, proto-PATIENT, and OTHER. This implementation is agnostic about the existence of finer-grained thematic roles at a semantic level. All it requires is that any finer-grained typology of thematic roles map to the three proto-roles necessary for the syntactic calculation. In this way, UTAH represents a categorical or absolute approach to the thematic system, where each proto-role is a fixed thematic category. Under this implementation, thematic roles that tend to involve internal causation map to proto-AGENT, roles that tend to involve external causation map to proto-PATIENT, and all other roles map to OTHER (Levin & Rappaport Hovav, 1995). Example (4) lists 13 common finer-grained thematic roles from the literature, and how they would map to the three proto-roles in this implementation.

(4) Example UTAH mapping with three fixed proto-roles

a. proto-AGENT: AGENT, CAUSER, EXPERIENCER (when internally-caused), POSSESSOR
b. proto-PATIENT: PATIENT, THEME, EXPERIENCER (when externally-caused), SUBJECT MATTER
c. OTHER: LOCATION, SOURCE, GOAL, BENEFACTOR, INSTRUMENT

Baker's (1997) implementation assumes that the proto-AGENT role maps to the syntactic *subject* position, the proto-PATIENT role maps to the syntactic *object* position, and that the OTHER role maps to oblique object positions (such as *object of PP*).

To see this UTAH implementation in action, we can apply it to examples of primary and exceptional patterns. For primary pattern sentences like *Jack cut the pie with a knife*, the *subject* is a proto-AGENT, the direct *object* is a proto-PATIENT, and the oblique object is OTHER. For psych-verbs, this implementation of UTAH leverages the internal-vs-external causation distinction: in *Lily fears spiders*, *Lily* is causing her own mental state, and is thus a proto-AGENT; in *Spiders frighten Lily*, *spiders* are causing Lily's mental state, and thus *Lily* is the proto-PATIENT. For the unergative *sneezed* in *Jack sneezed during the meeting*, *Jack* is the proto-AGENT, and mapped to the *subject.* For the unaccusative *arrived* in *The package arrived during the meeting,* this implementation would claim that *the package* enters the syntactic derivation as the *object* of *arrive*, thus respecting UTAH. *The package* would then be moved to the *subject* position by an additional mechanism (such as the movement operation in GB/Minimalism). Appendix A.1 in the supplementary materials provides a more explicit walk-through of this UTAH implementation.

2.3 The relativized Uniformity of Theta Assignment Hypothesis (rUTAH)

rUTAH (Larson, 1988, 1990; Grimshaw, 1990; Speas, 1990) also has two components: (i) a hierarchy of thematic roles that will be used for the calculation of syntactic position, and (ii) an expected mapping between the relative position of thematic roles on the hierarchy and syntactic positions. The basic idea is that for any given utterance, the rUTAH calculation requires the learner to first determine an ordering relation among the utterance's thematic roles, based on a previously-established thematic role hierarchy. This hierarchy is presumably based on a some sort of relative salience of the different thematic roles, possibly even outside of the domain of language itself (though most rUTAH-based analyses leave open the etiology of the thematic role hierarchy). The learner can then use that ordering relation of the utterance's roles to map each role to a syntactic position: the thematic role that is highest in the hierarchy will map to the (structurally) highest syntactic position, the next highest thematic role will map to the next highest syntactic position, and so on. Here, we created a thematic role hierarchy based on the hierarchies developed in Larson (1988, 1990) using the 13 common thematic roles from the literature mentioned above. This hierarchy is given in (5) and example utterance mappings are in (10a)-(10e). Note that some roles may not be strictly ordered with respect to each other in the hierarchy. For instance, LOCATION and SOURCE are equally salient in the hierarchy in (5).

For this implementation, we assume that syntactic *subjects* are structurally higher than syntactic *objects*, which in turn are higher than oblique *objects*.

(5) Hierarchy:
AGENT > CAUSER > EXPERIENCER > POSSESSOR >
SUBJECT MATTER > CAUSEE > THEME > PATIENT >
(LOCATION, SOURCE, GOAL, BENEFACTOR, INSTRUMENT)

For primary pattern sentences like *Jack cut the pie with a knife*, there are three thematic roles: AGENT, PATIENT, and INSTRUMENT. The thematic hierarchy places them in that order (AGENT > PATIENT > INSTRUMENT), so they map to *subject*, *object*, and oblique *object* positions respectively. For psych-verbs like *fear* in *Lily fears spiders*, rUTAH would posit that *Lily* is an EXPERIENCER, while *spiders* is a SUBJECT MATTER. As such, *Lily* will map to the *subject* position, and *spiders* will map to the *object* position. In contrast, for psych-verbs like *frighten* in *Spiders frighten Lily*, rUTAH would posit that *spiders* is now a CAUSER, though *Lily* is still an EXPERIENCER. Because CAUSER > EXPERIENCER, *spiders* will map to the *subject* position, and *Lily* will map to the *object* position. Finally, for the intransitive verbs *sneezed* in *Jack sneezed during the meeting* and *arrived* in *The package arrived during the meeting*, both verbs only have one syntactic position and one thematic role; so, the argument appears in *subject* position regardless of its thematic role. Appendix A.2 in the supplementary materials provides a more explicit walk-through of this rUTAH implementation.

2.4 UTAH vs. rUTAH

To be clear, the implementations of UTAH and rUTAH that we adopt here are just two of many possible implementations of these theories. We don't believe that there's anything special about the specific implementations that we chose (and future studies should investigate other implementations). What's critical for our purposes, because we intend to model the acquisition process, is that UTAH and rUTAH involve two distinct types of thematic systems that are developmentally plausible. That is, to map thematic roles onto syntactic positions, children are likely to either (i) make a small number of coarse intermediate categories of thematic roles corresponding to proto-roles, or (ii) view some roles as more salient than others, and order roles accordingly. In each case, the critical step is limiting the number of thematic roles that children must attend to and track statistically, either in absolute terms or in relative terms. That said, we do believe that the implementations of UTAH and rUTAH that we have chosen for our models are relatively representative of the theory types as a whole, at least as far as the two theories are represented in the theoretical literature.

2.5 Derived-mapping equivalents of UTAH and rUTAH

Derived-mapping approaches don't postulate any expected mapping between thematic roles and syntactic positions at the beginning of acquisition. Instead, some verbs and their linking patterns are first learned in isolation; then, over time, if enough verbs are learned with the same properties, a class is formed via general-purpose learning mechanisms that allows these linking patterns (and other verb behaviors) to generalize. In other words, over time, children will build verb classes that can be used to make predictions about novel verbs. In this way, derived-mapping approaches can capture both the regularities and the exceptions that we observe within and across languages. Both result from different verb classes derived in a bottom-up way from experience. More specifically, children learn patterns associated with individual verbs, create verb classes based off of those verbs, and then generalize to more abstract patterns (e.g., an expected linking pattern within a given verb class). So, children derive an expectation for linking pattern mappings over time, rather than being innately equipped with this expectation. We note that derived-mapping approaches would need to identify another source of the primary linking pattern's cross-linguistic robustness. That is, because knowledge of the explicit linking pattern isn't innate, the consistency of the primary linking pattern across languages must come from somewhere else under a derived-mapping approach. Moreover, we note that derived-mapping approaches clearly must assume some kinds of innate knowledge and abilities (e.g., the general-purpose learning mechanisms they rely on are typically considered to be innate). It's just that derived-mapping approaches don't assume that the explicit linking pattern knowledge itself is innate, the way innate-mapping approaches do.

2.6 Evaluating the expectation for a mapping

To evaluate the role of expected mappings in acquisition, we begin with the thematic role systems from either UTAH (an absolute set of 3 proto-roles) or rUTAH (a relative hierarchy), and manipulate the presence or absence of an expected link between thematic roles and syntactic positions. To reiterate, we focus on UTAH and rUTAH because they're prominent innate-mapping approaches, and can easily generate minimally different derived-mapping versions. Importantly, by manipulating whether a modeled learner has or doesn't have prior knowledge of a linking pattern between thematic roles and syntactic positions, we can evaluate whether having or not having this knowledge yields behavior that matches children's observable behavior with respect to verb classes. Any results can then be interpreted with respect to innate-mapping and derived-mapping approaches to solving the linking problem.

3 Verb classes as an evaluation metric

3.1 Verb classes defined by verb behaviors

To compare different approaches to solving the linking problem, we evaluate these approaches on a shared goal: the acquisition of developmentally observed verb classes. The predominant approach to defining verb classes in the literature (e.g., Levin 1993) is by verb behavior: which syntactic frames a verb can appear in, as well as the thematic role information of its arguments within each frame. For example, both *want* and *seem* can appear in the syntactic frame NP V $\text{IP}_{-finite}$ (e.g., *Jack wants/seems to laugh*). However, *want* gives the *subject* NP *Jack* an EXPERIENCER role while *seem* gives the *subject* NP no role (instead, that NP's role comes only from the embedded verb). We additionally include animacy information of a verb's arguments (e.g., *Jack* is +animate) as part of a verb's behavior. (See section 4.2 for the developmental motivation to include animacy information.) A verb class can then be defined as a distribution over verb behavior, i.e., the combination of syntactic frames, positional thematic roles, and animacy of arguments a verb appears with. For example, one verb class during the course of development may consist of the verbs that are known only to be passivizable by a certain age (+passive): they appear with the syntactic frame NP *be/get* $V_{participle}$ (e.g., *The cookie was/got eaten*), and in that frame, the *subject* NP is the PATIENT though the NP can be either +/-animate. Another verb class may consist of verbs known to both be passivizable and able to take a non-finite *to* sentential complement (+passive, +non-finite *to*): they exhibit the passive behavior noted above, and in addition allow the syntactic frame NP V $IP_{-finite}$.

3.2 Target states: Verb classes known by children at different ages

To evaluate the performance of our modeled learners, we need to establish a target knowledge state for them to reach. We're also interested in the developmental trajectory of verb class knowledge, and so want to assess a modeled learner's ability to capture child knowledge at different ages. Importantly, English child verb classes may well differ from English adult verb classes and so we use the experimental acquisition literature on children's comprehension and production of verbs as evidence of children's knowledge of verb classes.

To derive those verb classes, we first did a meta-analysis of 38 articles from the experimental acquisition literature (see Appendix B in the supplementary materials). Based on this, we extracted (i) the set of verbs that children comprehend and/or produce at different ages, and (ii) the set of verb behaviors that are associated with these verbs at those ages. This meta-analysis yielded 12 verb behaviors (see Table 1) for 86 verbs that can be used to define child verb classes in English.[2] The full results of our meta-analysis of the experimental literature can be found in Appendix B.

Because the input data available to our modeled learners from the CHILDES Treebank (Pearl & Sprouse, 2013b) range up to five years old, we focused on the verb classes children seem to know by age three, four, and five. (These corpus data are discussed in more detail in section 4.2.) We additionally restricted these classes to verbs appearing five or more times in the age-appropriate input sets for three- , four- , and five-year-olds, with the idea that a modeled learner could infer something from the distribution of verbs appearing at least this often. This process resulted in the verbs and derived verb classes characterized by different verb behaviors that are summarized in Table 2, for a total of 15-25 verb classes comprising 60-84 verbs from ages three to five. The full description of these child verb classes are in Appendix C of the supplementary materials (Tables 7, 8, and 9).

One important property of the child verb classes serving as the modeled learner target state is that a specific verb can change its verb class over time (based on child behavior with that verb); this therefore means the content of verb classes can change over time. For example, the class where verbs are known only to be passivizable ([+passive]) at age three contains 20 verbs, while the same class at age four contains 26 verbs (it adds 6 verbs over time). As another example, *see* belongs to the passivizable ([+passive]) class at age three, the passivizable class that also allows *that* complements ([+passive, +*that*-complement]) at age four, and the passivizable class that allows both *that* and *whether/if* complements ([+passive, +*that*-complement, +*whether/if*-complement]) at age five. We note that because these verb classes are derived from existing behavioral data, the changes to a verb's class represent either (i) development of verb class knowledge, or (ii) a (current) lack of

---

[2] We note that the earliest age documented in the experimental literature was used as the age of acquisition for the verb behavior associated with a specific verb.

Table 1: Verb behaviors associated with specific verbs from child behavioral study meta-analysis.

verb behavior	**example**	**description**
Unaccusative	The ice *melted*.	Intransitive (NP V) frame where *subject* is PATIENT.
Ditransitive	Jack *sent* Lily the apple.	Verb allows double object construction (NP V NP NP).
Passivizable	Jack was *tricked/laughed* at.	Verb allows passive frame (NP *be/get* $V_{participle}$) where *subject* is PATIENT of verb or verbal complex.
Control object	Lily *asked* him to escape.	Embedded *subject* is GOAL of matrix verb and AGENT of embedded verb in NP V NP $IP_{-finite}$ frame.
Raising object	Lily *wanted* him to escape.	Embedded *subject* is AGENT of embedded verb only in NP V NP $IP_{-finite}$ frame.
Control subject	Jack *tried* to escape.	*subject* is AGENT of matrix verb and embedded verb in NP V $IP_{-finite}$ frame.
Raising subject	Jack *happened* to escape.	*subject* is AGENT of embedded verb only in NP V $IP_{-finite}$ frame.
Psych: Subject experiencer	Jack *loved* Lily.	*subject* is EXPERIENCER of verb in NP V NP frame.
Psych: Object experiencer	The giant *frightened* Jack.	*object* is EXPERIENCER of verb in NP V NP frame.
Non-finite *to* complement	I *want* (him) to go.	Verb allows a non-finite *to* complement, with or without an embedded *subject* (NP V (NP) $IP_{-finite}$).
that-complement	Lily *hoped* that Jack escaped.	Verb allows finite complement headed by *that* (NP V CP_{that}).
whether/if-complement	Lily *wondered* whether Jack escaped.	Verb allows finite complement headed by *whether* or *if* (NP V $CP_{whether/if}$).

empirical data about knowledge of verb behavior at younger ages. Under the working assumption that these are developmental changes to verb class knowledge over time, we test our modeled learners at three ages, determining which modeled learners (representing different learning assumption combinations) can best match children's verb class knowledge development.

3.3 Assessing verb class learning

3.3.1 The Rand Index

Each modeled learner outputs a set of inferred verb classes, with each class containing one or more verbs, and each verb belonging to only one class. We want to assess how well these inferred verb classes match the true verb classes derived from observed child behavior. Because the output is similar to that of a clustering task (i.e., the modeled learner outputs clusters of verbs, which are the inferred verb classes), we consider evaluation metrics from the machine learning literature on clustering. For this study, we use the Rand Index (**RI**; Rand, 1971) because it's a common measure in the clustering literature and it has an intuitive absolute interpretation. Those readers

Table 2: Summary of verb classes derived from child behavioral data for three, four, and five-year-olds. This includes the number of derived verb classes, the number of verbs appearing five or more times in the dataset captured by those classes, and the verb behaviors that comprise the derived verb classes.

age	**# classes**	**# verbs**	**verb behaviors**
3yrs	15	60	unaccusative, ditransitive, non-finite *to* complement, passivizable, *that* complement
4yrs	23	76	unaccusative, ditransitive, non-finite *to* complement, passivizable, *that* complement, control object, control subject, psych object experiencer, psych subject experiencer, raising object, raising subject
5yrs	25	84	unaccusative, ditransitive, non-finite *to* complement, passivizable, *that* complement, control object, control subject, psych object experiencer, psych subject experiencer, raising object, raising subject, *whether/if* complement

already familiar with this metric should feel free to skip the explanation below.

The RI is a pairwise measure derived from signal detection theory. When considering a pair of verbs, there are two possible true states: the two verbs are clustered together into a single class in children's minds, or the two verbs are separated into two distinct classes. Similarly, there are two possible modeled learner output states: the two verbs are clustered into a single class in the modeled learner, or the two verbs are separated into two distinct classes. Crossing the true child and modeled learner output states leads to four possible combinations, as shown in Table 3. When two verbs are the same kind in the true state (*True child state: Together*), they should be clustered together in the modeled learner output. A true positive (TP) occurs when the modeled learner clusters these verbs together, while a false negative (FN) occurs when the modeled learner separates these verbs. When two verbs are not the same kind in the true child state (*True child state: Separate*), they should be separated by the modeled learner. A true negative (TN) occurs when the modeled learner does separate them, while a false positive (FP) occurs when the modeled learner clusters them together. The RI is the ratio of correct classifications (true positives and true negatives) to the total number of classifications made (true positives, true negatives, false positives, and false negatives): $\frac{TP+TN}{TP+TN+FP+FN}$. The intuitive appeal of this ratio is that credit is given both for correctly putting verbs together into the same class and for correctly keeping them separate. The RI ranges between 0 (no classifications are correct) and 1 (all classifications are correct): $0 \leq$ RI ≤ 1. The interpretation of the RI is intuitive in an absolute sense: an RI of .5 means that half of the classifications were correct; equivalently, for any randomly chosen verb pair, there is a probability of .5 that the modeled learner's output will agree with the true child state.

Table 3: Signal detection theory distinctions relevant for the Rand index (RI) when applied to a verb pair.

		Modeled learner output state	
		Together	*Separate*
True child state	*Together*	True Positive	False Negative
	Separate	False Positive	True Negative

3.3.2 Evaluating an RI score relative to chance

One limitation of the RI is that the distribution of RI scores for any given number of classes is not known; therefore we can't determine from the RI score alone if the RI score we obtain is particularly good or particularly bad. We might therefore want to perform some sort of test that compares the observed RI score to the distribution of RI scores expected by a null hypothesis (either chance, or some other null expectation). One solution to this problem is to use a randomization test that randomizes the three types of information (i.e., the parameters) that the modeled learners are inferring from their input: the number of verb classes, the size of each verb class, and the assignment of individual verbs to these classes. In particular, using the same generative process the modeled learners will use (described more fully in Section 4.3 and Appendix E.2 of the supplementary materials), we generate a random number of classes of random size, and randomly assign verbs to these classes. We can then calculate an RI score for this randomized set of classes, which is equivalent to an RI under the null hypothesis that the parameters are exchangeable. We can repeat this process some large number of times (e.g., 10,000) to estimate a distribution of RI scores under this null hypothesis. We can then calculate the probability of obtaining our observed RI score (or one more extreme) under the null hypothesis using this distribution. We report the observed RI and the threshold for significance at $p<.01$.[3]

4 Computationally modeling the acquisition of verb classes

4.1 The acquisition framework

We follow the view that language acquisition is an information-processing task, where children use their available input to build an internal system of linguistic knowledge whose behavioral output we can observe (Lidz & Gagliardi, 2015; Omaki & Lidz, 2015; Pearl, in press). The framework of Pearl (in press), building on that of Lidz and Gagliardi (2015) and Omaki and Lidz (2015), articulates several crucial components of this task, underscoring how theories of representation and theories of the learning process work together to create a complete theory of acquisition.

---

[3] Another common method for evaluating RI scores relative to chance is to convert RI scores into a new measure known as the adjusted Rand Index (**ARI**) (Hubert & Arabie, 1985). We discuss this option in more detail in Appendix D in the supplementary materials. We note that using the ARI instead of (or in combination with) the RI and randomization test described in the main text does not qualitatively change the results we report.

For our purposes, there are three crucial pathways. First, there's the **input-intake** pathway, where the external signal, the *input*, is encoded by the child into an internal mental representation we'll call the *linguistic intake*.[4] The parts of the linguistic intake that are identified by the acquisition system as relevant for acquisition are called the *acquisitional intake*. For example, an input utterance of *What's she climbing over?* might be encoded by the child as containing certain syntactic and conceptual information – this is the linguistic intake, which serves as the child's representation of that utterance at this stage of development. This encoding process will depend on the child's ability to deploy her existing linguistic and extralinguistic knowledge in real time, given her developing cognitive abilities. The acquisitional intake is the portion of that representation relevant for the acquisition task at hand – for example, perhaps only syntactic structure may be relevant for learning about certain constraints on *wh*-dependencies (as in Pearl & Sprouse, 2013a), but perhaps conceptual information may be relevant for learning about the verb argument structure of *climb*. The acquisitional intake is determined by the child's learning biases about what information is relevant in the linguistic intake. For verb class learning, this pathway will determine how the age-appropriate child-directed speech samples serving as input are transformed into different acquisitional intakes, depending on the modeled learner's learning assumptions.

The second pathway is the **intake-inference** pathway, which takes the acquisitional intake and does inference on that intake to generate the most up-to-date hypotheses or generalizations about the linguistic system encoded by the developing grammar. The exact update procedures used will depend on the child's current learning biases. For example, a child might use purely statistical inference within a hypothesis space defined in terms of clusters of salient features, or a hypothesis-testing approach within a hypothesis space defined in terms of linguistic parameters. For verb class learning, this pathway will involve hierarchical Bayesian learning that generates the verb classes in the modeled learner's developing grammar (i.e., the learner's inferred classes), based on the syntactic, conceptual, and linking information in the acquisitional intake.

The third pathway is the **grammar-behavior** pathway. This pathway describes how the child's internal representations (encoded by the linguistic intake of the moment and the developing grammar) are transformed into various types of external behavior that we can observe, such as utterance generation, truth-value judgments, or looking times. This depends both on the state of the child's internal representations and the production systems that operate on those representations to produce observable behavior. For example, an internal representation of *What's she climbing over?* that involves both syntactic and conceptual information might cause a child to generate the utterance *What's she dancing on?* using her developing grammar, because the new utterance has similar syntactic and conceptual properties to the utterance in the linguistic intake. For verb class learning, this pathway will involve how the verb classes in the modeled learner's developing grammar (i.e., the inferred classes in the learner's output) compare to the verb classes derived from observed child behavior at ages three, four, and five.

By using this framework – and, more specifically, these three pathways – we can make theories of acquisition (which involve both theories of representation and theories of the learning process)

---

[4] What we call the linguistic intake has been referred to in the framework mentioned above as "perceptual intake" because it's what the child is capable of perceiving from the available input at that point in development; we choose "linguistic" to highlight that this representation includes more than just perceptual information.

explicit and testable against available empirical data (Pearl, 2014; Pearl & Sprouse, 2015; Pearl, 2017, in press). Here, this means that we can evaluate different theories of how to solve the linking problem by how well they enable a modeled learner to learn verb classes the way children seem to. More specifically, each modeled learner implements a combination of learning assumptions that corresponds to different theoretical claims (e.g., an absolute vs. relative thematic system, prior knowledge of the mapping vs. no prior knowledge). By seeing if a given modeled learner can learn the verb classes children do at different ages, we can evaluate the utility of these assumptions for acquisition.

4.2 The input-intake pathway

4.2.1 Input

Children's input signal can include both linguistic information (e.g., spoken or signed productions) and non-linguistic information (e.g., contextual information about intended meaning). We take realistic samples of this input signal from the CHILDES Treebank (Pearl & Sprouse, 2013a), which contains speech directed at children between one and five years old, annotated with linguistic and non-linguistic information. In particular, around 180,000 child-directed speech utterances from the BrownEve, BrownAdam, and Valian corpora (Brown, 1973; Valian, 1991) have been annotated with syntactic, conceptual, and thematic information. First, these utterances have syntactic phrase structure, based on an adapted version of the Penn Treebank annotation system. This annotation was done using a combination of automated and hand annotation (see Pearl and Sprouse (2013a) and the included readme file at http://www.socsci.uci.edu/~lpearl/CoLaLab/CHILDESTreebank/childestreebank.html for details). Second, animacy for each NP argument was annotated by hand. We included animacy because a number of acquisition studies have demonstrated that animacy is a useful cue for learning verb classes (Scott & Fisher, 2009; Becker, 2009; Kirby, 2009a, 2010; Becker & Estigarribia, 2013; Becker, 2014, 2015; Hartshorne et al., 2015). Third, thematic roles for the arguments of each verb (except the copula *be*) were annotated by hand using 13 thematic role labels that are common in the literature (again, see the readme file mentioned above for details).

We divided these utterances into age ranges based on the age of the child the speech was directed at: less than 3 years of age, less than 4 years of age, and less than 5 years of age. We then constructed datasets representing the input to a child of a particular age. We note that the datasets used as input for models of older children (e.g., <4yrs, representing a four-year-old child) include the data directed at younger children (e.g., <3yrs + data directed at children between the ages of three and four). This is because we assume older children would learn from all the data they've heard up until that point. Table 4 provides a detailed summary of the statistics for each input dataset.

4.2.2 Linguistic intake

From the input signal, children extract their linguistic intake. The information they extract depends on what information is salient to them and what they can plausibly extract from the input in real

Table 4: Child-directed speech data used as input to modeled three-year-old, four-year-old, and five-year-old learners. This includes the sources of these data in the CHILDES Treebank, the number of children the speech was directed at, the age range of the children the speech was directed at, the total number of utterances and words, the total number of verb types, and the number of verb types appearing 5 or more times in the dataset.

dataset	**sources**	**# children**	**ages**	**# utt**	**# words**	**# vbs**	**# vbs >5**
<3yrs	BrownEve, Valian	22	1;6-2;8	≈39.8K	≈197K	555	239
<4yrs	BrownEve, Valian, BrownAdam3to4	23	1;6-4;0	≈50.7K	≈254K	617	267
<5yrs	BrownEve, Valian BrownAdam3to4 BrownAdam4up	23	1;6-4;10	≈56.5K	≈285K	651	285

time. We consider several types of information children could plausibly extract for learning about verb classes: one syntactic, one conceptual, and one linking conceptual and syntactic information.

Syntactic information seems plausible, as children are known to be adept at syntactic bootstrapping – that is, using the syntactic context – when learning about verbs (Landau & Gleitman, 1985; Gleitman, 1990; Gillette, Gleitman, Gleitman, & Lederer, 1999; Fisher, Gertner, Scott, & Yuan, 2010; Gutman, Dautriche, Crabbé, & Christophe, 2015; Harrigan, Hacquard, & Lidz, 2016). One way to implement syntactic information is via phrase structure, with verb argument positions like *subject* labeled, as shown in (6a).

Another plausible information source is the concept of animacy (e.g., a penguin is animate, while an ice cube isn't). Animacy is something young children are known to both be sensitive to as a general property and also use as a cue in experimental studies to predict how verbs will behave (Scott & Fisher, 2009; Becker, 2009; Kirby, 2009a, 2010; Becker, 2014, 2015; Hartshorne et al., 2015). Moreover, if children are able to harness animacy effectively in their input, it's possible to use the animacy of a verb's arguments (in particular, whether the argument is *inanimate*) to distinguish verb behaviors such as those associated with subject-raising, subject-control, object-raising, and object-control (Kirby, 2009a, 2010; Becker & Estigarribia, 2013; Becker, 2014). One way to implement this conceptual information is for the verb's NP arguments to be labeled as +/-animate, as in (6b).

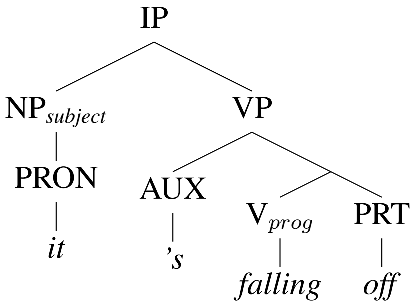
A third source of information corresponds directly to linking theories, as it concerns the link between conceptual information like thematic roles and syntactic position. More specifically, infants under a year old are sensitive to the presence of thematic roles (<10 months: Gordon, 2003; <6 months: Hamlin, Wynn, & Bloom, 2007; Hamlin, Wynn, Bloom, & Mahajan, 2011), making thematic roles a plausible information source for learning verb classes. UTAH and rUTAH assume a built-in mapping from the intermediate thematic representation (whether fixed proto-roles like

UTAH or an ordered hierarchy like rUTAH) to syntactic positions like *subject*; derived-mapping approaches using the same thematic systems don't assume this mapping is present initially. Importantly, all approaches require the child to extract the syntactic positions of the verb's arguments, and be aware of their thematic role, as shown in (6c).

Here, we make the simplifying assumption that the perceptual encoding process creating the linguistic intake is perfectly reliable (an assumption that can be relaxed in future work). Implementationally speaking, this means we assume that when given an input utterance like *it's falling off* from the BrownEve corpus in the CHILDES Treebank (Pearl & Sprouse, 2013a), we assume a linguistic intake that encodes syntactic and conceptual information, such as (6).

(6) Example linguistic intake for *it's falling off*.

a. Syntactic information for *falling*:



b. Animacy information:
it (*subject*$_{falling}$) = -animate

c. Thematic information:
it (*subject*$_{falling}$) = THEME$_{falling}$

4.2.3 The acquisitional intake

From this linguistic intake, the modeled learners extract their acquisitional intake. The exact acquisitional intake depends on the learning assumptions the learner is using.

For the syntactic information, syntactic frames encoding surface argument structure can be derived from the phrase structure of the verb usage. For example, the utterance *it's falling off* might yield a frame for *fall* involving the NP *subject* and the particle, either with or without the progressive morphology that surfaces on the verb itself (**+/-surface-morphology**), as in (7). Whether children heed the verbal surface morphology when encoding syntactic frames for their acquisitional intake is currently unknown, given available developmental data. Importantly, how the modeled learner deals with verbal morphology must be fixed before a modeled learner can be constructed. Since either option is plausible, we implement modeled learners of both kinds – that is, our modeled learners will also vary on whether they encode the verb's surface morphology in their syntactic frames.

(7) *fall* syntactic frame options for *it's falling off*

  a. +surface-morphology: NP V_{+prog} PRT
  b. -surface-morphology: NP V PRT

Another key point of variation is whether the mapping from the intermediate thematic representation is present or not. This affects whether the modeled learner expects a mapping *a priori* (**expect-mapping**). If the modeled learner expects a mapping (+expect-mapping), then it will be sensitive to violations of that expectation. Our modeled learners interpret these violations as instances of movement. That is, the learner will abstract away from the specific roles and positions, and instead only take in the fact that movement occurred, as shown in (8b). If instead the modeled learner doesn't yet expect a mapping (-expect-mapping), the learner will track the distribution of the intermediate thematic representation. That is, the learner will take in the details of which (proto-)role occurred in which position, as shown in (8c). In this way, the expectation of a mapping directly impacts the learner's acquisitional intake.

(8) Acquisitional intake for *The ice was melted by the girl*, using +/-expect-mapping

  a. *The ice* = PATIENT = *subject*
     the girl = AGENT = *object of PP*
  b. +expect-mapping
     (i) absolute (UTAH): proto-PATIENT = *subject*, proto-AGENT = *object of PP*
       Unexpected. Indicates +movement.
       Acquisitional intake: 2 instances of movement
     (ii) relative (rUTAH): 2ND-HIGHEST = *subject*, HIGHEST = *object of PP*
       Unexpected. Indicates +movement.
       Acquisitional intake: 2 instances of movement
  c. -expect-mapping
     (i) absolute: proto-PATIENT = *subject*, proto-AGENT = *object of PP*
       Acquisitional intake:
       1 proto-PATIENT as *subject*, 1 proto-AGENT as *object of PP*
     (ii) relative: 2ND-HIGHEST = *Subject*, HIGHEST = *object of PP*
       Acquisitional intake:
       1 2ND-HIGHEST as *subject*, 1 HIGHEST as *object of PP*

The different learning assumptions affecting the learner's acquisitional intake and their different combinations are shown in Table 5. Given the 3 binary choices (+/-surface-morphology, absolute/relative thematic system, and +/-expect-mapping), we implement 8 modeled learners: a +surface-morphology and -surface-morphology variant for learners using one of the two thematic systems and either expecting or not expecting a mapping. Note that all modeled learners use the animacy of a verb's arguments, in addition to syntactic frame information and thematic role information. Where they differ is how exactly they use the syntactic frame and thematic role information.

As an example, let's consider the different acquisitional intakes for the utterance *it's falling off*, whose linguistic intake was shown in (6). All learners encode one instance of an inanimate

Table 5: Example of a child's linguistic intake from (6) of the utterance *it's falling off* for the four modeled learner types. To save space, surface morphology is in parentheses to indicate the two modeling options for each learner type.

absolute thematic expected mapping	**absolute thematic no expected mapping**	**relative thematic expected mapping**	**relative thematic no expected mapping**
NP $\mathrm{V}_{(prog)}$ PRT	NP $\mathrm{V}_{(prog)}$ PRT	NP $\mathrm{V}_{(prog)}$ PRT	NP $\mathrm{V}_{(prog)}$ PRT
subject$_{-anim}$ = 1	*subject*$_{-anim}$ = 1	*subject*$_{-anim}$ = 1	*subject*$_{-anim}$ = 1
movement = 1	*subject*$_{proto-\text{PATIENT}}$ = 1	movement = 0	*subject*$_{\text{HIGHEST}}$ = 1

argument in *subject* position (*subject*$_{-anim}$). So, there's no difference in the acquisitional intake with respect to animacy. For those learners ignoring surface morphology on the verb, only the core verb frame would be extracted: NP V PRT. For learners heeding surface morphology, the fact that the verb is in the progressive would additionally be included: NP V_{prog} PRT.

The exact thematic information extracted depends on the thematic system (absolute/relative): with an absolute thematic system, the thematic role of the *subject* (THEME) is mapped to proto-PATIENT; with a relative thematic system, the learner uses the thematic role hierarchy to map the thematic role of the *subject* (THEME) to the HIGHEST role because it's the only thematic role present. If there's no expectation of mapping, the learner encodes the distribution of thematic representations (here, proto-PATIENT or HIGHEST in *subject* position). If there is in fact an expectation of mapping, the learner encodes whether the observed mapping obeys or violates that expectation. For the absolute thematic representation with a mapping expectation (UTAH), a proto-PATIENT in *subject* position violates the expected mapping and so is interpreted as movement; in contrast, for the relative thematic representation with a mapping expectation (rUTAH), the HIGHEST role in *subject* position obeys the expected mapping and so is interpreted as no movement.

4.3 The intake-inference pathway

Each modeled learner uses the acquisitional intake defined by its respective learning assumption combination to update its hypotheses about verb classes (i.e., its generalizations about which verbs behave alike in its developing grammar); a successful assumption combination will allow the learner to match children's observable behavior for verb classes. We implement this update process using hierarchical Bayesian inference, where the learner assumes the generative process depicted in Figure 1 (the generative process is represented with standard plate diagram notation for hierarchical Bayesian modeling). The observable verb data V in the acquisitional intake are generated by combining the available syntactic, animacy, and thematic information in the acquisitional intake, mediated by the latent representation of verb classes C. Below we provide high-level descriptions of the different components of the inference process shown in Figure 1. Interested readers can refer to Appendix E in the supplementary materials for more details on the inference implementation.

Observable data are available for each verb $v_j \in V$, in the form of the frames F that verb is used in, which include the syntactic structure, the animacy of the arguments, and the thematic roles

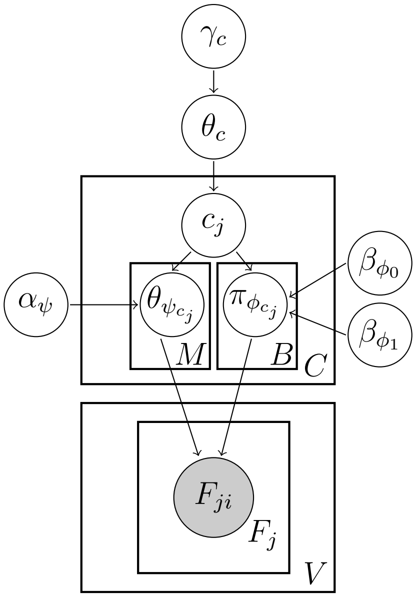


Figure 1: Plate diagram for a generative model of verb classes, based on syntactic, animacy, and thematic information from individual verbs in the input. Observable verb data V (and specifically verb frame instances F_{ji}) are generated based on the underlying verb class information C, which involves different characteristics M and B tracked by modeled learners (specifically, multinomial characteristics ψ like syntactic frame information and binomial characteristics ϕ like argument animacy information).

present. For example, the verb *fall* may appear multiple times, in instances such as *it's falling off*, *she fell down*, *don't fall!*, and *London Bridge is falling down*. Each frame instance F_j for a verb appears with some frequency F_{ji} – for example, *it's falling off* might occur three times.

The objective of the modeled learner is to infer the set of verb classes C that generate the observable verb data. Each verb v_j belongs to its verb class c_j. The learner doesn't know beforehand how many verb classes there are, what size they are, or which verb belongs to which. However, via the verb class hyperparameters θ_c and γ_c, the learner has a bias for classes distributed in a power law distribution, where a few classes have many verbs and the rest of the classes have few verbs.

Each verb class c_j has certain binomial characteristics B and multinomial characteristics M associated with it. Binary characteristics $\phi \in B$ include whether the *subject*, *object*, and oblique *object* are animate (+/-animate). If the modeled learner involves an expected mapping, then whether the mapping was violated (and so interpreted as movement) is also a binary characteristic. Each class will have some probability of preferring each option $\pi_{\phi_{c_j}}$. For example, a class c_j might prefer inanimate to animate subjects, with $\pi_{-animate_{subject}}$ = 0.70 and $\pi_{+animate_{subject}}$ = 0.30. During the course of learning, the learner infers these probabilities for each verb class. The hyperparameters (β_{ϕ_0}, β_{ϕ_1}) implement an initial uniform probability over the possible binary options, thereby implementing no bias *a priori*.

Multinomial characteristics $\psi \in M$ include which syntactic frame a verb appears in (e.g., NP

V PRT for *it's falling down*). If the modeled learner doesn't assume a mapping between thematic roles and syntactic positions, then the syntactic position is also a multinomial property (e.g., if the proto-AGENT appears in *subject*, *object*, or oblique *object* position). Each class will have some probability of preferring each option $\theta_{\psi_{c_j}}$. For example, a class c_j might primarily prefer the NP V PRT and NP V syntactic frames, giving them higher probabilities, and disprefer the frame NP V IP ($\theta_{NP\ V\ PRT}$ = 0.50, $\theta_{NP\ V}$ = 0.40, ..., $\theta_{NP\ V\ IP} \approx$ 0.00). During the course of learning, the learner infers these probabilities for each verb class. The hyperparameter α_ψ implements an initial uniform probability over the possible multinomial options, thereby implementing no bias *a priori*.

Importantly, the learner infers different verb classes precisely because the characteristics of verb classes differ sufficiently. In particular, given the observed instances of verb usage, the learner uses Bayesian inference to infer (i) how many verb classes there are, (ii) what the characteristics of each verb class are, and (iii) which class each verb belongs to. The best hypothesis is the one that maximizes the probability of the observed data, balanced against the prior preference for classes distributed in a power law distribution.

This inference is accomplished via Gibbs sampling operating over the data as a single batch (see Appendix E in the supplementary materials), which is guaranteed to converge on the optimal answer if given sufficient time to search the hypothesis space (i.e., Gibbs sampling is an optimal inference process). This is part of what makes the modeled learners *ideal learners* – the inference computation is implemented by an optimal inference process that is *not* intended to be realistically constrained. Instead, humans likely approximate this inference process to accomplish the same computation and execute inference incrementally as data are encountered.

A reasonable question is why we should use an ideal inference process rather than a realistically constrained process to model language acquisition. Typically, acquisition modelers will start with an optimal inference process in order to know if the mental computation specified by the model is a potential match to human behavior (here, child language acquisition behavior) (Pearl, in press). If not, this is a signal that the learning assumptions encoded in the model are unlikely to be right. That is, if a modeled learner can't get close to human behavior even when the mental computation is performed as perfectly as possible, then that computation is probably not the right one. This would mean the learning assumptions that circumscribe that mental computation (here: using syntactic, animacy, and thematic information in particular ways) are not useful.

In contrast, if a modeled learner using optimal inference can match human behavior, this suggests the learning assumptions are plausible. Subsequent work could then explore how acquisition unfolds when inference is non-optimal (e.g., subject to the cognitive constraints children have and the incremental nature of learning). In the meantime, the ideal learning model using optimal inference serves as a useful proof-of-concept in the search for learning assumptions that can potentially solve the acquisition problem under investigation. More generally, it's important to determine that learning assumptions are potentially useful to children before investigating if they're usable by children. This is the approach we pursue here.

4.4 The grammar-behavior pathway

This pathway determines how a modeled learner's output will be evaluated when the target is observed behavior. In section 3.2, we described the verb classes derived from observed child behavior. It's reasonable to believe that such verb classes are a legitimate target state reflecting children's underlying knowledge because of how we think of the grammar-behavior pathway. In particular, we assume here that if children's comprehension and/or production indicate that they treat two verbs similarly with respect to some verb behavior (e.g., being passivizable), this transparently reflects children's developing grammars – that is, the two verbs in question are clustered together in children's minds with respect to that verb behavior. If children's verb comprehension and/or production indicate two verbs are clustered together for all currently tested verb behaviors for those verbs, then we assume that the verbs are in the same class in children's developing grammars. This is how the verb classes in Tables 7-9 in Appendix C of the supplementary materials were derived.

We take these verb classes, derived for children ages three, four, and five, as representative of the developing grammars of children of these ages. So, modeled learner output is compared against them using the measures discussed in section 3.3. The modeled learners whose inferred classes best match these child verb classes at particular ages can be thought to encode the learning assumptions that children have at those ages.

5 Modeling results

Recall that each of the eight modeled learners uses a different combination of learning assumptions, based on how linguistic and non-linguistic information are used (Table 5). For each learner, we ran an ideal learner implementation 10 times over each age-based dataset (<3yrs, <4yrs, or <5yrs). The resulting ten sets of inferred verb clusterings were aggregated into a single set of verb classes, using a simple threshold: any verb pair together in more than 75% of the runs (i.e., >7 of 10) was put together in the aggregate verb clustering; similarly, any verb that was in a class of its own (a *singleton*) for more than 75% of the run was put as a singleton in the aggregate verb clustering for that modeled learner. Each modeled learner's aggregate verb clusterings for each dataset appear in Appendix F in the supplementary materials. Figure 2 shows RI when compared to the child verb classes relevant for different ages of acquisition (i.e., verb classes learned by age three for the <3yrs dataset; verb classes learned by age four for the <4yrs dataset; verb classes learned by age five for the <5yrs dataset). For the RI randomization tests, we use a threshold of p<.01 for significance (two-tailed). We indicated the threshold for the null hypothesis (randomizing all three parameters) with a solid horizontal line. We also added a single asterisk (*) to models that are significant under this null hypothesis. The learners surpassing the p<.01 threshold are summarized in Table 6.

Taken together, what stands out is that there are learners at each age who are doing better than chance, though the collection of learning assumptions that successful learners encode varies by age. Below we interpret these results with respect to the four linking theory proposals: the innate-mapping UTAH and rUTAH, and their derived-mapping equivalents.

UTAH assumes an absolute thematic system and innate knowledge of the linking pattern. This set of assumptions (absolute, +expect-mapping) is not compatible with the assumptions of successful modeled learners at three years old, though it is at four and five years old. So, UTAH would need to be coupled with a late-maturation developmental theory, where the absolute the-

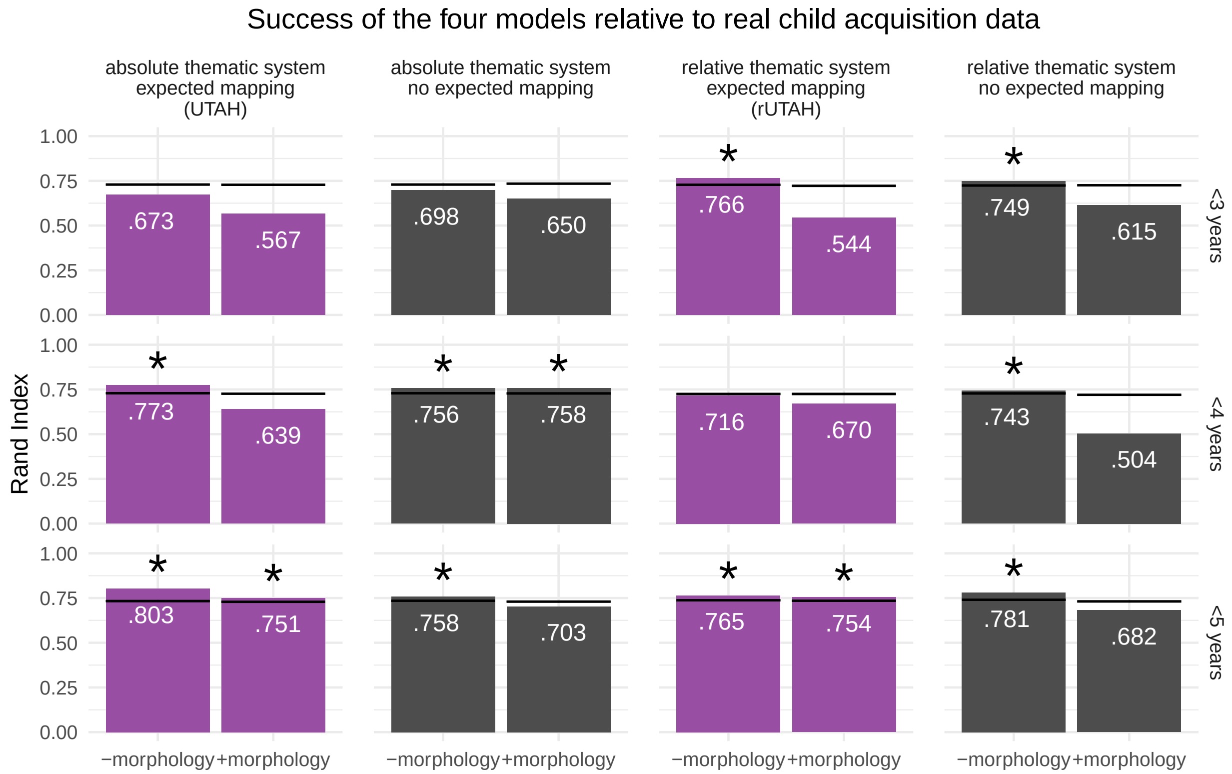


Figure 2: Rand Index (RI) scores for all modeled learners. The four major modeled learner types are organized into columns, with classic UTAH and rUTAH in purple. Surface morphology is nested within each major modeled learner type. The solid horizontal lines indicate the (upper) $p<.01$ threshold for the (two-tailed) randomization tests randomizing across all three parameters in the model: number of classes, size of classes, and verb assignment to classes. An asterisk (*) means the result was significant for the randomization test. The white numbers within each bar report the RI index value to three decimal places.

Table 6: Modeled learners by age passing the $p<.01$ threshold, based on RI scores. Learning assumptions shown are which thematic system is used (absolute/relative), whether a mapping from thematic roles to syntactic positions is expected (expect-mapping), and whether surface morphology on verbs is heeded for verb syntactic frames (surface-morphology). Each row represents a set of modeled learners above threshold at matching children's verb classes.

age	**thematic system**	**expect-mapping**	**surface-morphology**
<3yrs	relative	+/-	-
<4yrs	absolute	+/-	-
	absolute/relative	-	+
<5yrs	absolute/relative	+	+/-
	absolute/relative	-	-

matic knowledge and innate linking knowledge only become available at age four or later. We note that if such knowledge emerges at four, children would also need to ignore verb surface morphology.

rUTAH assumes a relative thematic system and innate knowledge of the linking pattern. This set of assumptions (relative, +expect-mapping) is compatible with the assumptions of successful learners at three and five years old, but not at four years old. So, rUTAH would need to be coupled with a developmental theory where the innate linking knowledge is either (a) inaccessible at age four for some reason, or (b) actually isn't accessible until age five, and so children at age three don't have access to the innate linking knowledge (relative, -expect-mapping). That is, rUTAH requires either a U-shaped developmental theory or a late-maturation developmental theory.

The derived-mapping variant of UTAH assumes an absolute thematic system and derived knowledge of the linking pattern. This set of assumptions (early: absolute, -expect-mapping; late: absolute, +expect-mapping) doesn't seem obviously compatible with the assumptions of successful learners at three years old. This is because all successful learners at this age rely on the relative thematic system. So, the child would need to derive both the absolute thematic system and the linking pattern after this age. At four and five however, there are successful learners relying on the absolute thematic system. So, a derived-mapping UTAH child could have derived both this thematic system and the linking pattern knowledge at four or five. If linking knowledge is derived by four, the child would be ignoring surface morphology; if linking knowledge is derived by five, the child could heed or ignore surface morphology.

The derived-mapping variant of rUTAH assumes a relative thematic system and derived knowledge of the linking pattern. This set of assumptions (early: relative, -expect-mapping; late: relative, +expect-mapping) is compatible with the assumptions of successful learners at all ages. For example, at three, children relying on a relative thematic system wouldn't expect a mapping (and would also ignore surface morphology); at four, they wouldn't expect a mapping (but now would heed surface morphology); at five, they would have derived linking knowledge and expect a mapping (and either heed or ignore surface morphology).

Taken together, our results highlight the connection between theories of representation and theories of development. While our results are compatible with both innate-mapping approaches (UTAH, rUTAH) and their derived-mapping equivalents, they argue against an early-maturation innate theory of development. That is, neither of the innate-mapping linking theory proposals seem immediately compatible with early-maturation innate knowledge. Instead, the linking knowledge (and sometimes the thematic knowledge) would need to develop later (late-maturation innate-mapping); conversely, the linking knowledge (and sometimes the thematic knowledge) could be derived from language experience, as in the derived-mapping approaches. To choose among these linking theory proposals, we therefore need more empirical data about the other learning assumptions (i.e., thematic systems and attention to surface morphology) that English children use at ages three, four, and five when creating verb classes. We return to this empirical need in the next section.

6 Discussion

We believe that a significant contribution of this work is the integrated quantitative framework itself, which provides a concrete way for linking theory proposals to both (i) generate developmental predictions, and (ii) be evaluated on those developmental predictions. In particular, the framework implements an acquisition process that is both empirically-grounded and theoretically-motivated. Empirical data determine the modeled learner's input, desired output, and motivate the probabilistic learning mechanism used for inference; theoretical proposals determine the representations that control how the learner's input is transformed into the intake that drives learning. Below we discuss the results of using this framework for the particular linking proposals investigated here, and also how this framework can be used in the future as (i) more empirical data become available, or (ii) different theories of representation and/or development are proposed.

6.1 Implications for theories of representation and development

Given the complexity of the learning problem of creating dozens of verb classes, and given that the learning assumptions implemented in the modeled learners here only involve a subset of all the possible information children could be using to solve that problem, our first noteworthy result is that *any* of the learning assumption combinations are successful. Though we started this project from the assumption that syntactic frames, thematic information, and animacy information from child-directed input would be sufficient to learn verb classes, there's no *a priori* reason to believe that this information would be sufficient. So, these results empirically support the common assumption in the literature that these specific pieces of information are sufficient to learn verb classes the way children seem to.

From the perspective of theories of linguistic representation, these results have two implications. First, both UTAH and rUTAH are reasonably accurate at capturing children's representations at some point in development. This provides developmental support that they are plausible representational theories. Moreover, because they are compatible with the oldest children's verb behavior (at five years old), they are also plausible representational theories for adults.

However, one particularly notable finding is that not all the options capture younger children's behavior equally (at three and four years old). The results here suggest three-year-olds are more likely to rely on a relative thematic system, while older children may not. This has real implications for what needs to be built in to yield the linguistic development we observe in children. Here, it seems that the conceptual categories corresponding to proto-roles are *not* required (which UTAH relies on); moreover, there may not need to be a built-in expectation of a specific mapping between thematic roles and syntactic positions (as early-maturation innate-mapping approaches would predict). Instead, both types of knowledge could potentially develop later (or be accessed later if children don't have sufficient cognitive resources to do so earlier in development for some reason).

Both innate-mapping and derived-mapping approaches are compatible with these results, but then require different promissory notes. For late-maturation innate-mapping approaches, a developmental account is needed either for (i) why the knowledge itself develops later, or (ii) why children's access to this knowledge develops later. Either avenue also requires evidence from developmental neurobiology. For derived-mapping approaches, the required knowledge would be derived from language experience, rather than being innately specified. Under this approach, it remains to be seen exactly how the conceptual categories of the absolute thematic system and the expectation of a mapping would be derived from children's input. That is, a viable derived-mapping approach should demonstrate what prior knowledge and abilities are needed in combination with children's input to derive both the appropriate conceptual categories and the appropriate mapping. Future computational modeling may be able to contribute to this investigation. Moreover, another future step would focus on the precise mechanisms that derived-mapping approaches can use to explain the distribution of linking patterns across languages. In particular, derived-mapping approaches would need to demonstrate how the linking patterns observed cross-linguistically can be derived from children's input in each language.

6.2 Open questions

There are a number of open questions that the current results highlight, both in terms of the empirical foundations and the theories of representation and development. Here we classify these questions into three types: experimental, computational, and theoretical avenues of inquiry.

6.2.1 Experimental avenues

One avenue for future experimental work is to increase the number of verbs and verb classes that are used in early acquisition studies. Though our corpus analysis yielded up to 285 verbs appearing five or more times (<3yrs: 239, <4yrs: 267, <5yrs: 285) in the CHILDES Treebank, the available experimental data about specific verb behaviors yielded far fewer verbs on which to evaluate our simulations (<3yrs: 60, <4yrs: 76, <5yrs: 84). This means there are nearly 200 verbs for each age group that we have model predictions for, but no behavioral data about (and therefore were not reported here). With more targeted child language experiments, we'll have a broader empirical basis to evaluate our acquisition theories against. For example, at three years old, there are two modeled learners who best match what we know currently about three-year-old verb classes. These learners both rely on the relative thematic system and ignore surface morphology on verbs. The learner which doesn't expect a mapping puts together *keep* and *stop* in one class and *miss* and *say* in a separate class, while the learner which does expect a mapping groups all these verbs together into the same class. Do three-year-olds expect different behaviors for these four verbs, or the same behaviors? Once we know, we can better choose between the two learning assumption combinations that currently best fit three-year-old behavior.

This lack of behavioral data also applies to the verb behaviors we know about – here, there were 12 attested verb behaviors, but there are many more where we need knowledge of how specific verbs behave (e.g., intransitivity, monotransitivity, unergativity, verbs taking non-finite complements with *-ing*, verbs taking small clause complements, *wager*-class verbs). Again, with a broader child behavioral foundation, we'll be better able to choose among the modeled learner options and the learning assumptions they encode.

6.2.2 Computational avenues

One avenue for computational work is complementary to the future experimental work with children. Each modeled learner here has generated a set of verb classes which is that learner's internal representation of which verbs behave like other verbs. Each verb class has a set of characteristics (involving syntactic and conceptual preferences) that can be used to generate precise predictions for any experimental setup. For example, we can calculate the probability distribution over verbs that a modeled child will prefer to use with a particular utterance that has certain syntactic and conceptual characteristics (e.g., *She* _ *to laugh* = $subject_{+anim}$, NP V IP$_{-finite}$, $subject_{\text{EXPERIENCER}}$). This corresponds to what might be observed in child productions. We can also calculate the probability distribution over utterances that a modeled child will prefer to use for a particular verb (e.g., *want* might have a high probability for *She* _ *to laugh* while *make* has a low probability). This corresponds to both the productions a child might generate, and also the ease with which a child would comprehend a verb used in a particular utterance. Both of these are examples of the modeled learner generating concrete behavioral predictions that can be experimentally evaluated. When the predictions diverge and only one matches children's behavior, we then have additional empirical support for whichever modeled child (and therefore, whichever specific combination of learning assumptions) was successful.

We can also make more sophisticated computational models that capture both the incremental nature of children's learning and children's cognitive constraints. Recall that the ideal learner model implementations here operate by learning from all the data at once that children of a certain age will have seen, and optimally executing the inference over those data. As mentioned, this is a first step in understanding the mental computations that occur during acquisition. Future work can relax some of the idealized assumptions present in the ideal modeled learners used here. For example, one option is to make more realistic learners that (i) learn from data as they're encountered one utterance at a time (rather than as a batch), and (ii) use an inference approximation, rather than Gibbs sampling, to converge on the final set of verb classes (e.g., see the learning approaches of Fazly, Alishahi, & Stevenson, 2010, Barak, Fazly, & Stevenson, 2014b, and Barak, Fazly, & Stevenson, 2014a). Unlike the ideal learner model implementations, these more realistic modeled learners would be executing a potential inference algorithm that children could be capable of – this makes these future models algorithmic-level (rather than computational-level) in the sense of Marr (1982).

The utility of algorithmic-level implementations is to see if the learning assumptions that were useful for a computational-level learner are *still* useful when incremental learning and cognitive constraints are present (Pearl, 2014; Phillips & Pearl, 2015; Pearl, in press). That is, algorithmic-level implementations can tell us if the learning assumptions that seem to be useful for ideal learners are actually *usable* by real children, who have various constraints on their acquisition computation. This isn't always the case – it could turn out that certain learning assumptions are less helpful to a cognitively constrained learner while other assumptions are more helpful (Phillips & Pearl, 2015; Pearl & Phillips, 2018).

Assuming that the developmental trajectory suggested by these results holds under future experimental and (incremental) modeling work, another open question that can be investigated via computational methods is how the primary linking pattern and the secondary exception patterns arise under a derived-mapping approach. That is, how could the expectation for the "right" mapping between thematic representations and syntactic positions develop between ages three and five? Without a built-in expectation of specific mappings, these patterns are dependent on the content of the input in combination with whatever prior knowledge and abilities children have. If there are a sufficient number of primary pattern verbs in the input (and/or verb classes) learned at early stages, then this will lead to the development of the primary pattern expectation. Mathematical analyses of children's input that predict when children will make a generalization vs. not, such as the Tolerance Principle (Yang, 2005; Legate & Yang, 2013; Schuler, Yang, & Newport, 2016; Yang & Montrul, 2017), can provide an answer. Such analyses can either support the ability of realistic input to help children derive the primary mapping or demonstrate the obstacles to be surmounted under the derived-mapping approach.

6.2.3 Theoretical avenues

From a theoretical perspective, there may be other solutions to the linking problem that we wish to investigate using this integrated quantitative framework. Here, we focused on two prominent options discussed in the theoretical literature (UTAH and rUTAH) that (i) take thematic roles as their basis, and (ii) involve either an absolute (UTAH) or relative (rUTAH) perception of these thematic roles. While these both seem plausible, other options are certainly available. For example, perhaps children abstract across thematic roles in different categorical or relative ways than the implementations explored here (more than 3 proto-roles, different definitions of protoroles, different orderings in the role hierarchy, etc). Relatedly, there could also be different thematic role distinctions at the basic conceptual level – the 13 roles here were chosen to make the CHILDES Treebank as useful as possible to the widest range of users (Pearl & Sprouse, 2013a). That said, there are a number of specific proposals for thematic role systems in the literature; the diversity of theories only increases when we consider that children's thematic distinctions might differ from adults' in complex ways (especially very young children's). It could also be that children begin by not abstracting over thematic roles at all. Instead, they might track mappings from the individual thematic roles directly to syntactic positions. Finally, as briefly mentioned in section 2.6, it's also possible that the source of the linking patterns we see lies outside of syntax (so, not in principles like UTAH or rUTAH), and is instead a consequence of a constraint on the types of semantic representations that language allows (Wood, 2015; Kastner, 2016; Myler, 2016). This is still a type of innate knowledge; therefore, the quantitative framework developed here could be modified to compare modeled learners with knowledge of that constraint versus modeled learners without knowledge of that constraint.

Related to the idea of different underlying thematic systems and how they might change during development, there may also be a change to the information children are sensitive to in the input. For example, while younger children may rely on syntactic frames, older children may rely on additional and/or more abstract syntactic information. For example, when encountering the utterance *She seemed to laugh*, a younger child might extract the syntactic frame NP __ $\text{IP}_{-finite}$ for *seem*. In contrast, an older child might also perceive the raising dependency, and so encode *seem*'s syntax as NP_1 __ $[_{IP}\ t_1\ \text{VP}_{-finite}]$. Knowing exactly what information children of different ages are able to both extract from their input and use for learning depends on having precise theories of acquisition that combine developing representations with developing abilities to use those representations in real time.

Moreover, it's also important to reconcile any current and future findings with existing child behavioral data. For example, both the late-maturation innate-mapping and derived-mapping approaches supported by our results here will need to account for data suggesting children do have some early mapping preferences (Naigles, 1990; Naigles & Kako, 1993; Bunger & Lidz, 2004; Gertner, Fisher, & Eisengart, 2006; Bunger & Lidz, 2008; Hartshorne et al., 2015). We leave this exciting theoretical work for the future.

7 Conclusion

To successfully learn language – and more specifically how to use verbs correctly – children must solve the linking problem: they must learn the mapping between the thematic roles specified by a verb's lexical semantics and the syntactic argument positions specified by a verb's syntactic frame. Here, we've constructed an argument from acquisition for different theoretical approaches to solving the linking problem. In particular, we've used acquisition of verb classes as an evaluation metric for theories of solving the linking problem, with the idea that a good theory will be able to account for children's developing knowledge of verb classes over time. We made different theoretical options concrete within an integrated quantitative framework of the acquisition process that relies on corpus analysis, experimental meta-analysis, and computational modeling. More specifically, we compared different underlying thematic representations to be linked to syntactic positions, as well as when prior knowledge of a mapping is available.

Our results allowed us to specify for the first time a developmental trajectory of mental representations and learning assumptions children may have when learning verb classes. Importantly, this specification is compatible with both innate-mapping and derived-mapping approaches to solving the linking problem, in combination with other learning assumptions about the thematic system and attention to verbal surface morphology. However, our results argue against early-maturation innate theories of development for either UTAH or rUTAH. An advantage of innate-mapping approaches like UTAH and rUTAH is that they can easily explain the cross-linguistic regularity of linking patterns. So, one fruitful avenue of future work for derived-mapping approaches is to understand how children derive the regularity we see in linking patterns from their input. Beyond this, our results support relative thematic representations in three-year-olds, with both absolute and relative thematic representations potentially available for four- and five-year-olds. More generally, our quantitative approach to language acquisition allows us to connect together theories of linguistic representation and theories of the learning process, and so better understand both as part of an integrated theory of language.

References

Baker, M. C. (1988). *Incorporation: A theory of grammatical function changing*. Chicago, IL: University of Chicago Press.

Baker, M. C. (1997). Thematic roles and syntactic structure. In *Elements of grammar* (pp. 73–137). Springer.

Barak, L., Fazly, A., & Stevenson, S. (2014a). Gradual acquisition of mental state meaning: a computational investigation. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 36).

Barak, L., Fazly, A., & Stevenson, S. (2014b). Learning verb classes in an incremental model. In *Proceedings of the Fifth Workshop on Cognitive Modeling and Computational Linguistics* (pp. 37–45).

Becker, M. (2006). There began to be a learnability puzzle. *Linguistic Inquiry*, *37*(3), 441–456.

Becker, M. (2007). Animacy, expletives, and the learning of the raising-control distinction. *Generative approaches to language acquisition North America*, *2*, 12–20.

Becker, M. (2009). The role of np animacy and expletives in verb learning. *Language Acquisition*, *16*(4), 283–296.

Becker, M. (2014). *The acquisition of syntactic structure: Animacy and thematic alignment* (Vol. 141). Cambridge University Press.

Becker, M. (2015). Animacy and the Acquisition of Tough Adjectives. *Language Acquisition*, *22*(1), 68–103.

Becker, M., & Estigarribia, B. (2013). Harder words: Learning abstract verbs with opaque syntax. *Language Learning and Development*, *9*(3), 211–244.

Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ-theory. *Natural Language & Linguistic Theory*, *6*(3), 291–352.

Bloom, L., Rispoli, M., Gartner, B., & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language*, *16*(01), 101–120.

Bloom, L., Tackeff, J., & Lahey, M. (1984). Learning *to* in complement constructions. *Journal of Child Language*, *11*(02), 391–406.

Bowerman, M. (1988). The 'No Negative Evidence' problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford, England: Blackwell.

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

Braine, M. D., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar.

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

Bunger, A., & Lidz, J. (2004). Syntactic bootstrapping and the internal structure of causative events. In *Proceedings of the annual boston university conference on language development* (Vol. 28, pp. 74–85).

Bunger, A., & Lidz, J. (2008). Thematic relations as a cue to verb class: 2-year-olds distinguish unaccusatives from unergatives. In (Vol. 14, p. 4).

Burzio, L. (1986). *Italian syntax: A government-binding approach* (Vol. 1). Springer Science & Business Media.

Campbell, A. L., & Tomasello, M. (2001). The acquisition of english dative constructions. *Applied Psycholinguistics*, *22*(02), 253–267.

Conwell, E., & Demuth, K. (2007). Early syntactic productivity: Evidence from dative shift. *Cognition*, *103*(2), 163–179.

Crain, S., Thornton, R., & Murasugi, K. (2009). Capturing the evasive passive. *Language Acquisition*, *16*(2), 123–133.

Déprez, V., & Pierce, A. (1993). Negation and functional projections in early grammar. *Linguistic Inquiry*, 25–67.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in english: A corpus-based analysis. *Cognitive Linguistics*, *12*(2), 97–141.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 547–619.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory.* New York: Holt, Rinehart, & Winston, Inc.

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149.

Gelman, S. A., & Koenig, M. A. (2001). The role of animacy in children's understanding of 'move'. *Journal of Child Language*, *28*(03), 683–701.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684–691.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). Introducing Markov Chain Monte Carlo. In *Markov chain Monte Carlo in practice* (p. 19). Suffolk.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure.* University of Chicago Press.

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

Goldberg, A. E. (2013). Argument structure constructions versus lexical rules or derivational verb templates. *Mind & Language*, *28*(4), 435–465.

Goldwater, S., & Griffiths, T. (2007). A fully Bayesian approach to unsupervised part-of-speech tagging. In *Annual meeting-Association for Computational Linguistics* (Vol. 45, p. 744).

Gordon, P. (2003). The origin of argument structure in infant event representations. In *Proceedings of the 28th Annual Boston University Conference on Language Development* (Vol. 1, pp. 189–198).

Gordon, P., & Chafetz, J. (1990). Verb-based versus class-based accounts of actionality effects in children's comprehension of passives. *Cognition*, *36*(3), 227–254.

Grimshaw, J. (1990). *Argument structure.* Cambridge, MA: The MIT Press.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in english. *Language*, 203–257.

Gutman, A., Dautriche, I., Crabbé, B., & Christophe, A. (2015). Bootstrapping the syntactic bootstrapper: Probabilistic labeling of prosodic phrases. *Language Acquisition*, *22*(3), 285–309.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559.

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences*, *108*(50), 19931–19936.

Harrigan, K., Hacquard, V., & Lidz, J. (2016). Syntactic Bootstrapping in the Acquisition of Attitude Verbs: think, want and hope. In *Proceedings of WCCFL* (Vol. 33).

Hartshorne, J. K., Pogue, A., & Snedeker, J. (2015). Love is hard to understand: The relationship between transitivity and caused events in the acquisition of emotion verbs. *Journal of Child Language*, *42*(03), 467–504.

Hirsch, C., & Wexler, K. (2007). The late development of raising: What children seem to think about seem. In *New horizons in the analysis of control and raising* (pp. 35–70). Springer.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218.

Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language*, *50*(2), 182–195.

Jackendoff, R. (1987). The status of thematic relations in linguistic theory. *Linguistic inquiry*, *18*(3), 369–411.

Kastner, I. (2016). *Form and meaning in the hebrew verb* (Unpublished doctoral dissertation). New York University, New York City.

Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development*, *21*(2), 93–107.

Kidd, E., Lieven, E. V., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences*, *32*(1), 132–142.

Kirby, S. (2009a). Do what you know:“semantic scaffolding” in biclausal raising and control. In *Annual meeting of the berkeley linguistics society* (Vol. 35, pp. 190–201).

Kirby, S. (2009b). *Semantic scaffolding in first language acquisition: The acquisition of raising-to-object and object control* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill, Chapel Hill, NC.

Kirby, S. (2010). Semantic scaffolding in l1a syntax: Learning raising-to-object and object control. In *Proceedings of the 2009 mind-context divide workshop* (pp. 52–59).

Kirby, S. (2011). Move over, control freaks: Syntactic raising as a cognitive default. *BUCLD 35 Online Proceedings Supplement. http://www.bu.edu/bucld/proceedings/supplement/vol35*.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child*. Harvard University Press.

Larson, R. K. (1988). On the double object construction. *Linguistic Inquiry*, *19*(3), 335–391.

Larson, R. K. (1990). Double objects revisited: Reply to Jackendoff. *Linguistic Inquiry*, *21*(4), 589–632.

Legate, J., & Yang, C. (2013). Assessing Child and Adult Grammar. In R. Berwick & M. Piatelli-Palmarini (Eds.), *Rich Languages from Poor Inputs* (pp. 168–182). Oxford, UK: Oxford University Press.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Levin, B., & Rappaport Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface* (Vol. 26). MIT press.

Lidz, J., & Gagliardi, A. (2015). How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics*, *1*(1), 333–352.

MacKay, D. J., & Peto, L. C. B. (1995). A hierarchical Dirichlet language model. *Natural Language Engineering*, *1*(3), 289–308.

Maratsos, M., Fox, D. E., Becker, J. A., & Chalkley, M. A. (1985). Semantic restrictions on children's passives. *Cognition*, *19*(2), 167–191.

Maratsos, M. P. (1974). Children who get worse at understanding the passive: A replication of bever. *Journal of Psycholinguistic Research*, *3*(1), 65–74.

Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.

Messenger, K., Branigan, H., McLean, J., & Sorace, A. (2009). Semantic factors in young children's comprehension and production of passives. In *Proceedings of the 33rd boston university conference on language development* (pp. 355–366).

Myler, N. (2016). *Building and interpreting possession sentences*. Cambridge, MA: MIT Press.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of child language*, *17*(02), 357–374.

Naigles, L., & Kako, E. (1993). First contact in verb acquisition: Defining a role for syntax. *Child development*, *64*(6), 1665–1687.

Nguyen, E., Lillo-Martin, D., & Snyder, W. (2016). Actionality speaks louder than felicity: Children's comprehension of long passives. In *Proceedings from the Generative Approaches to Language Acquisition.*

O'Brien, K., Grolla, E., & Lillo-Martin, D. (2006). Long passives are understood by young children. In *Proceedings from the 30th boston university conference on language development* (pp. 441–451).

Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, *22*(2), 158–192.

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, *105*(1), 125–165.

Pearl, L. (2014). Evaluating learning strategy components: Being fair. *Language*, *90*(3), e107–e114.

Pearl, L. (2017). Evaluation, use, and refinement of knowledge representations through acquisition modeling. *Language Acquisition*, *24*(2), 126–147.

Pearl, L. (in press). Modeling syntactic acquisition. In J. Sprouse (Ed.), *Oxford Handbook of Experimental Syntax.* Oxford University Press.

Pearl, L., Ho, T., & Detrano, Z. (2016). An argument from acquisition: Comparing English metrical stress representations by how learnable they are from child-directed speech. *Language Acquisition*, 1–36.

Pearl, L., & Phillips, L. (2018). Evaluating language acquisition models: A utility-based look at Bayesian segmentation. In A. Villavicencio & T. Poibeau (Eds.), *Language, Cognition and Computational Models* (p. 185-224). Cambridge: Cambridge University Press.

Pearl, L., & Sprouse, J. (2013a). Computational Models of Acquisition for Islands. In J. Sprouse

& N. Hornstein (Eds.), *Experimental Syntax and Islands Effects* (pp. 109–131). Cambridge: Cambridge University Press.

Pearl, L., & Sprouse, J. (2013b). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition*, *20*, 19–64.

Pearl, L., & Sprouse, J. (2015). Computational modeling for language acquisition: A tutorial with syntactic islands. *Journal of Speech, Language, and Hearing Research*, *58*, 740–753.

Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 4, pp. 157–190).

Perlmutter, D. M., & Postal, P. M. (1984). Impersonal passives and some relational laws. *Studies in Relational Grammar*, *2*, 126–170.

Phillips, L., & Pearl, L. (2015). The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive Science*, *doi:10.1111/cogs.12217*.

Postal, P. (1971). *Crossover phenomena*. New York: Holt, Rinehart and Winston.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850.

Resnik, P., & Hardisty, E. (2010). *Gibbs sampling for the uninitiated* (Tech. Rep.). University of Maryland Institute for Advanced Computer Studies.

Schuler, K., Yang, C., & Newport, E. (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting, Philadelphia, PA.*

Scott, R. M., & Fisher, C. (2009). Two-year-olds use distributional cues to interpret transitivity-alternating verbs. *Language and cognitive processes*, *24*(6), 777–803.

Snedeker, J., & Huang, Y. T. (2015). Sentence processing. In E. L. Bavin & L. R. Naigles (Eds.), *The Cambridge Handbook of Child Language.* Cambridge, England: Cambridge University Press.

Snyder, W., & Stromswold, K. (1997). The structure and acquisition of english dative constructions. *Linguistic Inquiry*, 281–317.

Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 859–890.

Speas, M. (1990). *Phrase structure in natural language* (Vol. 21). Springer Science & Business Media.

Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language*, *58*(2), 188–213.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based approach to child language acquisition*. Harvard University Press.

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

Wood, J. (2015). *Icelandic morphosyntax and argument structure*. Dordrecht: Springer.

Yang, C. (2005). On productivity. *Yearbook of Language Variation*, *5*, 333–370.

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

Yuan, S., & Fisher, C. (2009). “really? she blicked the baby?” two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*(5), 619–626.

Supplementary Materials

A UTAH and rUTAH implementations

Here we provide a more explicit walk-through of the UTAH and rUTAH implementations we investigated.

A.1 UTAH

(9) Example UTAH utterance mappings

a. Available roles in *Jack cut the pie with a knife*
= AGENT, PATIENT, INSTRUMENT
AGENT = proto-AGENT, PATIENT = proto-PATIENT, INSTRUMENT = OTHER

b. Available roles in *Lily fears spiders*
= EXPERIENCER, SUBJECT MATTER
EXPERIENCER = proto-AGENT, SUBJECT MATTER = proto-PATIENT

c. Available roles in *Spiders frighten Lily*
= CAUSER, EXPERIENCER
CAUSER = proto-AGENT, EXPERIENCER = proto-PATIENT

d. Available roles in *Jack sneezed during the meeting*
= AGENT
AGENT = proto-AGENT

e. Available roles in *The package arrived during the meeting*
= PATIENT
PATIENT = proto-PATIENT

For primary pattern sentences like (9a), the *subject* is a proto-AGENT, the direct *object* is a proto-PATIENT, and the oblique object is OTHER. For the psych-verbs in (9b)-(9c), this implementation of UTAH leverages the internal-vs-external causation distinction: in *Lily fears spiders*, *Lily* is causing her own mental state, and is thus a proto-AGENT; in *Spiders frighten Lily*, *spiders* are causing Lily's mental state, and thus *Lily* is the proto-PATIENT. For the unergative *sneezed* in (9d), *Jack* is the proto-AGENT, and mapped to the *subject*. For the unaccusative *arrived* in (9e), this implementation would claim that *the package* enters the syntactic derivation as the *object* of *arrive*, thus respecting UTAH. *The package* would then be moved to the *subject* position by an additional mechanism (such as the movement operation in GB/Minimalism).

A.2 rUTAH

(10) Example rUTAH utterance mappings

a. Available roles in *Jack cut the pie with a knife*
= AGENT, PATIENT, INSTRUMENT

AGENT > PATIENT > INSTRUMENT
AGENT = HIGHEST, PATIENT = 2ND-HIGHEST, INSTRUMENT = 3RD-HIGHEST

b. Available roles in *Lily fears spiders*
= EXPERIENCER, SUBJECT MATTER
EXPERIENCER > SUBJECT MATTER
EXPERIENCER = HIGHEST, SUBJECT MATTER = 2ND-HIGHEST

c. Available roles in *Spiders frighten Lily*
= CAUSER, EXPERIENCER
CAUSER > EXPERIENCER
CAUSER = HIGHEST, EXPERIENCER = 2ND-HIGHEST

d. Available roles in *Jack sneezed during the meeting*
= AGENT
AGENT = HIGHEST

e. Available roles in *The package arrived during the meeting*
= PATIENT
PATIENT = HIGHEST

For primary pattern sentences like (10a), there are three thematic roles: AGENT, PATIENT, and INSTRUMENT. The thematic hierarchy places them in that order (AGENT > PATIENT > INSTRUMENT), so they map to *subject*, *object*, and oblique *object* positions respectively. For psych-verbs like *fear* in (10b), rUTAH would posit that *Lily* is an EXPERIENCER, while *spiders* is a SUBJECT MATTER. As such, *Lily* will map to the *subject* position, and *spiders* will map to the *object* position. In contrast, for psych-verbs like *frighten* in (10c), rUTAH would posit that *spiders* is now a CAUSER, though *Lily* is still an EXPERIENCER. Because CAUSER > EXPERIENCER, *spiders* will map to the *subject* position, and *Lily* will map to the *object* position. Finally, for the intransitive verbs *sneezed* (10d)-(10e), both verbs only have one syntactic position and one thematic role; so, the argument appears in *subject* position regardless of its thematic role.

B Survey of child behavioral results

B.1 Child behavioral data

This is a synthesis of 38 behavioral acquisition studies relating to verb behaviors known by children by five years old. The specific verbs attested are used to identify which particular verbs ought (or ought not) to cluster together at different ages.

B.1.1 Passivizable, intransitive, & monotransitive.

A verb that's passivizable is often one that can be used transitively (i.e., it allows an *object*). For example, *eat* is both passivizable (*It was/got eaten*) and (optionally) transitive (*I ate it*). However, verbs can also be used in the passive form even if they're intransitive, because they can take an *indirect object*. We can see this with *sneeze*: *It was/got sneezed at* and *I sneezed at it*).

In terms of acquisition evidence, we should be able to look at studies that investigate comprehension of transitive verbs and transitive cues, as well as studies that investigate children's comprehension of passives. Moreover, for passives, comprehension of a "short" passive (e.g., *It was/got eaten*) should be sufficient, rather than requiring comprehension of a "long" passive (e.g., *It was/got eaten by the cat*). Also, if children comprehend the long passive, they should be able to comprehend the short passive for that verb.

The ability to comprehend the passive usage of a verb correctly seems to come significantly after the cues to transitivity are recognized. For example, by two years old, English children recognize that the frame *She's Xing the man* indicates X is a transitive verb and so expect a transitive meaning where one agent affects another; they also recognize the frame *She's Xing* indicates X is an intransitive verb, and so expect an intransitive (e.g., unaccusative or unergative) meaning with only a single agent (Naigles, 1990; Naigles & Kako, 1993; Yuan & Fisher, 2009). At 28 months, they also recognize cues involving multiple frames to identify verbs as optionally transitive vs. unaccusative (Scott & Fisher, 2009).

Maratsos 1974. M. P. Maratsos (1974) finds that children can pass an act-out task with full passives using the verbs *bump* and *push* by age four and a half.

Maratsos et al. 1985. M. Maratsos, Fox, Becker, and Chalkley (1985) found that children can comprehend long passives for these verbs by age 4: *find, hold, kick, kiss, push, wash*. They comprehend long passives for these verbs by age 5: *like, love*. They comprehend long passives for these verbs by age 9: *hate, remember, see*.

Gordon & Chafetz 1990. Gordon and Chafetz (1990) found that children can comprehend both short and long passives for these verbs by age 3: *drop, eat, carry, hold, hug, kick, kiss, shake, wash, watch*. However, they fail to comprehend either long or short passives for these by age 3 to 4: *believe, forget, hate, hear, know, like, remember, see*.

O'Brien et al. 2006 & Nguyen et al. 2016. O'Brien, Grolla, and Lillo-Martin (2006) found that children can comprehend long passives for these verbs by age 3 (and 4) when the pragmatic context makes the use of the passive more felicitous: *hug, chase, like, see*. However, Nguyen, Lillo-Martin, and Snyder (2016) found that three- and four-year-olds only seem to comprehend long passives for *hug* and *chase*, though they comprehend short passives for all four verbs.

Hirsch & Wexler 2007. Hirsch and Wexler (2007) found that three-year-old children can comprehend long passives for these actional verbs at greater than chance rates in a two-choice sentence picture-matching task: *hold, kick, kiss, push*. In contrast, only seven-year-old children (and older) comprehend long passives for these psychological verbs at greater than chance rates: *hate, love, remember, see*.

Crain et al. 2009. Crain, Thornton, and Murasugi (2009) report samples of elicited long passives from 9 three-year-olds, 22 four-year-olds, and one five-year-old. We can assume that if these children produce a passive structure, they have classified the verbs as passivizable. Based on the samples from the three-year-olds (3;4-3;11), by age three children should have grouped these verbs together: *bleed, crash, eat, hit, hurt, lick, knock, marry, punch, pick up, push, scratch, trip up, turn (over)*. Based on the samples from the four- and five-year-olds (4;1-5;0), by age four children should have additionally grouped these verbs together: *bite, hug, jump, kill, kick, kiss, knock, lick, ride, shoot*.

Messenger et al. 2009. Messenger, Branigan, McLean, and Sorace (2009) found that children can comprehend long passives for these verbs by age 3-4: *frighten, shock, annoy, upset, surprise, scare, pat, bite, pull, hit, carry, squash*. However, they were unable to comprehend long passives at age 3-4 for these verbs: *see, hear, love, ignore, remember, hate*.

B.1.2 Ditransitive.

Corpus analysis and experimental work discussed below suggests that some verbs that allow (or require) two *objects* are identified by age three (Gropen et al., 1989; Snyder & Stromswold, 1997; Campbell & Tomasello, 2001; Conwell & Demuth, 2007; Thothathiri & Snedeker, 2008), though there may be some overgeneralizations where verbs are assumed to have the ditransitive behavior for awhile that shouldn't (ex: *say*).

Gropen et al. 1989. Gropen et al. (1989) conducted a corpus analysis of five children's productions (including Brown-Eve) and found that the double-object construction was often first produced between ages 1;8 and 3;3. Their corpus analysis showed that at least two of the five children had used the following verbs in both dative constructions, with the earliest age of use in parenthesis: *give (1;9), get (2;0), read (2;0), bring (2;3), buy (2;11), show (3;0), make (3;4), tell (3;4)*. This same analysis showed that at least two of the five children had used the following verbs in the double-object dative construction, with the earliest age of use in parenthesis: *read (1;8), give (1;9), show (1;9), bring (1;10), get (2;0), buy (2;11), pour (2;11), tell (3;0), draw (3;4), make (3;4), teach (3;6), ask (4;7)*.

This suggests that these verbs may have the ditransitive behavior by these ages: two = *read, give, show, bring, get*; three = *buy, pour, tell*; four = *draw, make, teach*; five = *ask*.

In addition, Gropen et al. (1989) also observed overproductive uses of the double-object construction at these ages for these verbs: *write (2;3), say (2;8), eat (3;3), keep (3;8), spend (4;0), put on (4;1), finish (4;11), fix (5;2)*. So, the above clusterings may also include these errors (i.e., including these verbs at these ages). So, the groupings may look more like this for the ditransitive behavior: two = *read, give, show, bring, get*; three = *buy, pour, tell, write, say*; four = *draw, make, teach, eat, keep, spend*; five = *ask, put on, finish*; six = *fix*.

Snedeker & Huang 2015, Campbell & Tomasello 2001. Snedeker and Huang (2015), citing Campbell and Tomasello (2001), note that both constructions associated with the dative (i.e., *sub-*

ject verb $object_{indir}$ $object_{dir}$ [double-object]=*She gave him a penguin* and *subject verb* $object_{dir}$ *preposition* $object_{indir}$ [prepositional]=*She gave a penguin to him*) are acquired before age three. More specifically, Campbell and Tomasello (2001) conducted a corpus analysis of seven children in the CHILDES database and found that both dative constructions are first produced by age three (at the latest), and often before age two and a half (five of seven children). The specific verbs where the majority of children produced multiple dative construction types by age three are *bring, get, give, make, read,* and *show*. So, it is likely children should have grouped these verbs together by age three. Moreover, if we focus on the verbs that were used in the double-object construction specifically, these are grouped together: *bring, buy, get, give, make, read, show, tell*.

Huttenlocher et al. 2004. Huttenlocher et al. (2004) found that four- and five-year-olds had structural priming for both dative constructions across these verbs: *bake, bring, buy, deliver, feed, give, serve, show, teach, throw*. This suggests these verbs have been clustered together by age four into a class that allows the double-object dative and the prepositional dative.

Conwell & Demuth 2007. Conwell and Demuth (2007) find that three-year-olds demonstrate abstract structural knowledge of both the double-object and prepositional dative constructions, using elicited repetition with novel verbs. In particular, three-year-olds will use the prepositional construction for a novel verb when it's been modeled in the double-object construction, suggesting they understand these are related. So, it is likely three-year-olds have a verb class that includes the double-object construction (i.e., the ditransitive behavior).

Thothathiri & Snedeker 2008. Thothathiri and Snedeker (2008) note that naturalistic production isn't definitive about categorization – in particular, children's productions could be simple imitations of their input rather than generalizations formed via a verb class. Thothathiri and Snedeker (2008) conduct priming studies with three- and four-year-olds children to determine whether class-level knowledge exists for dative constructions (double-object and prepositional). This presumably also indicates which verbs belong to that class, as children who can generalize with a construction to a new usage are doing so because the verb belongs to the class.

In the first set of experiments, four-year-olds show both within-verb priming (with *give*) and across-verb priming (*show* priming *give* for the double-object construction, *bring* priming *give* for the prepositional dative). These results suggest that four-year-olds cluster together *show* and *give* for the double-object construction, and potentially also *bring* for datives in general.

In the second set of experiments, three-year-olds also show both within-verb and across-verb priming with these verbs: *bring, hand, pass, send, show, throw, toss*. The priming effect was stronger for the double-object construction, which suggests this class has definitely been formed. So, by three, children have likely clustered together these verbs (along with *give*) into a class that allows the ditransitive construction.

B.1.3 Unaccusatives and unergatives.

Both experimental evidence (Bunger & Lidz, 2004) and analysis of naturalistic productions (Déprez & Pierce, 1993; Snyder & Stromswold, 1997) suggest that English children have begun forming a class of unaccusative verbs by age two. Experimental evidence (Bunger & Lidz, 2008) additionally suggests that English children have begun forming a class of unergative verbs by age two.

Depréz & Pierce 1993. Déprez and Pierce (1993) investigated three English children's naturally produced speech between the ages of one and a half and two years old. They found that post-verbal *subjects* (i.e., VS order, like *going it*) occurred only with *be* and with unaccusatives such as *break, go, come*, and *fall*. This suggests that English children have grouped these unaccusatives together by age two (presumably noting the PATIENT-like role of the *subject*).

Snyder & Stromwold 1997. Snyder and Stromswold (1997) conducted a corpus analysis in the naturalistic productions of 12 English children to determine the age at which the verbs *break, come, fall, go, grow,* and *leave* were first produced in an "unaccusative context". This age range was found to be between 1;6 and 2;7. So, we might interpret this as children recognizing that these six verbs can be used unaccusatively by age two and clustering them together.

Gelman & Koenig 2001. Gelman and Koenig (2001) find that five-year-olds and adults use the animacy of a *subject* to determine how to interpret the verb *move* in intransitive uses such as *Was this X moving?* In particular, +animate *subjects* yield an unergative reading while -animate *subjects* yield an unaccusative reading. So, by five, children seemed to have created a distinction between the unergative and unaccusative classes that depends on the animacy of the subject. The three-year-olds in the study were trending towards this behavior, but their behavioral results weren't statistically significant.

Bunger & Lidz 2004, 2008. Bunger and Lidz (2004) demonstrate experimentally that 2-year-old English children are able to use syntactic distributional cues to determine that a verb is unaccusative. For example, when presented with *"The ball is pimming"* (sometimes accompanied with *"The girl is pimming the ball"*, 2-year-olds infer that *pim* refers to the "results-focused" action (like the ball bouncing) that unaccusatives have. This suggests that children have begun forming an unaccusative verb class by age two, and it has both syntactic and semantic cues associated with it.

Bunger and Lidz (2008) demonstrate experimentally that 2-year-old English children also are able to use both syntactic distributional cues and semantic role information to determine that a verb is unergative. For example, when presented with *"The boy is blicking"*, 2-year-olds infer that *blick* refers to the "means-focused" action (like pumping a bicycle pump to spin an attached flower) that unergatives have. This suggests that children have begun forming an unergative verb class by age two, and it has both syntactic and semantic cues associated with it.

B.1.4 Control object and raising object.

Kirby 2009, 2009, 2010 & Becker 2014. Becker (2014) discusses a set of experiments by Kirby (2009b) in which the raising object verbs *want* and *need* were investigated along with the control object verbs *ask* and *tell*. Kirby (2009b, 2011) found that four- and five-year-olds were able to correctly interpret raising object verb and control objects verb utterances when the embedded clause is active (e.g., "*He wanted Winnie the Pooh to kiss Patrick*", "*He asked the farmer to comb the horse*"). Both ages of children are also able to interpret raising object utterances when the embedded non-finite clause is a passive ("*He wanted Tigger to be called by Elmo.*"). However, only five-year-olds were above chance on interpreting control objects with embedded non-finite passives (*She told the policeman to be sniffed by the dog."*).

Additionally, as reported in Kirby (2009a, 2010), both four- and five-year-olds are sensitive to the animacy restrictions for control object verbs (and reject utterances like "*Elmo told the toys to be smaller*" as "weird" more often than chance). This behavior is different for raising object verbs (with utterances like "*The boy wanted the cake to be chocolate*"), where both ages of children were at chance for accepting the utterances as "okay". Similarly, as also reported in Kirby (2009a, 2010), four- and five-year-olds treat control object and raising object verbs differently when judging the acceptability of expletive subjects in the embedded clause (e.g., control object: **The girl told it to be warm*, raising object: *The girl wanted there to be cookies in the bag.*). Adult-like judgments do vary by age: four-year-olds only have adult judgments for control object verbs and reject expletive subjects while five-year-olds only have adult judgments for raising object verbs and accept expletive subjects. Still, the main point is that they recognize that these types of verbs behave differently from each other but similarly to ones of the same kind (i.e., *want* patterns with *need* and *ask* patterns with *tell*).

This suggests five-year-olds have distinguished raising object from control object verbs, and four-year-olds may have as well. More specifically, we might expect four- and five-year-olds to group together *want* and *need* in one class and *ask* and *tell* in a separate class.

B.1.5 Control subject and raising subject.

Becker 2006. The experiment with children in Becker (2006) suggests that five-year-olds have adult-like judgments for verbs allowing inanimate *subjects*. In particular, they accept inanimate *subjects* only for subject raising verbs like *seem* and *appear* (e.g., *The flower seems to be pink*), and not for control subject verbs like *want* and *try* (e.g. *The flower wants to be pink*). In contrast, three- and four-year-olds allow all verbs to have inanimate *subjects* (and so have not differentiated *want* and *try* from *seem* and *appear* in this respect). For purposes of verb classification, this suggests that children place *want* and *try* together, separately from *seem* and *appear* by five years old.

Hirsch & Wexler 2007. The comprehension experiments in Hirsch and Wexler (2007), which involved children choosing which of two pictures matched a spoken utterance, suggest that children can correctly interpret instances of *seem* that don't involve raising (e.g., *It seems to Homer that Marge is pushing a cart*) as young as three. However, they struggle to comprehend *seem* instances that involve raising (e.g., *Homer seems to Maggie to be bowling a ball.*. Because no other subject-raising verbs were tested, it is unclear if children would classify *seem* as being similar to other subject-raising verbs.

Becker 2007 and Becker 2009. The experiments in Becker (2007, 2009) suggests that three- and four-year-olds accept expletive subjects for *seem* and *appear* (e.g., *It seems to be windy*, *It appears to be warm*) either 83% (three-year-olds) or 91.7% (four-year-olds) of the time. This may indicate that they recognize that *seem* and *appear* are the same kind of predicate. However, these children also accepted expletive subjects for the control subject verbs *want* and *try* either 66.7% (three-year-olds) or 68.1% (four-year-olds) of the time. Becker (2009) suggests this is because children of both ages recognize the expletive subject as a cue for raising subject verbs – and so try to cast control subject verbs as raising subject verbs when they're used with expletive subjects (*It wants to be raining* = modal interpretation ≈ *It's going to rain.*

For the purposes of verb classification, only four-year-olds – but not three-year-olds – accepted raising subject verbs with expletive subjects more often than control subject verbs with expletive subjects. Becker (2009) suggests that three-year-olds are still determining the class membership for raising vs. control subject verbs. In particular, only four-year-olds recognize that *want* and *try* should be in a different class than *seem* and *appear* (the class that naturally allows expletive subjects).

Becker 2014. Becker (2014) discusses a pilot study (pp.206-207) with 5 five-year-olds, and finds that being used with an inanimate *subject* is a strong signal to expect that verb to also be used with the *there*-expletive (*Did there meb to be a banana in the soup?*). So, five-year-olds have a strong sense of cues to raising verb analyses.

B.1.6 Subject-experiencer & object-experiencer verbs.

Experimental studies suggest that children are still in the process of learning how to interpret subject-experiencer verbs like *like, love*, and *hate* in all contexts, even though they use these words frequently in their naturalistic output and hear them often in their input (Hartshorne et al., 2015). Five-year-olds, however, can interpret subject-experiencer verbs correctly above chance. In contrast, four-year-olds are better at sorting out object-experiencer verbs like *surprise* and *frighten* and interpreting them correctly above chance (Hartshorne et al., 2015) when both the *subject* and *object* are animate. This suggests that there may be an object-experiencer verb class by age 4, as well as an emerging subject-experiencer verb class.

Hartshorne et al. 2015. Hartshorne et al. (2015) conducted experimental studies testing children's comprehension of several subject-experiencer verbs (*admire, fear, hate, like, love, trust*) and object-experiencer verbs (*amaze, bore, confuse, frighten, scare, surprise*). They found that four-year-olds correctly interpreted three higher-frequency object-experiencer verbs above chance (*surprise, frighten, scare*) and no subject-experiencer verbs above chance when both *subject* and *object* were animate. In fact, two low frequency subject-experiencer verbs were interpreted as object-experiencer verbs above chance (*fear, trust*). This may suggest that four-year-olds have grouped these five verbs together: *fear, frighten, scare, surprise, trust*. When the *subject* was animate while the *object* was inanimate, four-year-olds are above chance at interpreting the three most frequent subject-experiencer verbs correctly: *like, love, hate*. This suggests that four-year-olds have grouped these verbs together, though they may not always interpret them correctly if the *object* is animate.

B.1.7 Complements: *-ing*, *to*, *that*, *whether/if*.

Bloom et al. 1984, 1989. Bloom, Tackeff, and Lahey (1984) conducted an analysis of child-produced speech from four children between the ages of two and three years old. They found suggestive evidence that children have productive use of non-finite *to* in "V-to-V" constructions, such as *I want to see Mommy* (though not in "V-NP-to-V" constructions like *I want Mommy to get balloon*). This suggests that by three years old, children have begun forming a class of verbs that take non-finite *to* as a complement. The specific verbs that involved complement verb contexts in child productions were these: *ask, forget, get, go, have, know, like, need, show, start, suppose, teach, tell, try, use, want*. A subset of these were used specifically with "how to" instructive contexts: *know, show, teach*. Bloom et al. (1984) note that no *-ing* complements were produced by any of the children they looked at. This suggests acquisition of *-ing* complements occurs after three years of age.

Bloom, Rispoli, Gartner, and Hafitz (1989) examined the naturalistic productions from these same children for evidence of productive use of (1) "S-complements" (including both non-finite *-ing* clauses such as *I see Mommy wash(ing) her hands* and finite *-that* clauses such as *I see that Mommy is washing her hands*, and (2) *wh*-complements, such as *I know what the little bear's eating*. They found that *think* and *see* were produced with S-complements while *know* and *see* were produced with *wh*-complements. However, their evidence did not indicate that these children realized there was a general rule or class of such verbs. Instead, it seemed "children learned this for each matrix verb separately". This suggests that verb classes involving S-complements and *wh*-complements (i.e., interrogative clauses, which would include *whether/if* clauses) aren't formed till after three years of age.

Diessel & Tomasello 2001. Diessel and Tomasello (2001) conducted a corpus analysis of seven English-speaking children's spontaneous speech between the ages of 1;2 and 5;2. They identified seven verbs taking an *if*-complement: *see* (69/98), *tell* (14/98), *wonder, ask, care, know, happen* (15/98). This suggests that by five, children have clustered together *see* and *tell*, and may have also clustered together *ask, care, happen, know,* and *wonder*.

Papafragou et al. 2007. Papafragou et al. (2007) investigated the cues children (and adults) use to identify the meaning of a verb, including the syntactic cue of taking a tensed clausal complement introduced by complementizer *that* (e.g., *Matt gorps that his grandmother is under the covers.*). When eliciting guesses about verb meaning from 34 children ages 3;7-5;9, they found that this syntactic frame increased the chances of children guessing belief verbs. In particular, children's guesses included these verbs: *be surprised, dream, fall for, guess, know, lie, pretend, think, trick*. This suggests that by four to five, children are forming classes of verbs that allow a tensed sentential complement (specifically here, the belief verb class).

Kidd et al. 2006, 2010. Kidd et al. (2006) conducted a sentence recall/lexical priming study on three-, four-, and five-year-olds involving both high and low frequency verbs that take finite complements (e.g., *think: "I think she is riding away on the horse."*)[5]. Interestingly, children often made substitutions with different verbs, instead of repeating the verbs previously presented to them. Three-year-olds substituted in these verbs more than once: *think* (179/266), *bet* (34/266), *hope* (24/266), *see* (11/266), *know* (9/266), *say* (4/266), and *hear* (4/266). Four- and five-year-olds substituted in the same verbs, though with higher incidences for some lower frequency verbs (*know, say*) than the three-year-olds. This suggests that by age three, children may have clustered together these *that*-complement taking verbs: *bet, hear, hope, know, say, see, think*.

Kidd et al. (2010) conducted a similar sentence recall/lexical priming study on four- and six-year-olds involving both high and low frequency verbs that take finite complements (e.g., *say: 'Mickey says that Minnie is wearing a lovely dress."*). Again, children often made substitutions with different verbs, instead of repeating the verbs previously presented to them. Four-year-olds substituted in *think* nearly 88% of the time (126/143), and used *get* (1/143), *see* (3/143), *tell* (3/143), and *wish* (6/143) for the remaining 12%. Six-year-olds substituted *think* nearly 75% of the time (102/135), and used *find* (2/135), *laugh* (1/135), *like* (2/135), *persuade* (1/135), *reckon* (13/135), *see* (5/135), *tell* (1/135), and *wish* (1/135) for the remaining 25%. If we assume that any verb substituted in more than once is reliably affiliated with the class of complementizer-*that* verbs, we have the following clusters: at four, children have clustered together *see, tell, think*, and *wish*; at six, they have additionally clustered in *find, like*, and *reckon*.

B.2 Verb classes derived from behavioral data

This is a synthesis of the 33 behavioral studies describing the behavior of specific verbs at specific ages. We note that the earliest age of acquisition documented was used for the verb behavior associated with a specific verb.

B.2.1 Passivizable.

By age three, children should cluster together these: *bleed, carry, chase, crash, drop, eat, hit, hold, hug, hurt, kick, kiss, knock, lick, like, marry, punch, pick up, push, scratch, see, shake, trip up, turn (over), wash, watch* (Gordon & Chafetz, 1990: short and long passives, O'Brien et al., 2006: pragmatically felicitous long passives, Nguyen et al., 2016: short passives; Hirsch & Wexler, 2007; Crain et al., 2009: long passives). Moreover, they should keep them separate from these: *believe, forget, hate, hear, ignore, know, love, remember* (Gordon & Chafetz, 1990: short and long passives; Messenger et al., 2009: long passives).

---

[5] Note that the primes didn't involve complementizer *that* even though the verbs could all allow it. That is, all the verbs belonged to the class that optionally allows complementizer *that*.

By age four, children should additionally cluster in these: *annoy, bite, bump, find, frighten, jump, kill, pat, pull, ride, scare, shock, shoot, squash, surprise, upset* (M. P. Maratsos, 1974: long passives, M. Maratsos et al., 1985: long passives, Messenger et al., 2009: long passives, Crain et al., 2009: long passives).

By age five, children should additionally include these: *love* (M. Maratsos et al., 1985: long passives).

By age seven, children should additionally include these: *hate, remember* (Hirsch & Wexler, 2007: long passives).

B.2.2 Unaccusative.

By two, children have recognized that these verbs can be used unaccusatively and so should be clustered together: *break, come, fall, go, grow, leave* (Déprez & Pierce, 1993; Snyder & Stromswold, 1997: naturalistic production).

B.2.3 Control object and raising object.

The study by Kirby (2009b) suggests that children should group *want* and *need* together (raising object), and cluster together *ask* and *tell* (control object) separately by age four.

B.2.4 Control subject and raising subject.

Studies by Becker suggest that children should group *want* and *try* (control subject) together, and separate them from *seem* and *appear* (raising subject), by either four years old (Becker, 2007, 2009) or five years old (Becker, 2006).

B.2.5 Subject-experiencer and object-experiencer.

By age four, children have grouped these high-frequency subject-experiencer verbs together: *like, love, hate* (Hartshorne et al., 2015: truth-value judgment task). They have also grouped together several object-experiencer verbs (*frighten, scare, surprise*) with lower-frequency subject-experiencer verbs (*fear, trust*) (Hartshorne et al., 2015: truth-value judgment task).

B.2.6 Complements: *to*, *that*, *whether*.

***to*-complement.** By age three, children have likely clustered together these verbs that can take a non-finite *to* complement: *ask, forget, get, go, have, know, like, need, show, start, suppose, teach, tell, try, use, want*. A subset of these may be in a separable “how to” instructive context class: *know, show, teach* (Bloom et al., 1984: corpus analysis of naturalistic productions).

***that*-complement.** By age three, children have likely clustered together these verbs that optionally take a *that*-complement: *bet, hear, hope, know, say, see, think* (Kidd et al., 2006: sentence recall/lexical priming substitution). By age four, children have likely clustered together these verbs as well: *tell, wish* (Kidd et al., 2010: sentence recall/lexical priming substitution). By age five, children have likely grouped in these belief verbs together that take a sentential complement introduced by complementizer *that*: *be surprised, dream, fall for, guess, lie, pretend, trick* (Papafragou et al., 2007: elicited verbs). By age six, they have additionally clustered in *find, like*, and *reckon* (Kidd et al., 2010: sentence recall/lexical priming substitution).

***whether/if*-complement.** By age five, children have clustered together these verbs that allow *if*-complements: *see* and *tell*. They may have also clustered together *ask, care, happen, know,* and *wonder* (Diessel & Tomasello, 2001: spontaneous speech productions).

C Verb classes derived from child behavioral data

Table 7: Attested verb behaviors and verb classes derived from child behavioral data for verbs in the <3yrs dataset appearing 5 or more times.

Verb behavior	**Attested verbs**
Unaccusative	**6**: break, come, fall, go, grow, leave
Ditransitive	**14**: bring, buy, get, give, make, pass, pour, read, say, send, show, tell, throw, write
Non-finite *to* complement	**16**: ask, forget, get, go, have, know, like, need, show, start, suppose, teach, tell, try, use, want
Passivizable	**27**:
	+: carry, chase, crash, drop, eat, hit, hold, hurt, jump, kick, kiss, knock, lick, like, punch, push, scratch, see, shake, turn, wash, watch
	–: believe, forget, hear, know, remember
***that*-complement**	**9**: bet, hear, hope, know, say, see, tell, think, wish
Verb classes	**15 classes, 60 verbs**
[+unaccusative]	**5**: break, come, fall, grow, leave
[+ditrans]	**10**: bring, buy, give, make, pass, pour, read, send, throw, write
[+non-finite *to*]	**9**: ask, have, need, start, suppose, teach, try, use, want
[+passive]	**20**: carry, chase, crash, drop, eat, hit, hold, hurt, jump, kick, kiss, knock, lick, punch, push, scratch, shake, turn, wash, watch
[-passive]	**2**: believe, remember
[+*that*-comp]	**4**: bet, hope, think, wish
[+unaccusative, +non-finite *to*]	**1**: go
[+ditrans, +non-finite *to*]	**2**: get, show
[+ditrans, +*that*-comp]	**1**: say
[+ditrans, +non-finite *to*]	**1**: tell
[+passive, +non-finite *to*]	**1**: like
[+passive, +*that*-comp]	**1**: see
[-passive, +non-finite *to*]	**1**: forget
[-passive, +*that*-comp]	**1**: hear
[-passive, +*that*-comp, +non-finite *to*]	**1**: know

Table 8: Attested verb behaviors and verb classes derived from child behavioral data for verbs in the <4yrs dataset appearing 5 or more times.

Verb behavior	**Attested verbs**
Unaccusative	**6**: break, come, fall, go, grow, leave
Ditransitive	**21**: bake, bring, buy, draw, eat, feed, get, give, keep, make, pass, pour, read, say, send serve, show, teach, tell, throw, write
Non-finite *to* complement	**16**: ask, forget, get, go, have, know, like, need, show, start, suppose, teach, tell, try, use, want
Passivizable	**38**: +: bite, bump, carry, chase, crash, drop, eat, find, frighten, hit, hold, hurt, jump, kick, kill, kiss, knock, lick, like, pull, punch, push, ride, scare, scratch, see, shake, shoot, surprise, turn, wash, watch –: believe, forget, hear, know, love, remember
***that*-complement**	**9**: bet, hear, hope, know, say, see, tell, think, wish
Control object	**2**: ask, tell
Control subject	**2**: try, want
Psych: Obj Exp	**3**: frighten, scare, surprise
Psych: Subj Exp	**2**: frighten, like, love
Raising object	**2**: need, want
Raising subject	**1**: seem
Verb classes	**23 classes, 76 verbs**
[+unaccusative]	**5**: break, come, fall, grow, leave
[+ditrans]	**15**: bake, bring, buy, draw, feed, give, keep, make, pass, pour, read, send, serve, throw, write
[+non-finite *to*]	**4**: have, start, suppose, use
[+passive]	**26**: bite, bump, carry, chase, crash, drop, find, hit, hold, hurt, jump, kick, kill, kiss, knock, lick, pull, punch, push, ride, scratch, shake, shoot, turn, wash, watch
[-passive]	**2**: believe, remember
[+*that*-comp]	**4**: bet, hope, think, wish
[+raising-subj]	**1**: seem
[+unaccusative, +non-finite *to*]	**1**: go
[+ditrans, +non-finite *to*]	**3**: get, show, teach
[+ditrans, +passive]	**1**: eat
[+ditrans, +*that*-comp]	**1**: say
[+ditrans, +*that*-comp, +non-finite *to*, +control-obj]	**1**: tell
[+non-finite *to*, +control-obj]	**1**: ask
[+non-finite *to*, +control-subj]	**1**: try
[+non-finite *to*, +raising-obj]	**1**: need
[+non-finite *to*, +raising-obj, +control-subj]	**1**: want
[+passive, +non-finite *to*, +psych-subj]	**1**: like
[+passive, +*that*-comp]	**1**: see
[+passive, +psych-obj]	**2**: scare, surprise
[-passive, +non-finite *to*]	**1**: forget
[-passive, +*that*-comp]	**1**: hear
[-passive, +*that*-comp, +non-finite *to*]	**1**: know
[-passive, +psych-subj]	**1**: love

Table 9: Attested verb behaviors and verb classes derived from child behavioral data for verbs in the <5yrs dataset appearing 5 or more times.

Verb behavior	**Attested verbs**
Unaccusative	**6**: break, come, fall, go, grow, leave
Ditransitive	**23**: ask, bake, bring, buy, draw, eat, feed, finish, get, give, keep, make, pass, pour, read, say, send, serve, show, teach, tell, throw, write
Non-finite *to* complement	**16**: ask, forget, get, go, have, know, like, need, show, start, suppose, teach, tell, try, use, want
Passivizable	**38**: +: bite, bump, carry, chase, crash, drop, eat, find, frighten, hit, hold, hurt, jump, kick, kill, kiss, knock, lick, like, love, pull, punch, push, ride, scare, scratch, see, shake, shoot, surprise, turn, wash, watch –: believe, forget, hear, know, remember
***that*-complement**	**13**: bet, dream, guess, hear, hope, know, lie, pretend, say, see, tell, think, wish
Control object	**2**: ask, tell
Control subject	**2**: try, want
Psych: Obj Exp	**3**: frighten, scare, surprise
Psych: Subj Exp	**2**: frighten, like, love
Raising object	**2**: need, want
Raising subject	**1**: seem
***whether/if*-complement**	**7**: ask, care, happen, know, see, tell, wonder
Verb classes	**25 classes, 84 verbs**
[+unaccusative]	**5**: break, come, fall, grow, leave
[+ditrans]	**16**: bake, bring, buy, draw, feed, finish, give, keep, make, pass, pour, read, send, serve, throw, write
[+non-finite *to*]	**4**: have, start, suppose, use
[+passive]	**25**: bite, bump, carry, chase, crash, drop, find, hit, hold, hurt, jump, kick, kill, kiss, knock, lick pull, push, ride, scratch, shake, shoot, turn, wash, watch
[-passive]	**2**: believe, remember
[+*that*-comp]	**8**: bet, dream, guess, hope, lie, pretend, think, wish
[+raising-subj]	**1**: seem
[+*whether/if*-comp]	**3**: care, happen, wonder
[+unaccusative, +non-finite *to*]	**1**: go
[+ditrans, +non-finite *to*]	**3**: get, show, teach
[+ditrans, +passive]	**1**: eat
[+ditrans, +*that*-comp]	**1**: say
[+ditrans, +*that*-comp, +non-finite *to*, +control-obj, +*whether-if*-comp]	**1**: tell
[+ditrans, +non-finite *to*, +control-obj, +*whether-if*-comp]	**1**: ask
[+non-finite *to*, +control-subj]	**1**: try
[+non-finite *to*, +raising-obj]	**1**: need
[+non-finite *to*, +raising-obj, +control-subj]	**1**: want
[+passive, +non-finite *to*, +psych-subj]	**1**: like
[+passive, +*that*-comp, +*whether-if*-comp]	**1**: see
[+passive, +psych-obj]	**2**: scare, surprise
[+passive, +psych-obj]	**1**: love
[-passive, +non-finite *to*]	**1**: forget
[-passive, +non-finite *to*, +*that*-comp, *whether/if*-comp]	**1**: know
[-passive, +*that*-comp]	**1**: hear
[-passive, +psych-subj]	**1**: love

D ARI as a clustering evaluation metric

Another common method for evaluating RI scores relative to chance is to convert RI scores into a new measure known as the adjusted Rand Index (**ARI**) (Hubert & Arabie, 1985). The ARI uses a specific randomness model (the generalized hypergeometric distribution) to calculate an expected value for the specific number and size of classes returned by the modeled learner. The RI is then scaled relative to this expected value such that the expected value is 0, an RI of 1 is still 1, and an RI of 0 becomes -1. The result is the ARI, which ranges between -1 and 1 ($-1 \leq \text{ARI} \leq 1$), with scores less than 0 indicating worse than chance performance relative to the randomness model, 0 indicating chance performance relative to the randomness model, and scores greater than 0 indicating better than chance performance. Note that 1 is still perfect performance, as in the original RI. We implemented the ARI calculation using the python function *sklearn.metrics.adjusted_rand_score*.

We note that the ARI does not specify how far above 0 counts as a significant departure from the value expected by chance; it also assumes a specific model of randomness that may or may not hold for a given empirical domain.

Here we briefly summarize the modeling results when ARI is used as the evaluation metric. Because the ARI builds in a randomness model, such that an ARI of 0 represents chance performance, we didn't perform any statistical tests on the ARI results. The main finding is that *all* modeled learners at all ages have an ARI above 0, suggesting that all modeled learners are identifying verb classes that match children's verb classes more often than chance. Thus, in this sense, the ARI aligns with all the RI results – all learners performing above chance according to the RI are performing above chance according to the ARI. What differs is how much above 0 an individual learner's ARI is. Below we compare the relative magnitude of the ARI results across learners.

For the three-year-old modeled learners (<3yrs), all four that ignore surface morphology (-morphology) perform better than the four that heed surface morphology (-morphology: .193-.238, +morphology: .053-.144). This accords well with the RI results, where both learners who performed better than chance ignored surface morphology. For the four-year-old modeled learners (<4yrs), the best performing one is one of the ones above chance for the RI results: heeding surface morphology, using the absolute thematic system, and not expecting a mapping (ARI: 0.261). We note that its ARI score is far better than any of the other modeled learners at this age: ARI = .080-.143). For the five-year-old modeled learners (<5yrs), the two that appear to substantially outperform the others heed surface morphology (+morphology), use either the absolute or relative thematic system, and expect a mapping (+expect-mapping) (absolute: .256, relative: .279, all other 6 modeled learners: .087-.149). These learners are among those that are above chance according to the RI results, again supporting the RI results.

E Gibbs sampling in the generative model

The plate diagram for the generative model is presented here again for ease of reference, and Table 10 describes the variables in the plate diagram. Gibbs sampling is done following the process laid out in the plate diagram and described in more detail in the rest of this appendix. Depending on the modeled learner, between 2000 and 4000 iterations of Gibbs sampling were run, with each iteration involving the sampling of all classes, class properties, and hyperparameters, as described below. The number of iterations was determined by doing several sample runs across a range of iteration counts for each intake set, and seeing which one performed best according to the learner's log probability of the intake data. The iteration counts used for the results reported in the main text are as follows:

(i) <3yrs:
+morphology: UTAH (3000), rUTAH (3000), absolute-no-mapping (2000), relative-no-mapping (2000)
-morphology: UTAH (3000), rUTAH (3000), absolute-no-mapping (3000), relative-no-mapping (4000)
(ii) <4yrs:
+morphology: UTAH (2000), rUTAH (2000), absolute-no-mapping (3000), relative-no-mapping (2000)
-morphology: UTAH (2000), rUTAH (4000), absolute-no-mapping (2000), relative-no-mapping (2000)
(iii) <5yrs:
+morphology: UTAH (3000), rUTAH (2000), absolute-no-mapping (2000), relative-no-mapping (3000)
-morphology: UTAH (3000), rUTAH (2000), absolute-no-mapping (4000), relative-no-mapping (2000)

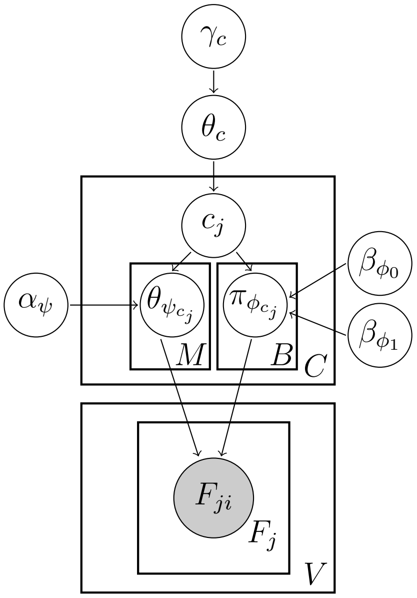


Figure 3: Plate diagram for the generative model of verb classes.

E.1 How Gibbs sampling works

The general form of the sampler for a situation where outcome x_i takes value k, given previous outcomes x_{-i} = (x_1, ...x_{i-1}), and hyperparameter/pseudocount β for all K possible outcomes and multinomial distribution θ is:

$$P(k|x_{-i},\beta) = \int P(k|\theta)P(\theta|x_{-i},\beta)d\theta = \frac{n_k+\beta}{i-1+K\beta} \tag{1}$$

Variable	Interpretation
V	total number of verbs
F_j	total frames for verb *j*
F_{ji}	frequency of frame *i* for verb *j*
C	number of classes
c_j	class of verb *j*
θ_c	distribution over class labels
γ_c	hyperparameter (pseudocounts) for class labels
B	number of binary (binomial distribution) properties being tracked
	ex: +anim-subj, +anim-obj, +anim-iobj, +mvmt = 4
$\pi_{\phi_{c_j}}$	distribution over binary property ϕ for class c_j
	ex: +anim-subj, +anim-obj, +anim-iobj, +mvmt
β_{ϕ_1}, β_{ϕ_0}	hyperparameters for Beta distribution prior for property ϕ
M	number of multinomial properties tracked
	ex: syntactic frames, thematic-subj distr, thematic-obj distr, thematic-iobj distr = 4
$\theta_{\psi_{c_j}}$	distribution over multinomial property ψ for class c_j
	ex: syntactic frames, thematic-subj distr, thematic-obj distr, thematic-iobj distr
α_ψ	hyperparameter for Dirichlet distribution prior for property ψ

Table 10: Variables used in the plate diagram for the generative model and in the details of the Gibbs sampling process.

according to the derivation in Goldwater and Griffiths (2007), who refer to the derivation in MacKay and Peto (1995), where n_k is the number of times k occurred in x_{-i}.

This is the basis that we can use to do the sampling for all categorical and binary variables that correspond to a single frame instance (ex: verb classes, +animate-subject, etc). The basic sampling process involves removing the information of the current item being considered, and then calculating the sampling probabilities as directed. Once sampling is complete and the item's information is updated, we add the new information back in. See Resnik and Hardisty (2010) for an excellent tutorial about Gibbs sampling.

E.2 Probability of the category label for a given verb

For all currently existing categories c_j in C and the possibility γ_c of creating a new category, calculate

$$p_{cat_j} = P(c_j|c_{-j}, \gamma_c) = \frac{n_{c_j} + \gamma_c}{n_{all} + C\gamma_c} \tag{2}$$

where n_{c_j} is the number of verbs with category c_j (excluding the verb whose class is currently being sampled (c_{-j}), n_{all} is the total number of verbs excluding the one being sampled, and C is the total number of classes currently. So, the equivalent for the possibility of creating a new category would be

$$P(c_{new}|c_{-j},\gamma_c) = \frac{\gamma_c}{n_{all} + C\gamma_c} \tag{3}$$

These should then all sum to 1.

Note: When generating a random assignment of verbs to classes, use the above process to generate a class for each verb. That is, for each verb, generate a class (either an existing one or a new one) according to the probability above. In the results reported in the main text for the randomization tests, we use the default value γ_c=1 for this random generation process, which is the value all modeled learners were initialized with.

E.3 Sampling binary properties

These will include +anim-subj, +anim-obj, and +anim-iobj for all model types. For the model assuming the mapping is already known, +mvmt would indicate whether the observed syntactic positions obey r/UTAH transparently or have moved. All of these apply to the probability of an individual frame f_{ji} appearing with the property within a given verb class c_j for a verb v_j, with $F_{c_{-j}}$ referring to all the frames in class c_j except for the ones from v_j. These properties are represented by π_{ϕ_c} in the plate diagram.

$$p_{\phi_{i,c_j}} = P(f_{ji}|c_j, c_{-j}, F_{c_{-j}}, \pi_{\phi_{c_j}}, \beta_{\phi_1}, \beta_{\phi_0}) = \frac{n_{\phi_{c_j}} + \beta_{\phi_1}}{n_{all_{c_j}} + \beta_{\phi_1} + \beta_{\phi_0}} \tag{4}$$

where $n_{\phi_{c_j}}$ is the number of frames in category c_j that have the property, $n_{all_{c_j}}$ is all the frames in c_j, β_{ϕ_1} is the pseudocount for frames exhibiting the property, and β_{ϕ_0} is the pseudocount for frames not exhibiting the property.

The probability for the entire set of frames F_j is the product of each individual frame token's probability. We can use each individual frame's frequency F_{ji} to calculate this, assuming π_ϕ captures the complete distribution (both ϕ_1 and ϕ_0 probabilities) for property ϕ:

$$p_{\phi_{c_j}} = \prod_i p_{\phi_{i,c_j}}^{F_{ji}} \tag{5}$$

for all i frame types in verb v_j. In particular, if the frame f_{ji} has the property and appears F_{ji} times, its contribution is $p_{\phi_{i,c_j}}^{F_{ji}}$; if frame f_{ji} doesn't have the property and appears F_{ji} times, its contribution is $(1 - p_{\phi_{i,c_j}})^{F_{ji}}$.

The joint probability of all B binary properties is the product of p_{ϕ_c} for all binary properties ϕ:

$$p_{binary_{c_j}} = \prod_{\phi=1}^{B} p_{\phi_{c_j}} \tag{6}$$

E.4 Sampling multinomial properties

This works similarly to the sampling of binary properties – it's just that there are more than two options available. This would apply for the -expect-mapping modeled learner that infers the thematic-syntactic mappings for each verb class. Each underlying intermediate representation's thematic role (e.g.., absolute: AGENT-like or thematic: HIGHEST) would map with some probability to a syntactic position (*subject*, *object*, *oblique (indirect) object*). So, there would be three probabilities for semantic arguments: $p_{subj-pos}$, $p_{obj-pos}$, and $p_{iobj-pos}$. Each of these has three options of where to appear: syntactic *subject*, syntactic *object*, or syntactic *oblique object*. The probabilities of appearing in those positions are a multinomial distribution.

Syntactic frames are another example that applies to all modeled learners, as a frame can be one of the F frame types available.

This property would be captured by the θ_{ψ_c} in the plate diagram.

$$p_{\psi_{i,c_j}} = P(f_{ji}|c_j, c_{-j}, F_{c_{-j}}, \theta_{\psi_{c_j}}, \alpha_\psi) = \frac{n_{\psi_{c_j}} + \alpha_\psi}{n_{all_{c_j}} + O * \alpha_\psi} \tag{7}$$

where f_{ji} is the ith frame token in the total frames F_j for verb v_j, c_j is the category of v_j, $\theta_{\psi_{c_j}}$ is the distribution over the O options, α_ψ is the pseudocount for all options, $n_{\psi_{c_j}}$ is the number of frames in class c_j that have this value, $n_{all_{c_j}}$ is all the frames in class c_j that exhibit any of the values, and O is the number of options of available (e.g., number of syntactic frame types or number of syntactic positions.).

The probability for the entire set of frames F_j is the product of each individual frame token's probability. We can use each individual frame's frequency F_{ji} to calculate this, assuming θ_ψ captures the complete distribution for property ϕ:

$$p_{\psi_{c_j}} = \prod_i p_{\psi_{i,c_j}}^{F_{ji}} \tag{8}$$

for all i frames in verb v_j.

The joint probability of all M multinomial properties is the product of p_{ψ_c} for all multinomial properties ψ:

$$p_{multinomial_{c_j}} = \prod_{\psi=1}^{M} p_{\psi_{c_j}} \tag{9}$$

E.5 Complete sampling equation for new verb class for verb

Let λ be the set of hyperparameters needed for this calculation, including the hyperparameters for category selection, binary properties, and multinomial properties. The complete equation for selecting a new verb class is:

$$p_{c_j} = P(c_j|c_{-j}, \gamma_c, F_{-j}, \lambda) =$$
$$p_{cat_j} * p_{binary_{c_j}} * p_{multinomial_{c_j}}$$

Adjust these probabilities with an annealing temperature if desired (see section E.7 on annealing below), and then roll a C+1 sided die weighted according to the calculated probabilities. Then, select the category label that comes up.

E.6 Hyperparameter sampling

We follow the approach in Goldwater and Griffiths (2007) to sample hyperparameters. In particular, there are priors over each of the hyperparameters (we can assume an improper uniform prior), and use a single Metropolis-Hastings update (Gilks, Richardson, & Spiegelhalter, 1996) to resample the value of each hyperparameter after each iteration of the Gibbs sampler through all individual verbs.

To update the value of hyper parameter κ, we can sample a proposed new value κ' from a normal distribution with $\mu = \kappa$ and $\sigma = .1\kappa$. Then, we calculate the probability of the data (all verbs), given κ vs. given κ'.

E.6.1 Verb category hyperparameter γ_c

Let V be the the total number of observed verb types with their collection of observed frames, C be the current verb classes, and $\lambda_{-\gamma}$ be the set of hyperparameters except for γ_c. The current hyperparameter value is γ_c while the proposed one would be γ_c'. However, for purposes of the calculation (which we'll have to do for both), we can represent both of them below as γ.

$$\begin{aligned} p(C|V, \lambda_{-\gamma}, \gamma) &= \frac{p(V|C, \lambda_{-\gamma}, \gamma) * p(C|\gamma)}{P(V, \lambda_{-\gamma}, \gamma)} \\ &\propto p(V|C, \lambda_{-\gamma}, \gamma) * p(C|\gamma) \end{aligned}$$

We can disregard the denominator since we're just comparing these two values for γ_c and γ_c', and they'll have the same denominator.

Let's look at each term in turn.

$$p(V|C, \lambda_{-\gamma}, \gamma) = \prod_{j=1}^{V} p_{binary_{c_j}} * p_{multinomial_{c_j}}$$

Note that none of these terms depend on γ – so these will be the same for both γ calculations. (Effectively, this term is a constant, and all the action happens in $p(C|\gamma)$.)

$$\begin{aligned} p(C|\gamma) &= \prod_{i=1}^{C} p_{cat_i} \\ &= \prod_{i=1}^{C} \frac{n_{c_i} + \gamma}{V + C * \gamma} \end{aligned}$$

where n_{c_i} is the number of verbs in class i and V is the total number of verbs.

E.6.2 Binomial property hyperparameters β_{ϕ_1}, β_{ϕ_0}

Let V be the the total number of observed verb types with their collection of observed frames, C be the current verb classes, and $\lambda_{-\beta}$ be the set of hyperparameters except for the one β value being sampled. The current hyperparameter value is β_{ϕ_x} while the proposed one would be β'_{ϕ_x}. However, for purposes of the calculation (which we'll have to do for both), we can represent both of them below as β.

$$\begin{aligned} p(C|V, \lambda_{-\beta}, \beta) &= \frac{p(V|C, \lambda_{-\beta}, \beta) * p(C|\gamma)}{P(V, \lambda_{-\beta}, \beta)} \\ &\propto p(V|C, \lambda_{-\beta}, \beta) * p(C|\gamma) \end{aligned}$$

(We can disregard the denominator since we're just comparing these two values for β_{ϕ_x} and β'_{ϕ_x}, and they'll have the same denominator.)

Let's look at each term in turn.

$$p(V|C, \lambda_{-\beta}, \beta) = \prod_{j=1}^{V} p_{binary_{c_j}} * p_{multinomial_{c_j}}$$

Since $p_{multinomial_{c_j}}$ doesn't depend on β, it's a constant for both β calculations. For $p_{binary_{c_j}}$, only the calculation for binary property ϕ will be affected by this β calculation – all the other binary properties will have different β_ϕs that will remain constant for this β calculation. So, the calculation for $p_{\phi_{i,c_j}}$ is the one we pay attention to, calculating that for each verb in V and taking the product:

$$\begin{aligned} p(V|C, \lambda_{-\beta}, \beta) &= \prod_{j=1}^{V} p_{\phi_{i,c_j}} \\ &= \prod_{j=1}^{V} \prod_{i=1}^{F_j} p_{\phi_{i,c_j}}^{F_{ji}} \end{aligned}$$

where F_j is the number of frame types in verb v_j, F_{ji} is the frequency of the frame type i displaying a value of the desired property in class c_j, and p_{ϕ_i,c_j} is the distribution over the two options (so that the appropriate probability is used depending on which value the frame displays):

$$p_{\phi_i,c_j} = \frac{n_{\phi_{c_j}} + \beta}{F + \beta + \beta_{other}}$$

where $n_{\phi_{c_j}}$ is the number of frames in class c_j that have the appropriate property value for the β being calculated, F is the total number of frames in class c_j, and β_{other} is the β not being sampled.

$$p(C|\gamma) = \prod_{i=1}^{C} p_{cat_i}$$

Note that p_{cat_i} doesn't depend on β – so these will be the same for both β calculations. (Effectively, this term is a constant, and all the action happens above in $p(V|C, \lambda_{-\beta}, \beta)$.)

E.6.3 Multinomial property hyperparameters α_ψ

This process is going to be very similar to the binary hyperparameter sampling above.

Let V be the the total number of observed verb types with their collection of observed frames, C be the current verb classes, and $\lambda_{-\alpha}$ be the set of hyperparameters except for the α value being sampled. The current hyperparameter value is α_ψ while the proposed one would be α'_ψ. However, for purposes of the calculation (which we'll have to do for both), we can represent both of them below as α.

$$\begin{aligned} p(C|V, \lambda_{-\alpha}, \alpha) &= \frac{p(V|C, \lambda_{-\alpha}, \alpha) * p(C|\gamma)}{P(V, \lambda_{-\alpha}, \alpha)} \\ &\propto p(V|C, \lambda_{-\alpha}, \alpha) * p(C|\gamma) \end{aligned}$$

(We can disregard the denominator since we're just comparing these two values for α_ψ and α'_ψ, and they'll have the same denominator.)

Let's look at each term in turn.

$$p(V|C, \lambda_{-\alpha}, \alpha) = \prod_{j=1}^{V} p_{binary_{c_j}} * p_{multinomial_{c_j}}$$

Since $p_{binary_{c_j}}$ doesn't depend on α, it's a constant for both α calculations. For $p_{multinomial_{c_j}}$, only the calculation for multinomial property ψ will be affected by this α calculation – all the other multinomial properties will have different α_ψs that will remain constant for this α calculation. So, the calculation for $p_{\psi_{i,c_j}}$ is the one we pay attention to, calculating that for each verb in V and taking the product:

$$\begin{aligned} p(V|C, \lambda_{-\alpha}, \alpha) &= \prod_{j=1}^{V} p_{\psi_{i,c_j}} \\ &= \prod_{j=1}^{V} \prod_{i=1}^{F_j} p_{\psi_{i,c_j}}^{F_{ji}} \end{aligned}$$

where F_j is the number of frame types in verb v_j, F_{ji} is the frequency of the frame type i displaying a value of the desired property in class c_j, and $p_{\psi_{i,c_j}}$ is the distribution over the various options (so that the appropriate probability is used depending on which value the frame displays):

$$p_{\psi_i,c_j} = \frac{n_{\psi_{c_j}} + \alpha}{F + O * \alpha}$$

where $n_{\psi_{c_j}}$ is the number of frames in class c_j that have the appropriate property value for the α being calculated, F is the total number of frames in class c_j, and O is the number of multinomial options there are for this property.

$$p(C|\gamma) = \prod_{i=1}^{C} p_{cat_i}$$

As with the binary properties, p_{cat_i} doesn't depend on α – so these will be the same for both α calculations. (Effectively, this term is a constant, and all the action happens above in $p(V|C, \lambda_{-\alpha}, \alpha)$.)

E.6.4 Metropolis-Hastings update

We can now do the Metropolis-Hastings update: the probability of accepting the new value depends on the ratio between p(C | V, $\lambda_{-\kappa}$, κ) and p(C | V, $\lambda_{-\kappa}$, κ'), with a term correcting for the asymmetric proposal distribution.

1. Calculate a_1:

$$a_1 = \frac{p(C|V, \lambda_{-\kappa}, \kappa)}{p(C|V, \lambda_{-\kappa}, \kappa')} \tag{10}$$

2. Calculate a_2:

$$a_2 = \frac{p(\kappa|\kappa')}{p(\kappa'|\kappa)} \tag{11}$$

   where $p(\kappa|\kappa')$ is the probability of drawing κ from a normal distribution with μ = κ' and σ = $.1\kappa'$, while $p(\kappa'|\kappa)$ is the probability of drawing κ' from a normal distribution with μ = κ and σ = $.1\kappa$.

3. Calculate a = $a_1 a_2$.

   If $a \geq 1$, accept κ'.

   Otherwise, flip a weighted coin (and we can anneal with temperature T if desired). With probability $\frac{a^T}{a^T+(1-a)^T}$, choose κ'. With probability $\frac{(1-a)^T}{a^T+(1-a)^T}$, keep the original κ. If no annealing (or T=1 for that iteration), this defaults to probability a and $1 - a$ for the two options.

E.7 Annealing

To help the Gibbs sampler converge faster, a simulated annealing regime is typically used (e.g., in Goldwater & Griffiths 2007) to force more exploration early on by flattening the probabilities and less exploration later on by sharpening the probabilities. The way this is done is by raising the calculated probability to a power (which is the $temperature\ T$).

$$annealed\, probability = probability^T \tag{12}$$

As the Gibbs sampler does more iterations, the temperature T lowers so the calculated probabilities are sharpened. (Basically, the sampler is more confident about its calculated probabilities later on in learning.) We follow Goldwater and Griffiths (2007) and use a range of T = 2 lowered down to 0.8 over the course of all iterations.

F Filtered verb classes

Tables 11-16 show the filtered verb classifications for each strategy implemented by a computational-level Bayesian learner. Each modeled learner ran 10 times for each dataset (<3yrs, <4yrs, <5yrs), and these ten verb clusterings were aggregated into an aggregate verb clustering for each learner. Any verb pair together in more than $\frac{3}{4}$ of the learner runs (>7 out of 10) was put together in the aggregate verb clustering. Similarly, any verb that was in a class of its own (a *singleton*) for more than $\frac{3}{4}$ of the learner runs was put as a singleton in the aggregate verb clustering.

Table 11: <3yrs dataset: Aggregate inferred classes over 10 runs, given 4 strategies that involve the use of surface morphology (+surface-morphology), an intermediate representation (absolute/relative), and an expectation of a mapping between the intermediate representation and observable syntactic positions (+/-expect-mapping).

+surface-morphology			
absolute		**relative**	
+expect-mapping	**-expect-mapping**	**+expect-mapping**	**-expect-mapping**
singletons: begin, confuse, dress, figure, fill, fold, give, happen, knock, lean, mix, name, pick, plug, rain, roll, seem, send, teach, tell, tip, wait, wake, wonder	*singletons: begin, confuse, dress, figure, fill, hurry, rain, seem, tip, wait, wonder*	*singletons: begin, confuse, dress, figure, fill, fold, knock, lean, mix, name, pick, plug, rain, seem, send, teach, tell, tip, wait, wake, wonder*	*singletons: begin, confuse, dress, figure, fill, knock, plug, rain, seem, tip, wait, wonder*
1: answer, beat, bet, bite, blow, bother, break, bring, brush, build, burn, buy, carry, catch, change, chase, check, chew, chop, clean, close, color, count, cover, crack, cut, decide, do, draw, drink, dry, dump, eat, feel, find, finish, fix, forget, guess, have, hear, help, hit, hold, hook, hope, hurt, keep, kick, kiss, know, leave, lose, love, make, match, mean, meet, melt, mind, miss, move, mow, need, open, paint, pass, pinch, poke, pour, pretend, pull, push, put, reach, read, recognize, record, remember, save, say, scare, scratch, see, shake, share, shave, shoot, shut, sing, smack, snap, spell, spill, spray, squeeze, squish, start, stir, stop, study, swing, take, tape, tear, think, throw, tickle, tie, touch, turn, understand, untie, use, wash, watch, wear, wipe, wish	1: answer, beat, bite, blow, bother, bring, brush, build, bump, buy, catch, change, chase, check, clean, close, color, cool, count, crack, cut, decide, do, draw, drink, drop, dry, eat, feel, find, finish, fix, forget, get, have, hear, help, hold, hurt, keep, kick, know, leave, lose, love, make, mean, mind, miss, move, open, paint, pinch, pour, pull, push, reach, read, recognize, record, remember, rock, roll, save, say, scare, see, shake, share, shave, shut, sing, snap, spell, spill, spray, squeeze, stick, stir, stop, study, suppose, swing, tape, tease, test, tickle, touch, understand, use, watch, wear, wipe	1: answer, beat, bet, bite, blow, bother, break, bring, brush, build, burn, buy, carry, catch, change, chase, chop, clean, close, color, count, cover, crack, cut, decide, do, draw, drink, dry, dump, eat, find, finish, fix, forget, guess, have, hear, help, hit, hold, hook, hope, hurt, keep, kick, kiss, know, leave, lose, love, make, match, mean, meet, mind, miss, move, mow, need, open, paint, pass, pinch, pour, pretend, pull, push, put, reach, read, recognize, record, remember, save, say, scare, scratch, see, shake, share, shave, shoot, shut, sing, smack, snap, spell, spill, spray, squeeze, squish, start, stir, stop, study, swing, take, tape, tear, think, throw, tickle, tie, touch, turn, understand, untie, use, wash, watch, wear, wipe, wish	1: answer, beat, bite, blow, bother, break, bring, brush, build, bump, burn, buy, carry, catch, change, chase, check, chew, chop, clean, close, color, cool, count, cover, crack, cut, do, draw, drink, drop, dry, dump, eat, feel, find, finish, fix, forget, get, have, hear, help, hit, hold, hook, hurt, keep, kick, leave, lose, make, match, meet, mind, miss, move, mow, need, open, paint, pass, pinch, poke, pour, pull, push, put, reach, read, recognize, record, save, say, scare, scratch, see, set, shake, share, shave, shoot, shut, sing, smack, smell, snap, spell, spill, spray, squeeze, squish, start, stick, stir, stop, study, suppose, swing, take, tape, tear, test, throw, tickle, tie, touch, turn, untie, use, wash, watch, wear, wipe
			2: decide, know, love, mean, remember, understand
2: get, smell, suppose, test		2: get, suppose, test	
3: rock, taste, tease		3: rock, taste	3: rock, tease
4: belong, bob, bounce, care, cook, crawl, cry, dance, drive, fit, fly, grow, hammer, hang, hide, hop, jump, laugh, lay, listen, live, look, peek, play, ride, ring, row, run, sit, sleep, slip, sound, squeak, stand, stay, step, swim, talk, visit, walk, wind, work, write	2: bang, belong, bob, bounce, care, chirp, climb, come, cook, crawl, cry, dance, drive, fall, fit, fly, go, grow, hammer, hang, hide, hop, jump, laugh, lay, let, lie, listen, live, look, peek, peepee, play, point, pop, ride, ring, row, run, sit, sleep, slip, sound, speak, squeak, stand, stay, step, swim, talk, try, walk, whisper, work, write	4: bang, belong, bob, bounce, care, cook, cry, dance, drive, fit, fly, grow, hammer, hang, hop, jump, laugh, lay, lie, listen, live, look, peek, play, point, ride, ring, row, run, set, sit, sleep, slip, sound, stand, stay, step, swim, talk, tease, visit, walk, wind, work, write	4: bang, belong, bounce, care, cook, crawl, drive, fit, fly, grow, hammer, hang, hide, hop, jump, laugh, listen, live, look, peek, play, ride, ring, run, sit, sleep, slip, sound, squeak, stay, step, swim, talk, walk, work, write
		5: come, peepee, pop, speak	5: climb, lay, lie, peepee, point, stand
5: fall, hurry			
	3: bet, guess, hope, think, wish		6: bet, guess, hope, pretend, think, wish
	4: break, burn, meet, scratch, squish, untie		
6: bump, call, drop, serve	5: call, serve	6: bump, call, drop, feel, serve	7: call, serve
	6: carry, chew, chop, cover, dump, hit, hook, match, mow, pass, poke, put, shoot, take, tear, throw, tie, turn, wash		
7: dig, juggle	7: dig, juggle	7: dig, juggle	8: dig, juggle
	8: drip, smile		
8: entertain, excuse, thank	9: entertain, excuse, thank	8: entertain, excuse, thank	9: entertain, excuse, thank
9: freeze, smile	10: freeze, melt	9: freeze, smile	
10: learn, worry	11: learn, worry		
11: like, want	12: like, want	10: like, want	10: like, want
12: bless, pet, press	13: pet, press	11: bless, pet, press	11: pet, press
13: cool, set	14: set, start		
	15: smell, taste		
			12: bob, cry, row
14: chirp, climb, come, go, let, lie, peepee, point, pop, speak, try, whisper		12: chirp, climb	13: chirp, come, dance, go, let, pop, speak, try, whisper
		13: go, let, try, whisper	
15: feed, show		14: feed, show	14: feed, give, offer, show
16: lick, peel		15: lick, peel	15: lick, peel

Table 12: <3yrs dataset: Aggregate inferred classes over 10 runs, given 4 strategies that don't involve the use of surface morphology (-surface-morphology), an intermediate representation (absolute/relative), and an expectation of a mapping between the intermediate representation and observable syntactic positions (+/-expect-mapping).

-surface-morphology			
absolute		**relative**	
+expect-mapping	**-expect-mapping**	**+expect-mapping**	**-expect-mapping**
singletons: begin, figure, hurry, seem, tip	*singletons: begin, figure, seem, tip*	*singletons: begin, figure, seem, tip*	*singletons: begin, figure, seem, tip*
1: answer, ask, bite, bother, bring, brush, build, bump, burn, buy, call, carry, catch, change, chase, check, chop, close, color, count, cover, crack, cut, dig, do, draw, drink, drop, dry, dump, eat, feed, find, finish, fix, fold, get, give, have, hear, hit, hold, hook, juggle, keep, kick, kiss, knock, leave, lose, make, match, miss, mix, move, mow, name, need, offer, open, pass, peel, pet, pick, pinch, plug, poke, pour, press, pull, push, put, reach, read, recognize, record, rock, roll, save, say, scare, scratch, send, serve, shake, share, shoot, show, shut, sign, sing, smack, spell, spill, spray, stir, stop, study, swing, take, taste, teach, tear, tell, test, throw, tickle, tie, touch, turn, untie, visit, wash, watch, wear, wind, wipe	1: answer, ask, bite, blow, bother, break, bring, brush, build, bump, burn, buy, call, carry, catch, change, chase, check, chop, clean, close, color, count, cover, crack, cut, dig, do, draw, drink, drop, dry, dump, eat, find, finish, fix, fold, get, have, hear, hit, hold, hook, hurt, juggle, keep, kick, kiss, knock, leave, lose, love, make, match, meet, mind, miss, mix, move, mow, need, open, paint, pass, peel, pet, pick, pinch, plug, poke, pour, press, pull, push, put, reach, read, recognize, record, rock, roll, save, say, scare, scratch, see, send, shake, share, shoot, shut, sign, sing, smack, snap, spell, spill, spray, squeeze, squish, stir, stop, study, swing, take, tape, taste, tear, test, throw, tickle, tie, touch, turn, untie, visit, wash, watch, wear, wind, wipe	1: answer, bite, bother, build, bump, burn, buy, call, carry, catch, change, chase, close, color, count, crack, cut, dig, do, draw, drink, drop, dry, dump, eat, find, finish, fix, get, have, hear, hit, hold, juggle, keep, kick, kiss, leave, lose, make, miss, move, mow, need, open, peel, pet, pinch, poke, press, push, reach, read, recognize, record, rock, say, scare, scratch, send, shake, share, shut, sign, sing, spell, spill, spray, stir, stop, swing, taste, test, tickle, touch, untie, watch, wear, wipe	1: answer, bite, bring, build, bump, burn, buy, call, carry, catch, change, chase, close, count, crack, cut, dig, do, draw, drink, drop, dry, dump, eat, find, fix, get, have, hear, hit, juggle, kick, kiss, leave, lose, love, make, miss, move, mow, need, open, peel, pet, pinch, poke, pour, press, put, reach, read, recognize, record, save, say, scare, scratch, send, shake, share, shoot, shut, sign, sing, spell, spill, spray, squeeze, stir, swing, take, taste, test, throw, tickle, touch, untie, wash, watch, wear, wipe, 1: ask, feed, give, name, offer, serve, show, smack, teach, tell
			2: bother, keep, stop
		2: ask, beat, blow, bring, check, chew, chop, clean, cover, fold, hook, knock, lick, match, mix, pass, pick, plug, pour, pull, put, roll, save, shoot, slide, smack, take, tear, throw, tie, turn, wash, wind	
	2: bang, cook, cool, hammer, listen, play, ride, set, write		3: bang, cook
2: beat, lick	3: beat, lick		4: beat, knock
3: bang, belong, climb, come, cook, crawl, feel, fill, grow, hammer, hang, lay, lean, lie, listen, look, peek, play, point, ride, run, set, sit, slide, smell, sound, stand, step, stick, talk, wake, worry, write	4: belong, climb, crawl, dress, fall, feel, fill, grow, hang, jump, lay, lie, look, peek, peepee, point, pop, run, sit, sleep, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, wake, worry	3: bang, belong, bounce, climb, cook, cool, crawl, dress, drive, fall, feel, fill, fit, fly, grow, hammer, hang, jump, lay, lean, lie, listen, look, peek, peepee, play, point, pop, ride, run, set, sit, sleep, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, wake, worry, write	5: belong, bounce, climb, cool, crawl, dress, fall, feel, fill, fit, fly, grow, hammer, hang, hop, jump, lay, lean, lie, listen, look, peek, peepee, play, point, pop, ride, run, set, sit, sleep, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, wake, worry, write
4: bet, confuse, guess, hope, pretend, think, wish	5: bet, confuse, guess, hope, pretend, think, wish	4: bet, confuse, guess, hope, pretend, think, wish	6: bet, confuse, guess, hope, pretend, think, wish
5: bless, entertain, excuse, thank	6: bless, entertain, excuse, thank	5: bless, entertain, excuse, thank	7: bless, entertain, excuse, thank
6: blow, chew, clean			8: blow, check, chew, chop, clean, color, cover, fold, hold, hook, lick, match, mix, pass, pick, plug, pull, push, roll, tear, tie, turn, wind
7: bob, bounce, care, chirp, cool, cry, dance, dress, drip, drive, fall, fit, fly, happen, hide, jump, laugh, live, melt, peepee, pop, ring, sleep, slip, speak, squeak, stay, swim, walk, work	7: bob, bounce, care, chirp, cry, dance, drip, drive, fit, fly, happen, hide, laugh, live, melt, ring, smile, squeak, walk, work	6: bob, care, chirp, cry, dance, drip, happen, laugh, live, melt, ring, squeak, walk, work	9: bob, care, chirp, cry, dance, drip, happen, laugh, live, melt, ring, squeak, walk, work
8: break, mind, paint, snap, squeeze, squish		7: break, brush, hurt, meet, mind, see, snap, squeeze, squish, study, tape	10: break, brush, drive, finish, hide, meet, paint, rock, snap, squish, visit
	8: come, learn		
9: decide, know, mean, remember, tease, understand, wonder	9: decide, know, mean, remember, understand, wonder	8: decide, know, mean, remember, understand, wonder	11: decide, hurt, know, mean, mind, remember, see, study, tape, understand, wonder
	10: feed, give, name, offer, serve, show, teach, tell	9: feed, give, name, offer, serve, show, teach, tell	
10: forget, go, help, let, like, start, suppose, try, use, want	11: forget, like, suppose, use, want	10: forget, go, learn, let, like, start, suppose, try, use, want	12: forget, go, help, learn, let, like, start, suppose, try, use, want
	12: go, let, start, try		
11: hurt, meet, see, tape			
12: manage, smile	13: manage, shave	12: manage, shave	
13: row, wait	14: row, wait	13: row, wait	13: row, wait
14: freeze, rain		14: freeze, rain	14: freeze, rain

Table 13: <4yrs dataset: Aggregate inferred classes over 10 runs, given 4 strategies that involve the use of surface morphology (+surface-morphology), an intermediate representation (absolute/relative), and an expectation of a mapping between the intermediate representation and observable syntactic positions (+/-expect-mapping).

+surface-morphology			
absolute		relative	
+expect-mapping	**-expect-mapping**	**+expect-mapping**	**-expect-mapping**
singletons: begin, confuse, dress, figure, land, rain, seem, wonder	*singletons: begin, confuse, dress, figure, land, rain, seem, wonder*	*singletons: begin, confuse, dress, figure, hurry, land, rain, seem, tip, wait, wonder*	*singletons: begin, confuse, dress, figure, hurry, land, plug, rain, seem, tip, wait, wonder*
1: answer, attach, back, bear, believe, bend, bite, blow, bother, break, bring, build, bump, burn, buy, carry, catch, change, chase, check, chew, chop, clean, close, color, count, cover, crack, cross, cut, decide, do, draw, drink, drop, dry, dump, eat, examine, feel, find, finish, fix, fold, forget, get, hang, have, hear, help, hit, hold, hook, hurt, keep, kick, kiss, know, leave, lose, make, match, mean, meet, mind, miss, move, mow, open, paint, park, pinch, poke, pour, press, pretend, pull, push, put, reach, read, recognize, record, remember, roll, save, say, scare, scratch, screw, see, sell, send, set, shake, share, shoot, shut, sing, smell, snap, spell, spill, squish, stick, stir, study, surprise, swing, take, tape, tear, test, throw, tickle, tie, touch, trade, turn, understand, untie, use, visit, wash, watch, wear, win, wind, wipe	1: answer, back, beat, bend, bite, blow, bother, break, bring, build, bump, burn, buy, carry, catch, change, chase, check, chew, chop, clean, close, color, count, cover, crack, cross, cut, do, draw, drink, drop, dump, eat, fix, hang, hit, hold, hook, kick, kiss, knock, leave, lose, match, miss, move, mow, open, paint, park, pass, pinch, poke, pour, press, pull, punch, push, put, reach, read, record, roll, scare, scratch, screw, sell, set, shake, share, shave, shoot, shut, sing, smack, spell, spill, squish, stir, study, swing, take, tape, tear, thank, throw, tickle, tie, touch, trade, turn, untie, use, visit, wash, wear, wind, wipe	1: answer, attach, bear, believe, bite, bother, break, bring, build, bump, burn, buy, call, catch, change, chase, check, chop, clean, close, count, crack, cross, decide, do, draw, drink, drop, dry, eat, examine, feel, find, finish, fix, forget, get, have, hear, help, hit, hold, hurt, keep, kill, kiss, know, leave, lose, make, mean, meet, mind, miss, move, name, open, paint, park, peel, pinch, pretend, reach, read, recognize, record, remember, save, say, scare, scratch, see, sell, shake, share, shoot, shut, sing, smell, snap, spell, spill, spray, squish, stick, stir, study, suppose, surprise, tape, test, tickle, touch, trade, understand, untie, use, visit, watch, wear, win	1: answer, ask, attach, back, bear, believe, bend, bite, blow, bother, break, bring, build, bump, burn, buy, carry, catch, change, chase, check, chew, chop, clean, close, color, count, cover, crack, cross, cut, decide, do, draw, drink, drop, dry, dump, eat, examine, feed, feel, find, finish, fix, fold, forget, get, give, hang, have, hear, help, hit, hold, hook, hurt, keep, kick, kill, kiss, knock, know, leave, lick, lose, love, make, match, mean, meet, mind, miss, move, mow, need, offer, open, paint, park, pass, peel, pinch, poke, pour, press, pretend, pull, punch, push, put, reach, read, recognize, record, remember, roll, save, say, scare, scratch, screw, see, sell, send, set, shake, share, shave, shoot, show, shut, sing, smack, smell, snap, spell, spill, spray, squish, stick, stir, study, suppose, surprise, swing, take, tape, tear, test, throw, tickle, tie, touch, trade, turn, understand, untie, use, visit, wash, watch, wear, win, wind, wipe, wish
	2: attach, believe, decide, dry, examine, feel, find, finish, forget, get, have, hear, help, hurt, keep, know, make, mean, meet, mind, pretend, recognize, remember, save, say, see, smell, snap, suppose, surprise, test, understand, watch, win		
		2: back, bend, blow, carry, chew, color, cover, cut, dump, fold, hang, hook, kick, knock, match, mow, pass, poke, pour, press, pull, punch, push, put, roll, screw, send, set, swing, take, tear, throw, tie, turn, wash, wind, wipe	
2: ask, smack, spray, thank	3: ask, spray	3: ask, smack	
3: bang, belong, bob, bounce, care, climb, cook, crash, crawl, cry, dance, dream, drip, drive, fall, fight, fit, fly, grow, hammer, hide, hop, jump, laugh, lay, lean, lie, listen, live, look, manage, march, melt, peek, peepee, play, point, pop, ride, ring, row, run, sit, sleep, slip, smile, sound, speak, squeak, stand, stay, step, swim, talk, tease, wake, walk, work, write	4: bang, belong, bob, bounce, care, climb, cook, crash, crawl, cry, dance, dream, drive, fall, fight, fit, fly, grow, hammer, hide, hop, jump, laugh, lay, lie, listen, live, look, march, peek, peepee, play, point, pop, ride, ring, row, run, sit, sleep, slip, smile, sound, speak, squeak, stand, stay, step, swim, talk, taste, wake, walk, work, write	4: bang, belong, bob, bounce, care, climb, cook, cool, crash, crawl, cry, dance, dream, drip, drive, fall, fight, fit, fly, grow, hammer, hide, hop, jump, laugh, lay, lean, lie, listen, live, look, march, peek, peepee, play, point, pop, ride, ring, row, run, sit, sleep, slip, smile, sound, speak, squeak, stand, stay, step, swim, talk, taste, tease, wake, walk, work, write	2: bang, belong, bob, bounce, care, climb, cook, cool, crash, crawl, cry, dance, dream, drive, fall, fight, fit, fly, grow, hammer, hide, hop, jump, laugh, lay, lean, lie, listen, live, look, march, peek, play, point, pop, ride, ring, rock, row, run, sit, sleep, slide, slip, smile, sound, speak, squeak, stand, stay, step, swim, talk, taste, tease, wake, walk, work, write
4: beat, knock, pass, punch			
5: bet, guess, hope, think	5: bet, guess, hope, think, wish	5: bet, guess, hope	3: bet, guess, hope, think
6: bless, excuse	6: bless, excuse	6: bless, excuse	4: bless, excuse
7: call, name, serve	7: call, serve		5: call, name, serve
8: chirp, disappear	8: chirp, disappear		
9: come, go, let, shop, try, whisper	9: come, go, let, melt, shop, try, whisper	7: chirp, come, go, let, melt, shop, try, whisper	6: chirp, come, go, let, peepee, shop, try, whisper
			7: disappear, happen, manage, melt
	10: freeze, ski		
		8: frighten, measure	8: frighten, measure
10: give, offer, show	11: feed, show	9: feed, give, offer, show, tell	
			9: learn, worry
11: like, want	12: like, want	10: like, want	10: like, want
12: love, need	13: love, need	11: love, need	
13: peel, pet			
	14: rock, tease	12: rock, shave	
		12: sign, squeeze	
14: start, stop	15: start, stop	14: start, stop	11: start, stop
15: suppose, wish			

Table 14: <4yrs dataset: Aggregate inferred classes over 10 runs, given 4 strategies that don't involve the use of surface morphology (-surface-morphology), an intermediate representation (absolute/relative), and an expectation of a mapping between the intermediate representation and observable syntactic positions (+/-expect-mapping).

-surface-morphology			
absolute		**relative**	
+expect-mapping	**-expect-mapping**	**+expect-mapping**	**-expect-mapping**
singletons: begin, figure, row, seem, wait	*singletons: begin, figure, row, seem, wait*	*singletons: begin, figure, seem, tip, wait*	*singletons: begin, figure, row, seem, tip, wait*
1: answer, bite, bother, build, bump, catch, chase, check, close, dig, drink, dry, eat, frighten, kill, lose, measure, meet, move, mow, open, peel, reach, read, scare, scratch, shake, shut, sing, spell, spray, squeeze, tickle, touch, trade, visit	1: answer, bake, bother, break, brush, build, bump, buy, call, catch, change, chase, check, close, cook, crack, cross, dig, do, draw, drop, dry, eat, examine, find, finish, fix, frighten, get, have, hear, hide, keep, kill, kiss, lose, love, make, mean, measure, meet, miss, move, mow, name, need, open, paint, peel, pet, pinch, reach, read, recognize, say, scare, scratch, sell, shake, share, sign, sing, snap, spell, spill, spray, squeeze, stir, stop, study, test, tickle, touch, trade, untie, use, visit, watch, wear, win	1: answer, bake, bite, bother, break, brush, build, bump, burn, buy, call, carry, catch, change, chase, check, close, color, cook, count, crack, cut, dig, do, draw, drink, drop, dry, dump, eat, find, finish, fix, forget, frighten, get, have, hear, help, hit, hold, keep, kick, kill, leave, lose, love, make, mean, measure, meet, move, mow, name, need, open, park, peel, poke, reach, read, recognize, record, say, scare, scratch, screw, sell, shake, share, shoot, shut, sing, spell, spill, spray, stir, stop, study, swing, test, tickle, tie, touch, trade, use, watch, wear, wipe	1: answer, bake, bite, bother, break, brush, build, bump, burn, buy, call, carry, catch, change, chase, check, close, color, cook, count, crack, cut, dig, do, draw, drink, drop, dry, dump, eat, find, finish, fix, frighten, get, have, hear, help, hit, hold, keep, kick, kill, leave, lose, love, make, mean, measure, meet, move, mow, open, park, peel, plug, poke, reach, read, recognize, record, say, scare, scratch, screw, sell, shake, share, shoot, shut, sign, sing, spell, spill, spray, squeeze, stir, stop, study, swing, test, tickle, touch, trade, use, watch, wear, wipe
2: ask, bake, break, bring, brush, buy, call, carry, change, cook, crack, do, draw, drop, find, fix, get, have, hear, hit, keep, love, make, mean, mind, name, need, paint, park, recognize, say, see, share, shoot, smack, spill, stir, stop, study, tape, test, use, watch, wear, win			
3: attach, beat, bend, blow, burn, chop, clean, color, count, cover, cut, dump, fold, hang, hold, hook, kick, knock, leave, lick, match, mix, pass, pick, plug, poke, pour, press, pull, punch, push, put, record, roll, save, screw, slide, spank, swing, take, tear, throw, tie, turn, wash, wind, wipe	2: attach, beat, bend, bring, cover, cut, fill, hold, hook, knock, lick, mix, pass, pick, plug, pour, press, pull, punch, push, put, save, spank, take, throw, turn, wash, wind	2: attach, beat, bend, hook, knock, lick, mix, pass, pick, pour, press, punch, push, put, save, take, tear, throw, wash, wind	2: attach, beat, bend, bring, hook, knock, lick, mix, pass, pick, pour, press, punch, push, put, save, spank, take, throw, wash, wind
	3: bite, carry, color, count, dump, hit, kick, leave, park, poke, screw, shoot, swing, wipe		
	4: blow, burn, chew, chop, clean, drink, drive, fold, hang, match, record, rock, roll, shut, slide, tear, tie	3: blow, chew, chop, clean, cover, drive, fold, hang, match, paint, pull, rock, roll, slide, turn	3: blow, chew, chop, clean, cover, fill, fold, hang, match, pull, roll, slide, tear, tie, turn
4: back, hurry	5: back, hurry		
		4: bear, live	4: bear, live
5: bang, bounce, climb, cool, crash, crawl, dance, dress, fall, feel, fight, fit, fly, grow, hammer, hop, jump, lay, lean, lie, listen, look, peek, peepee, play, point, pop, ride, run, set, shave, sit, sleep, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, taste, wake, walk, worry, write	6: bang, fight, grow, hammer, play, ride, set, shave, taste, write	5: bang, bounce, climb, cool, crash, crawl, dance, fall, feel, fight, fit, fly, grow, hammer, happen, hop, hurry, jump, laugh, lay, lean, lie, listen, look, peek, play, point, pop, ride, run, set, shave, sit, sleep, slip, smell, sound, speak, squeak, stand, stay, step, stick, swim, talk, taste, wake, walk, worry, write	5: bang, bounce, climb, cool, crash, crawl, dance, fall, feel, fight, fit, fly, grow, hammer, happen, hop, jump, lay, lean, lie, listen, look, peek, play, point, pop, ride, rock, run, set, shave, sit, sleep, slip, smell, sound, speak, squeak, stand, stay, step, stick, swim, talk, taste, visit, wake, walk, worry, write
	7: bounce, climb, cool, crash, crawl, dance, fall, feel, fit, fly, hop, jump, lay, lean, lie, listen, look, peek, peepee, point, pop, run, sit, sleep, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, wake, walk, worry		
6: believe, decide, know, remember, surprise, understand, wonder	8: believe, decide, hurt, know, mind, remember, see, surprise, tape, understand, wonder	6: believe, decide, hurt, know, mind, miss, remember, see, squeeze, surprise, tape, understand, win, wonder	6: believe, decide, hurt, know, mind, miss, remember, see, surprise, tape, understand, wonder
7: bet, guess, hope, think	9: bet, guess, hope, think	7: bet, confuse, guess, hope, think	7: bet, confuse, guess, hope, think
8: bless, entertain, excuse, pay, thank,	10: bless, entertain, excuse, pay, thank	8: bless, entertain, excuse, thank	8: bless, entertain, excuse, thank
9: bob, chirp, ski	11: bob, chirp, ski	9: bob, chirp, ski	9: bob, chirp, ski
10: care, cry, disappear, dream, drip, laugh, march, melt, smile, work	12: care, cry, disappear, dream, drip, laugh, manage, march, melt, ring, smile, work	10: care, cry, disappear, dream, drip, manage, march, melt, ring, smile, work	10: care, cry, disappear, dream, drip, manage, march, melt, ring, smile, work
11: chew, fill			
12: come, shop			
13: cross, examine, kiss, pet, pinch, sell, sign, snap, squish		11: cross, examine, kiss, pet, pinch, untie	11: cross, examine, kiss, pet, pinch, untie
		12: snap, squish	
14: feed, give, offer, send, serve, show, teach, tell	13: ask, feed, give, offer, send, serve, show, smack, teach, tell	13: ask, bring, feed, give, offer, send, serve, show, smack, spank, teach, tell	12: ask, feed, give, name, offer, send, serve, show, smack, teach, tell
15: finish, hide, hurt, miss			
16: forget, help, like	14: forget, like		13: forget, like, need
17: freeze, rain	15: freeze, rain	14: freeze, rain	
18: go, land, learn, let, start, suppose, try, want	16: come, go, land, learn, let, shop, start, suppose, try, want	14: come, go, land, learn, let, peepee, shop, start, try	15: come, go, land, learn, let, shop, start, try
		15: suppose, want	16: suppose, want
19: happen, squeak	17: happen, squeak		
		16: hide, whisper	
20: juggle, tease, untie	18: juggle, squish, tease	17: juggle, tease	17: juggle, tease
21: pretend, wish	19: pretend, wish	18: pretend, wish	18: pretend, wish

Table 15: <5yrs dataset: Aggregate inferred classes over 10 runs, given 4 strategies that involve the use of surface morphology (+surface-morphology), an intermediate representation (absolute/relative), and an expectation of a mapping between the intermediate representation and observable syntactic positions (+/-expect-mapping).

+surface-morphology			
absolute		**relative**	
+expect-mapping	**-expect-mapping**	**+expect-mapping**	**-expect-mapping**
singletons: begin, bob, confuse, dig, dress, figure, hurry, land, rain, rinse, seem, tip, wait, wonder	*singletons: begin, confuse, dress, figure, hurry, land, rain, rinse, seem, tip, wonder*	*singletons: begin, confuse, dress, figure, rain, seem, wonder*	*singletons: begin, confuse, dress, figure, rain, seem, wonder*
1: answer, attach, bear, beat, bend, bite, blow, bother, break, bring, build, burn, buy, carry, catch, change, chase, check, chew, chop, clean, close, color, count, cover, crack, cross, cut, do, draw, drill, drink, drive, dump, eat, find, finish, fix, fold, follow, frighten, grind, hang, hit, hold, hook, hurt, keep, kick, kill, kiss, knock, leave, lick, lose, match, measure, miss, move, mow, open, paint, park, pass, peel, pinch, plug, poke, pour, press, pull, punch, push, put, reach, read, record, rewind, rock, roll, scare, scratch, screw, sell, set, shake, share, shave, shoot, shut, sing, slide, smack, spell, spill, squeeze, squish, stick, stir, straighten, stretch, swallow, sweep, swing, take, tape, tear, throw, tickle, tie, touch, trade, turn, twist, untie, visit, wash, wear, wind, wipe, wrap	1: answer, attach, bear, believe, bite, bother, break, bring, build, bump, burn, buy, catch, change, check, close, cook, count, cover, crack, cross, decide, do, draw, drink, drop, dry, dump, eat, examine, find, finish, fix, follow, forget, get, have, hear, help, hit, hold, hurt, keep, kill, kiss, know, leave, lose, make, match, mean, meet, mind, miss, move, open, paint, peel, pinch, poke, pretend, reach, read, recognize, record, remember, rewind, save, say, scare, scratch, see, sell, send, shake, share, sharpen, shave, shoot, sign, sing, smack, smell, snap, spell, spill, spray, squeeze, squish, stick, stir, straighten, suppose, surprise, swallow, swing, tape, test, tickle, touch, trade, twist, understand, untie, use, visit, watch, wear, win	1: answer, attach, bear, bend, bite, blow, bother, break, bring, brush, build, bump, burn, buy, carry, catch, change, chase, check, chop, clean, close, color, count, cover, crack, cross, cut, do, draw, drill, drink, drop, dump, eat, examine, find, finish, fix, fold, follow, frighten, get, hang, hear, hit, hold, hook, hurt, keep, kick, kill, kiss, leave, lick, lose, make, match, measure, meet, miss, move, mow, open, paint, park, peel, pet, pinch, poke, pour, press, pull, punch, push, put, reach, read, record, rewind, rock, roll, scare, scratch, screw, sell, send, set, shake, share, shoot, shut, sign, sing, slide, smell, spell, spill, spray, squeeze, squish, stick, stir, straighten, stretch, study, suppose, swallow, swing, take, tape, tear, test, throw, tickle, tie, touch, trade, turn, twist, untie, use, visit, wash, watch, wear, wind, wipe, wrap	1: answer, attach, bear, believe, bite, blow, bother, break, build, bump, burn, buy, carry, catch, change, chase, check, chop, clean, close, color, count, cover, crack, cross, cut, decide, do, draw, drink, drive, drop, dry, dump, eat, examine, feel, find, finish, fix, follow, forget, frighten, get, have, hear, help, hit, hold, hurt, keep, kill, kiss, know, leave, lose, make, match, mean, meet, mind, miss, move, open, paint, peel, pinch, poke, pretend, pull, reach, read, recognize, record, remember, rewind, roll, say, scare, scratch, screw, see, sell, shake, share, sharpen, shave, shoot, shut, sign, sing, smell, snap, spell, spill, spray, squeeze, squish, stick, stir, straighten, stretch, suppose, surprise, swallow, swing, tape, test, tickle, tie, touch, trade, turn, twist, understand, untie, use, visit, watch, wear, win, wipe
			2: bake, pet
		2: ask, smack, spank	
	2: beat, chew, knock, pass, plug, sweep	3: beat, chew, fill, grind, knock, mix, pass, pick, plug	
	3: bend, blow, brush, carry, chase, clean, color, cut, drill, fold, grind, hang, hook, kick, lick, measure, mow, park, pour, press, pull, punch, push, put, roll, screw, set, shut, slide, stretch, take, tear, throw, tie, turn, wash, wind, wipe, wrap		3: beat, bend, bring, chew, drill, fold, grind, hang, hook, kick, knock, lick, measure, mow, park, pass, plug, pour, press, punch, push, put, send, set, slide, smack, take, tear, throw, wash, wind, wrap
	4: drive, rock		4: rock, sweep
	5: frighten, pet		
2: back, belong, bounce, care, climb, come, cook, crash, crawl, cry, dance, dream, end, fall, fight, fit, fly, go, grow, hammer, hide, hop, jump, laugh, lay, lean, let, lie, listen, live, look, march, melt, peek, peepee, play, point, pop, rest, ride, ring, run, shop, sit, sleep, slip, smile, sneeze, sound, speak, squeak, stand, start, stay, step, stop, swim, talk, taste, try, wake, walk, whisper, work, worry, write	6: back, climb, come, cool, crash, end, fall, go, grow, lay, let, lie, peepee, point, pop, run, shop, stand, try, wake, whisper	4: back, bang, belong, bounce, care, chirp, climb, cook, cool, crash, crawl, cry, dance, disappear, dream, end, fall, fight, fit, fly, grow, hammer, happen, hide, hop, jump, laugh, lay, lean, lie, listen, live, look, manage, march, melt, peek, peepee, play, point, pop, rest, ride, ring, run, sit, sleep, slip, smile, sneeze, sound, speak, squeak, stand, stay, step, swim, talk, taste, trip, wake, walk, work, worry, write	5: back, chirp, climb, crash, end, fall, grow, lay, lie, melt, point, pop, run, slip, stand, wake
	7: bang, belong, bounce, crawl, fight, fit, fly, hammer, hop, jump, lean, listen, look, peek, play, rest, ride, ring, sit, slip, sound, speak, squeak, stay, step, swim, talk, taste, walk, worry, write		6: bang, belong, bounce, care, cook, crawl, cry, dance, dream, drip, fight, fit, fly, hammer, hide, hop, jump, laugh, lean, listen, live, look, march, peek, peepee, play, rest, ride, ring, sit, sleep, smile, sneeze, sound, speak, squeak, stay, step, swim, talk, taste, trip, walk, work, worry, write
	8: bob, cry, disappear, laugh, manage, march, sneeze		
	9: care, dream, live, sleep, trip, work		
		5: come, go, let, shop, try, whisper	7: come, go, let, shop, try, whisper
	10: dance, hide, start, stop		
	11: happen, melt		8: disappear, happen
		6: start, stop	9: start, stop
3: believe, decide, drop, forget, have, hear, help, know, love, make, mean, mind, need, pretend, recognize, remember, say, see, snap, surprise, understand, use, win, wish	12: love, need	7: believe, decide, forget, have, know, mean, mind, pretend, recognize, remember, say, see, sharpen, snap, surprise, understand, win	
4: bet, guess, hope, think	13: bet, guess, hope, think, wish	8: bet, think	10: bet, guess, hope, wish
		9: guess, hope	
5: bless, excuse, thank	14: bless, excuse, thank	10: bless, excuse, thank	11: bless, excuse, thank
6: bump, examine, sign, spray			
	15: call, name, serve	11: call, name, serve	12: call, serve
7: drip, row		12: drip, row, ski	
8: dry, feel, watch		13: drive, dry, feel, shave	
9: feed, show	16: feed, show	14: feed, give, offer, pay, save, show, tell	13: feed, give, offer, save
10: freeze, ski			
11: get, meet, send, smell, test			
12: learn, study, tease	17: learn, tease		
13: like, want	18: like, want	15: like, love, need, want	14: like, want
			15: love, need

Table 16: <5yrs dataset: Aggregate inferred classes over 10 runs, given 4 strategies that don't involve the use of surface morphology (-surface-morphology), an intermediate representation (absolute/relative), and an expectation of a mapping between the intermediate representation and observable syntactic positions (+/-expect-mapping).

-surface-morphology			
absolute		**relative**	
+expect-mapping	**-expect-mapping**	**+expect-mapping**	**-expect-mapping**
singletons: bear, begin, figure, row, seem, tip, wait	*singletons: begin, figure, row, seem*	*singletons: figure*	*singletons: figure*
1: answer, bite, brush, build, bump, burn, catch, change, chase, check, close, count, cover, dig, draw, drink, drive, eat, fix, frighten, kill, match, move, mow, open, peel, pull, reach, read, record, rewind, scratch, screw, set, shake, shoot, shut, sign, spray, stir, swallow, swing, tie, touch, trade, visit	1: answer, bake, bother, break, build, bump, buy, carry, catch, change, chase, check, close, cook, count, crack, dig, do, draw, drink, drive, drop, dry, eat, find, finish, fix, follow, forget, frighten, get, have, hear, help, hide, hit, keep, kick, kill, learn, like, lose, love, make, mean, measure, meet, move, mow, need, open, paint, peel, read, recognize, rewind, say, scare, scratch, sell, share, shoot, sing, spell, spill, stir, swallow, swing, tape, test, tickle, touch, trade, untie, use, watch, wear	1: answer, bake, bite, bother, bring, build, bump, burn, buy, carry, catch, change, chase, check, close, color, count, crack, do, draw, drink, drive, drop, eat, find, fix, follow, frighten, get, have, hear, hold, keep, kick, kill, leave, lose, love, make, measure, move, mow, need, open, peel, read, recognize, record, scare, scratch, screw, sell, share, shoot, shut, sign, spank, spell, spill, squeeze, stir, swallow, swing, tape, test, tie, touch, trade, untie, use, visit, watch, wear	1: answer, bake, bite, bother, build, bump, burn, buy, call, carry, catch, change, chase, check, close, color, count, cover, crack, cut, dig, do, draw, drink, drive, drop, eat, find, fix, follow, frighten, get, have, hear, hit, hold, keep, kick, kill, leave, lose, love, make, measure, move, mow, name, need, open, peel, reach, read, recognize, record, scratch, screw, sell, set, share, shoot, shut, sign, spell, spill, spray, stir, swallow, swing, test, tie, touch, trade, use, visit, watch, wear
			2: cook, paint, untie
		2: forget, like	3: forget, like
		3: meet, miss	
2: ask, feed, give, offer, pay, send, serve, show, smack, teach, tell	2: ask, call, feed, give, offer, pay, send, serve, show, smack, teach, tell	4: ask, call, feed, give, hit, name, offer, pay, send, serve, show, smack, teach, tell	4: ask, feed, give, offer, pay, send, serve, show, smack, teach, tell
3: attach, beat, bend, bring, chew, chop, clean, color, cut, drill, dump, fill, fold, hang, hold, hook, kick, knock, leave, lick, lift, mix, park, pass, pick, plug, poke, pour, press, punch, push, put, rinse, roll, save, slide, spank, take, tear, throw, turn, wash, wind, wipe, wrap	3: attach, bend, bite, bring, burn, chop, clean, color, cover, cut, drill, dump, fold, hold, hook, knock, leave, lick, mix, park, pick, plug, poke, pour, press, pull, punch, push, put, reach, record, rinse, roll, save, screw, set, shake, shut, sign, slide, spank, spray, squeeze, take, tear, throw, tie, turn, visit, wash, wind, wipe, wrap	5: attach, beat, bend, brush, chew, chop, clean, cover, cut, drill, dump, fill, fold, hang, hook, knock, lick, lift, mix, park, pass, pick, plug, poke, pour, press, pull, punch, push, put, rinse, roll, save, set, slide, take, tear, throw, turn, wash, wind, wipe, wrap	5: attach, beat, bend, bring, brush, chop, clean, drill, dump, fold, hook, knock, lick, mix, park, pass, pick, plug, poke, pour, press, pull, punch, push, put, rinse, roll, save, spank, take, tear, throw, turn, wash, wind, wipe
	4: chew, hang		6: chew, fill, hang, slide, wrap
4: back, bang, bounce, climb, cool, crash, crawl, dance, end, fall, feel, fight, fit, fly, grow, hammer, hop, jump, lay, lie, listen, look, peek, play, point, pop, rest, ride, run, shave, sit, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, taste, wake, walk, worry, write,	5: back, fall, fit, grow, hammer, hop, hurry, lay, lie, listen, peek, play, point, pop, rest, ride, run, shave, sit, slip, smell, stand, stay, step, stick, talk, taste, wake, write	6: back, bounce, climb, come, cool, crash, crawl, dance, dress, end, fall, feel, fit, fly, grow, hop, hurry, jump, lay, lean, lie, listen, look, peek, peepee, play, point, pop, rest, ride, run, shave, shop, sit, slip, smell, sound, speak, stand, stay, step, stick, swim, talk, taste, wake, walk, worry, write	7: back, bang, blow, bounce, climb, cool, crash, crawl, dance, end, fall, feel, fight, fit, fly, grow, hammer, hop, hurry, jump, lay, lie, listen, look, match, peek, play, point, pop, rest, ride, rock, run, shave, sit, slip, smell, sound, speak, stand, stay, step, stick, sweep, swim, talk, taste, wake, walk, worry, write
	6: bang, fight, fly, worry	7: bang, fight, hammer	
	7: bounce, climb, cool, crash, crawl, dance, end, jump, speak, squeak, swim, walk		
5: bake, bother, buy, call, carry, crack, do, drop, find, follow, get, have, hear, help, hit, keep, lose, love, make, measure, name, need, recognize, sell, share, spell, spill, tape, test, use, watch, wear			
6: believe, decide, finish, hurt, know, mean, mind, miss, pretend, remember, say, see, study, surprise, understand, wish, wonder	8: believe, decide, hurt, know, mind, miss, pretend, remember, see, study, surprise, understand, wish, wonder	8: believe, decide, hurt, know, mean, mind, pretend, remember, say, see, study, surprise, understand, wish, wonder	8: believe, break, decide, dry, finish, hurt, know, mean, meet, mind, miss, remember, rewind, say, scare, see, sing, squeeze, stop, study, surprise, tape, tickle, understand, win, wonder
			9: pretend, wish
7: belong, trip	9: belong, trip	9: belong, trip	10: belong, trip
8: bet, guess, hope, think	10: bet, confuse, guess, hope, think	10: bet, guess, hope, think	11: bet, guess, hope, think
9: bless, entertain, excuse, thank	11: bless, entertain, excuse, thank	11: bless, entertain, excuse, thank	12: bless, entertain, excuse, thank
10: blow, rock, sweep	12: blow, match, rock, sweep	12: blow, match, rock, sweep	
11: bob, chirp, ski, sneeze	13: bob, chirp, ski, sneeze	13: bob, chirp, freeze, rain, ski, sneeze	13: bob, chirp, ski, sneeze
			14: freeze, rain
12: break, cook, dry, hide, meet, paint, scare, sing, squeeze, tickle, untie		14: break, cook, finish, paint, reach, rewind, shake, sing, spray, stop, win	
13: care, cry, dream, drip, happen, laugh, live, march, melt, ring, sleep, smile, work		15: care, cry, disappear, dream, drip, happen, laugh, live, march, melt, ring, sleep, smile, squeak, work	15: care, cry, dream, drip, happen, laugh, live, march, melt, ring, sleep, smile, squeak, work
14: come, dress, go, land, lean, learn, let, peepee, shop, start, stop, try	14: care, cry, dream, drip, happen, laugh, live, march, melt, ring, sleep, smile, work	16: go, land, learn, let, try, whisper	16: come, dress, land, lean, peepee, shop
	15: come, go, land, lean, let, peepee, start, try, whisper		
15: cross, examine, kiss, pet, pinch, sharpen, snap, squish, stretch, twist	16: cross, examine, kiss, pinch, snap, squish, stretch, twist	17: cross, examine, kiss, pet, pinch, sharpen, stretch, twist	17: cross, examine, pet, pinch, sharpen
	17: dress, feel, look, shop, sound		
17: grind, straighten	18: grind, straighten	18: grind, straighten	18: grind, straighten
18: juggle, tease	19: juggle, tease	19: juggle, snap, squish, tease	19: juggle, kiss, snap, squish, stretch, tease, twist
	20: pet, sharpen		
19: suppose, want	21: suppose, want	20: suppose, want	20: go, learn, let, start, suppose, try, want