The acquisition of linking theories:
A Tolerance Principle approach to
deriving UTAH and rUTAH

Lisa Pearl & Jon Sprouse

Abstract

We investigate concrete acquisition theories for a derived approach to linking theory development, and explore to what extent two prominent linking theories in the syntactic literature – UTAH and rUTAH – can be derived from the data English children encounter. We leverage a conceptual acquisition framework that specifies key aspects of the child's acquisition task, including realistic child-directed input and a cognitively-motivated mechanism for inference (the Tolerance Principle). We find a learnability advantage for rUTAH over UTAH if children derive their linking theories from their input as specified here. We discuss the implications of these results for both syntactic theory and acquisition theory.

keywords: linking problem, UTAH, rUTAH, argument from acquisition, Tolerance Principle

1   Introduction

Consider the following sentence: "The little girl blicked the kitten on the stairs." Even if we don't know what "blick" means, we still have preferences about how to interpret this sentence. In particular, out of all the logically possible interpretations involving the little girl, the kitten, and the stairs, we prefer an interpretation where the little girl is doing something (blicking) to the kitten, and that event is happening on the stairs. The reason we as adults have this preferred interpretation is because we've solved the *linking problem*. That is, we have *linking theories* that link the thematic roles specified by a verb's lexical semantics to the syntactic argument positions specified by that verb's syntactic frame. Moreover, our linking theories are so well-developed that they can impose these links even when we don't know a verb's specific lexical semantics (as we see here with "blick").

Solving the linking problem – that is, developing linking theories – is a fundamental component of verb learning for children. The linking problem also appears to be relatively constrained in that there is one specific linking pattern (a **primary pattern**) that emerges 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 IN-STRUMENT/SOURCE/GOAL-like roles tend to appear in oblique syntactic positions such as *indirect*

object or *object of a prepositional phrase.* All linking theories must attempt to explain this very frequent pattern either by building it into the human mind in the form of innate knowledge (Fillmore, 1968; Baker, 1988; Perlmutter & Postal, 1984; Larson, 1990; Speas, 1990; Grimshaw, 1990), or by appealing to a specific interplay between the input that children receive and the mechanisms that underlie verb learning (Bowerman, 1988; Goldberg, 1995, 2006; Boyd & Goldberg, 2011; Goldberg, 2013).

The Uniformity of Theta Assignment Hypothesis (**UTAH**) and the relativized Uniformity of Theta Assignment Hypothesis (**rUTAH**) are two prominent linking theories in the syntactic literature that are typically associated with the innate approach to the linking problem. They therefore expect development of a linking theory to be the result of neurobiology, rather than informed by the language input children receive. In contrast, derived approaches to linking theory development posit the opposite: children use their input to generate their linking theories. However, to our knowledge, there are no specifications of derived approaches – that is, a concrete acquisition theory that assumes children derive linking theory knowledge from their input. We therefore take UTAH and rUTAH as two well-defined theories involving specific proposals for thematic representations (discussed in more detail in section 2), and investigate how the appropriate linking theories based on these thematic representations could be derived. That is, we aim to provide an existence proof for how linking theories could be derived from children's input, and thereby offer a concrete acquisition proposal for derived approaches to linking theory development.

UTAH and rUTAH make ideal case studies for exploring the acquisition of linking theories because (i) they are well-defined theories that can serve as the target of acquisition, (ii) they represent cognitively plausible linking theories, and (iii) they represent distinct locations in the space of possible linking theories (particularly along the fixed-relativized dimension, as discussed in more detail in section 2). Moreover, there is little modeling research on the specific mechanisms underlying the acquisition of linking theories. We aim to fill this gap here. In particular, we explore to what extent the Tolerance Principle, a cognitively motivated decision criterion, could be used to learn the linking theories specified by UTAH and rUTAH from realistic samples of English child-directed speech data.

To concretely investigate derived approaches for linking theories, we first identify possible acquisition targets, in the form of linking theory variants defined by UTAH and rUTAH. We then discuss how these acquisition targets impact the potential acquisition process. We then review key acquisition modeling components and how they are implemented in the acquisition task of deriving linking theories. This includes (i) the learner's initial knowledge state that defines the hypothesis space for linking theories, (ii) the data the learner utilizes for acquisition, drawn from syntactically- and thematically-annotated English child-directed speech, and (iii) the inference process that leverages the Tolerance Principle to yield the appropriate target knowledge state.

Our results suggest that relativized approaches to linking theories like rUTAH are possible to derive from realistic English child-directed speech using the acquisition process we specify; this contrasts with fixed approaches like UTAH, where derivation fails. These acquisition modeling results complement the emerging empirical and theoretical debates surrounding UTAH and rUTAH in potentially interesting ways. We therefore conclude by discussing the implications of our results for both syntactic theory and acquisition theory.

2 Linking theories as the target of acquisition

Linking theories must have (at least) three components: a specification of the thematic roles in the grammatical system, a specification of the syntactic positions in the grammatical system, and at least one principle that governs the mapping between thematic roles and syntactic positions. Here, we will decompose the two linking theories used as case studies in this project – UTAH and rUTAH – into their three components, and review how they account for the primary linking pattern.

2.1 UTAH

Thematic roles. The UTAH linking theory assumes a finite number of thematic roles that are typically defined in terms of semantic features, although there is quite a bit of debate about what those features should be, and even whether such a specification is possible (Fillmore, 1968; Perlmutter & Postal, 1984; Jackendoff, 1987; Baker, 1988; Grimshaw, 1990; Speas, 1990; Dowty, 1991; Baker, 1997). For concreteness, here we follow the specific UTAH implementation from Baker (1997). Baker's implementation posits three fixed thematic macroroles (similar to Dowty's (1991) proto-roles), which we will indicate with small caps: AGENT, PATIENT, and OTHER. It 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 macroroles. For example, for Baker (1997), thematic roles that tend to involve internal causation (Levin & Rappaport Hovav, 1995) map to AGENT, roles that tend to involve external causation (Levin & Rappaport Hovav, 1995) map to PATIENT, and all other roles map to OTHER. Example (1) lists 13 common finer-grained thematic roles from the literature, and how they would map to the three macroroles in this implementation.

(1) Baker's (1997) three fixed macroroles and 13 common finer-grained thematic roles

    a. AGENT: agent, causer, experiencer (when internally-caused), possessor
    b. PATIENT: patient, theme, experiencer (when externally-caused), subject matter
    c. OTHER: location, source, goal, benefactor, instrument

Syntactic positions. Baker's (1997) formulation of UTAH similarly posits three syntactic positions, which are defined by specific syntactic features – again, with much debate about the details of the syntactic theory. For the purposes of this study, we can abstract away from many of these details. What matters is that there is regularity in the syntactic positions that can be mapped to regularity in the thematic roles. To that end, we'll simply call the syntactic positions in this implementation of UTAH *subject*, *object*, and *oblique* (such as the object of a prepositional phrase), and use italics to indicate these are theory-specific labels. We don't intend to imply that subject, object, and oblique are theoretical primitives, but instead use these as cover terms that readers can substitute with any relevant syntactic analysis (e.g., specifier of TP for *subject*).[1]

---

[1]The VP-internal subject hypothesis does add a layer of complexity to the problem, in that the *subject* is not the starting position for any thematic roles. But this complexity is regular: AGENTs always appear in *subject* position, whereas PATIENTs can appear in either *subject* or *object* position. So, this regularity allows us to abstract away from the details of the VP-internal subject hypothesis for this study.

Linking principle. Baker's (1997) formulation posits a linking principle that governs the mapping between thematic roles and syntactic positions: the AGENT role maps to the syntactic *subject* position, the PATIENT role maps to the syntactic *object* position, and the OTHER role maps to *oblique* positions.[2] That is, there are three links that together form a single 3-link theory.

Accounting for the primary linking pattern. With these three components in place, Baker's UTAH implementation can account for the primary linking pattern. In sentences such as *Jack cut the pie with a knife*, the AGENT appears in *subject* position, the PATIENT appears in *object* position, and the OTHER (the instrument) appears in an *oblique* position. Exceptions to this pattern, as in the *The package arrived*, where a PATIENT appears in *subject* position, are handled by a derivational grammatical system that includes a movement operation. The NP *the package* enters the derivation in *object* position, in accordance with Baker's UTAH system, and then is moved to the *subject* position at a later point in the derivation. In this way, apparent exceptions to the primary pattern are only exceptions on the surface; at an early stage of the syntactic derivation, UTAH is indeed respected.

2.2 rUTAH

Thematic roles. The mapping between thematic roles and syntactic positions in UTAH is *fixed* in the sense that each thematic role will map to the same syntactic position in every construction; in contrast, one of the defining features of rUTAH is that the mapping between thematic roles and syntactic positions is *relative* (hence the name – *relativized* Uniformity of Theta Assignment Hypothesis: Larson, 1988, 1990; Grimshaw, 1990; Speas, 1990). To achieve this, rUTAH first assumes that thematic roles are arranged in a hierarchy, such that certain thematic roles are "higher" or "lower" on the hierarchy than other roles. Example (2) lists 13 common finer-grained thematic roles in a hierarchy derived from Larson (1988, 1990). One interesting feature of the Larson hierarchy is that finer-grained roles need not be strictly ordered relative to one another. We indicate this by placing unordered roles in parentheses.

(2) Hierarchy derived from Larson (1988, 1990):
agent > causer > experiencer > possessor >
subject matter > causee > theme > patient >
(location, source, goal, benefactor, instrument)

Given this hierarchy, any given thematic role in a specific sentence can be relatively defined within that specific sentence as the HIGHEST, SECOND HIGHEST, THIRD HIGHEST, etc. To avoid the repetition of the word highest, we'll call these FIRST, SECOND, and THIRD here.

---

[2]Though it's not typically discussed in the syntax literature, creating an explicit model using UTAH requires deciding whether the links are unidirectional or bidirectional. Here we assume bidirectional links (e.g., AGENT links to *subject* and *subject* links to AGENT). This seems most plausible given that links can be used both for production (where a link from thematic role to syntactic position is useful) and for comprehension (where a link from syntactic position to thematic role is useful). That said, readers who prefer a unidirectional version of UTAH can simply evaluate the portion of our results for the preferred directionality.

Syntactic positions.   rUTAH similarly assumes a relative hierarchy for syntactic positions, often defined in structural terms (e.g., by c-command relations). For example, one common c-command-based hierarchy applied to the Baker-style syntactic positions would be *subject > object > oblique*. Here, we'll refer to the relative syntactic positions as *first-syn*, *second-syn*, and *third-syn*.

Linking principle.   rUTAH posits a linking principle that governs the mapping between the relativized thematic roles and the relativized syntactic positions: the FIRST thematic role maps to the *first-syn* syntactic position, the SECOND thematic role maps to the *second-syn* syntactic position, and so on.[3] So, as with UTAH, there are three links that together form a single 3-link theory. The difference is that in rUTAH, the links are defined over thematic roles and syntactic positions that are relative, rather than fixed.

Accounting for the primary linking pattern.   One interesting feature of rUTAH is that, by implementing a relativized system, many of the apparent exceptions to the primary linking pattern cease to be exceptions. For example, the sentence *The package arrived* is an apparent exception to UTAH that requires a derivational grammar and a movement operation in Baker's (1997) system. But, under rUTAH, it's a paradigm example of the rUTAH mapping: the one and only thematic role, patient, is the FIRST in the sentence, and it's mapped to the *first-syn* syntactic position, which is the *subject* position. There's no need for a movement operation (or, indeed, even a derivational grammar). This sentence is simply an example of the primary linking pattern. The fact that many of the exceptions to the linking patterns under UTAH become paradigmatic cases of the linking pattern under rUTAH will be particularly relevant for our acquisition models, as the Tolerance Principle is directly concerned with the ratio of exceptions to paradigmatic cases for any hypothesized rule.

2.3   UTAH and rUTAH as case studies

We focus on UTAH and rUTAH as case studies for modeling the acquisition of linking theories for several reasons. The most important reason is that these theories are specified in fine enough detail to make the scope of the acquisition task clear. In particular, every acquisition theory for linking theories must include a specification of the thematic roles and syntactic positions in the system. These roles and positions then jointly contribute to a hypothesis space of potential links between roles and positions. Every acquisition theory must also include a bias to attend to links between roles and positions (i.e., the need to solve the linking problem must already be present in the child).

Moreover, every acquisition theory involving derivation of the linking theory from the child's input data (i.e., a derivational acquisition theory) must additionally include a procedure for generating explicit linking hypotheses to evaluate. We note that these hypotheses could be about basic links like AGENT $\rightarrow$ *subject* or the complex 3-link patterns that UTAH and rUTAH ultimately specify (see section 2.4 for more discussion). Similarly, every derivational acquisition theory must specify a procedure for evaluating those hypotheses relative to the data. So, though concrete derivational acquisition theories may make different choices in the details of each component (different

---

[3]As with the UTAH linking principle, we assume the rUTAH links are bidirectional.

roles, different positions, different hypothesis generation or evaluation procedures), the overall complexity of a derivational acquisition theory, in terms of the number of components, is unlikely to vary too much.

Beyond clarifying the acquisition problem, UTAH and rUTAH have a number of properties that should make them of interest to researchers working in either innate or derived knowledge frameworks. First, UTAH and rUTAH both involve complex linking patterns that are typically claimed to be innate; so, UTAH and rUTAH both offer the opportunity to explore how complex linking patterns could be learned from more general mechanisms, which is of interest to theorists exploring both types of approaches to solving the linking problem.

Second, UTAH and rUTAH both already include simplifying assumptions about the number of thematic roles and syntactic positions. These simplifying assumptions are cognitively plausible from the perspective of language development; this is because, in order to map thematic roles onto syntactic positions, children are likely to either (i) map the wide variety of fine-grained thematic roles in language to a small number of macroroles (as in UTAH), or (ii) view some roles as more salient than others, and order roles accordingly (as in rUTAH). Therefore, UTAH and rUTAH are likely to reveal information that is relevant to theorists exploring systems of differing complexity.

Third, UTAH and rUTAH are fairly far apart in the linking theory hypothesis space to the extent that fixed systems and relative systems are categorical opposites, and to the extent that UTAH and rUTAH are pure instantiations of these systems. That is, UTAH is fixed for all thematic roles and all syntactic positions; rUTAH is relative for all thematic roles and all syntactic positions. They thus define two poles in the hypothesis space, and will likely reveal information that is relevant to theorists exploring fixed systems, theorists exploring relative systems, and potentially even theorists exploring hybrid systems. In summary, we believe that UTAH and rUTAH are excellent case studies because they're well-specified in the literature, cognitively plausible, and likely to return useful information about whether the relevant linking knowledge can be derived from the input using a fixed vs. a relative system approach.

2.4 Linking theory options and acquisition implications

2.4.1 One 3-link theory means two acquisition stages

The linking principles described above for UTAH and rUTAH assume one 3-link theory. That is, this linking theory is a unit applicable to the verbs of the language, and it's made up of three individual links. So, if a verb doesn't follow any of the three links, it doesn't obey the linking theory.

To derive this linking theory, it seems that a child would need to go through two stages. First, she needs to derive the three individual links comprising the theory; second, she needs to derive the linking theory as a unit. How might children accomplish these two steps? One way is for children to assess the reliability of the linking hypotheses under consideration. Linking hypotheses that are reliable enough are maintained; linking hypotheses that aren't are discarded. In the first stage, a child could assess the reliability of individual links for the verbs of her language. If all goes well, this process would yield the appropriate three individual links out of all the logically possible ones; these individual links could then be composed into a single 3-link theory. In the second stage, a

child could then assess the reliability of the 3-link theory for the verbs of her language.

2.4.2 Three 1-link theories means one acquisition stage

An alternative approach is to simply have three 1-link linking theories. That is, the individual links are themselves separable linking theories that apply to all the verbs of the language. So, for example, a verb might follow two individual links (e.g., AGENT ↔ *subject*, OTHER ↔ *oblique*) while not following a third (PATIENT ↔ *object*); in this case, the verb would obey two of the 1-link theories but not the third one.

   Because the goal is three 1-link theories, there's no further need to create a more complex unit comprised of these individual links. So, to derive these 1-link linking theories, a child would need to go through only a single stage. She derives the three individual linking theories (i.e., the individual links), and she's finished. More specifically, she assesses the reliability of individual links against the verbs of her language. If all goes well, this process yields the three individual links of the appropriate 1-link linking theories. So, deriving three 1-link linking theories involves a simpler acquisition process than deriving one 3-link linking theory.

3   Acquisition components for deriving linking theories

The linking theory variant – whether one 3-link theory or three 1-link theories – serves as the target state for acquisition. This is the knowledge the child is attempting to derive from her input. To concretely investigate the acquisition process leading to that target state, we need to additionally specify other key acquisition components (Pearl & Sprouse, 2015; ?, ?; Pearl, in press): (i) the **initial state**, comprising the knowledge, biases, and abilities the child begins with, (ii) the **acquisitional intake**, representing the data the child uses for updating her hypotheses, and (iii) the **inference** process, representing how the child updates her hypotheses on the basis of that acquisitional intake. We describe each in turn below, as they relate to deriving linking theories from an English child's input.

3.1   Initial state

As mentioned in section 2.3, a child trying to derive linking theories must already have a bias to attend to links between thematic roles and syntactic positions. Additionally, the child must begin already with certain knowledge: knowledge of the relevant thematic representations (fixed macroroles for UTAH or relativized roles for rUTAH), and knowledge of relevant syntactic positions (fixed syntactic positions for UTAH or relativized positions for rUTAH). With this in mind, the child can then define a hypothesis space of possible linking theories that connects those thematic representations (involving 3 roles) to those 3 syntactic positions.

   Here, we further constrain the child's hypothesis space by allowing the child to have the following knowledge in her initial state:

- Links in the hypothesis space are unidirectional, and either go from role to position (e.g., AGENT → *subject*) or from position to role (e.g., *subject* → AGENT).

- Roles and positions can only participate in one link at a time. So, for example, AGENT and PATIENT can't both map to *subject* via the same unidirectional link, such as AGENT OR PATIENT → *subject*. Relatedly, *subject* and *object* can't both map to AGENT via the same unidirectional link, such as *subject or object* → AGENT.

These constraints lead to 18 individual links in the child's hypothesis space (3 positions x 3 roles x 2 directions). Children may then form 1-link or 3-link linking theories from these individual links.

In addition to the bias to attend to links and this knowledge defining the link types available, children also need whatever cognitive abilities are required to deploy this knowledge in real time, extract relevant information from their input, and perform inference over that information to update their linking theory hypotheses. The exact nature of these cognitive abilities is outside the scope of this paper, but we note the necessity of these abilities for the acquisition process described here.

3.2 Data intake

We estimate English children's input from the child-directed speech data in the CHILDES Treebank (Pearl & Sprouse, 2013, 2019), summarized in Table 1. This dataset contains realistic samples of speech directed at American English children between one and five years old, annotated with linguistic and non-linguistic information. The portion of the CHILDES Treebank we used for this investigation involved ≈140K child-directed speech utterances from the BrownEve, BrownAdam, and Valian corpora (Brown, 1973; Valian, 1991) annotated with phrase structure information, animacy information, and the 13 mid-level thematic roles discussed in section 2. We divided these ≈140K utterances into age ranges based on the age of the child the speech was directed at: less than 3 years old (<3yrs), less than 4 years old (<4yrs), and less than 5 years old (<5yrs). We then constructed datasets representing the input to a child of a particular age.[4] We note that the datasets used as input for 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 that older children would learn from all the data they've heard up until that point.

To derive linking theories, the relevant information for a verb's use are the thematic roles present and which syntactic position each role appears in. Therefore, we allow the modeled child's acquisitional intake to be the syntactic information corresponding to syntactic positions assumed by UTAH or rUTAH (*subject*, *first-syn*, etc.), the thematic information corresponding to the thematic roles assumed by UTAH and rUTAH (AGENT, FIRST, etc), and the syntactic positions the thematic roles appear in for each verb use, such as in (3).

---

[4] We extracted the verb lemmas by using python's WordNetLemmatizer package. The extracted lemmas were then manually checked by the first author, and child register verbs (e.g., *squoosh, squooshed, squooshing*) were resolved.

Table 1: Child-directed speech data to three-year-old, four-year-old, and five-year-old English children. 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 with 5 or more link instances in the dataset.

Dataset	Sources	children	ages	utterances	words	verbs	verbs >5
<3yrs	BrownEve, Valian	22	1;6-2;8	≈39.8K	≈197K	555	231
<4yrs	BrownEve, Valian, BrownAdam3to4	23	1;6-4;0	≈50.7K	≈254K	617	260
<5yrs	BrownEve, Valian BrownAdam3to4 BrownAdam4up	23	1;6-4;10	≈56.5K	≈285K	651	275

(3)   Examples of acquisitional intake

    a.   "The little girl kissed the kitten on the stairs."

        (i)   UTAH: *subject*=AGENT, *object*=PATIENT, *oblique*=OTHER

        (ii)   rUTAH: *first-syn*=FIRST, *second-syn*=SECOND, *third-syn*=THIRD

    b.   "The water is falling"

        (i)   UTAH: *subject*=PATIENT

        (ii)   rUTAH: *first-syn*=FIRST

To minimize data sparseness problems with respect to assessing link reliability, we restrict our analyses – and therefore the child's acquisitional intake – to verbs that occur with at least 5 argument uses in the corpus. For example, consider a verb occurring in 2 utterances, one utterance with arguments in *subject* and *object* position (*She kissed the kitten*), and one utterance with arguments in *subject*, *object*, and *oblique object* position (*She kissed the penguin at the zoo*). This would yield 5 (2 + 3) total arguments across all utterances for this verb, and so this verb would be included in our analysis. Since each occurrence of an argument yields evidence for a link, we refer to an argument use of this kind as a "link instance" and we only include verbs with 5 or more link instances in our analyses.[5]

From this corpus sample, we extrapolate the input that children of these ages encounter – in

---

[5]An anonymous reviewer notes that this means infrequent intransitive verbs will be more likely to be left out of our abridged corpus sample, as they'll have fewer argument uses than transitive verbs. This is true: 47.8-50.3% of transitive verbs are excluded, while 64.7-68.7% of intransitive verbs are excluded across our three datasets. However, we note that the total number of excluded transitive verbs is higher than the total number of excluded intransitive verbs: 243-271 transitive vs. 45-57 intransitive verbs, across our three datasets. Future work can draw on larger corpora where fewer verbs will need to be excluded due to low frequency.

particular the quantity and distribution of verb link instances – using the procedure detailed in Appendix A. This procedure draws on Hart and Risley (1995, 2003) to estimate the amount of input children hear per hour, and Davis, Parker, and Montgomery (2004) to estimate how many hours per day children of different ages are awake. With this estimated child input, we can then apply the Tolerance Principle to assess different linking hypotheses children could consider.

3.3 Inference

3.3.1 The Tolerance Principle as a decision criterion for reliability

Recall that the acquisition process we posited in section 2.4 involves the child assessing the reliability of different linking hypotheses on the basis of her input. One way to concretely model this process is to use the Tolerance Principle (Yang, 2005, 2016). The Tolerance Principle is a formal approach for determining when a child would choose to adopt a "rule", generalization, or default pattern to account for a set of items. This principle is based on cognitive considerations of knowledge storage and retrieval in real time, incorporating how frequently individual items occur, the absolute ranking of items by frequency, and serial memory access. Importantly for our purposes, this principle is designed precisely for data where there are exceptions to the potential rule; the principle then determines how many exceptions a rule can "tolerate" in the data before it's not worthwhile to have that rule at all. Therefore, this principle provides a precise threshold for the number of tolerable exceptions to a hypothesized rule or generalization.

The intuition behind the Tolerance Principle is that the child is optimizing retrieval time. More specifically, suppose a child is considering a rule that connects an item to some other information, such as a root connecting to its past tense form (Yang, 2005, 2016), a word connecting to its metrical stress pattern (Legate & Yang, 2013; Pearl, Ho, & Detrano, 2017), or thematic roles connecting to their syntactic positions (what we implement here). The potential rule compactly encodes some regularity – this is the pattern that several items in the dataset under consideration follow (e.g., default past tense morphology, a default stress pattern, or a default linking pattern).

When does it become useful to have a rule? One answer is that it's useful when having a rule makes the average retrieval time for any item in the dataset faster (Yang, 2005, 2016). That is, it's useful to have a past tense rule in order to retrieve a regular past tense form, it's useful to have a metrical stress rule to retrieve a predictable metrical stress pattern, and it's useful to have a linking rule (i.e., a linking theory) to retrieve a thematic role reliably associated with a syntactic position or a syntactic position reliably associated with a thematic role. However, if the past tense is too irregular, the metrical stress is too unpredictable, or the link is too unreliable, it's not useful to have the rule: retrieving the target information takes too long on average. Therefore, the child's decision about whether a rule should be adopted (i.e., past tense morphology viewed as regular, a metrical stress pattern viewed as predictable, or a link viewed as reliable) is based on its tolerance threshold – a rule can only tolerate so many exceptions before the child decides not to bother with it. Yang (2005, 2016) specifies this tolerance threshold by considering how long it would take to access an item's target information with vs. without the rule. The retrieval process is assumed to involve serial search, which accords with current psycholinguistic data reviewed by Yang (2005, 2016).

The threshold for adopting the rule is determined by a fairly complex equation (see Yang 2005, 2016), but is well approximated by the much simpler equation $\frac{N}{ln(N)}$, where N is the number of items the rule could potentially apply to. That is, if there are $\frac{N}{ln(N)}$ or fewer exceptions in the set of items the rule could apply to, it's useful in terms of retrieval time to adopt the rule. In other words, the Tolerance Principle requires a certain number of items that match a rule in order for that rule to be adopted as useful: that number is $N - \frac{N}{ln(N)}$. If there aren't that many items that match the rule, the rule isn't useful because adopting the rule slows down the average retrieval time.

Interestingly, this means that a potential rule needs to apply to a "super-majority" of items in order to be adopted. For example, a rule that could apply to 100 items allows only 21 exceptions (21%), and thus requires 79 items that match the rule under consideration. A rule that could apply to 1000 items allows only 144 exceptions (14.4%), and therefore requires 856 items that match the rule under consideration. A rule that could apply to 10000 items allows only 1085 exceptions (10.85%), and therefore requires 8915 items that match the rule under consideration. This has the practical effect of allowing only one option to be the rule (i.e., this disallows two or more "rules" for a set of items); this is because, by definition, only one option can ever hold a majority – let alone a super-majority.

So, to summarize, the Tolerance Principle provides a formal threshold for adopting a rule, i.e., when a child would choose to view a certain pattern as dominant or reliable for a set of items and therefore its default pattern. The Tolerance Principle has been used for investigating the acquisition of a default pattern or rule for a variety of linguistic knowledge types, including English past tense morphology (Yang, 2005, 2016), English noun pluralization (Yang, 2016), German noun pluralization (Yang, 2005), English nominalization (Yang, 2016), English metrical stress (Legate & Yang, 2013; Yang, 2015; Pearl et al., 2017), English *a*-adjective morphosyntax (Yang, 2015, 2016), English dative alternations (Yang, 2016), and noun morphology in an artificial language (Schuler, Yang, & Newport, 2016). We will use it here for evaluating linking hypotheses at both the individual link level (1-link theories and the first stage of 3-link theories) and the multi-link level (the second stage of 3-link theories).[6]

3.3.2 Evaluating possible linking theories with the Tolerance Principle

Evaluating linking theories over all verbs.   Linking theories are meant to be generalizations about how the verbs of the language behave. So, a linking theory, whether a 1-link theory or a 3-link theory, can be evaluated by how many verb types obey it. If enough verb types obey the linking theory, a child using the Tolerance Principle would decide the linking theory is reliable enough for verbs of the language. Here, we can apply this to the English verb types in English children's input (from Table 1: <3yrs = 231, <4yrs = 260, <5yrs = 275). The Tolerance Principle provides the threshold ($N - \frac{N}{ln(N)}$) for how many verbs must obey the linking theory: <3yrs = 183, <4yrs = 209, <5yrs = 220. If at least that many individual verbs obey the linking theory, a child of that age could successfully derive the linking theory for English verbs from her input by using the Tolerance Principle.

---

[6]The evaluation process described in the next section can be run using the code and input sets available at https://github.com/lisapearl/linking-problem-code.

If the target state is three 1-link linking theories, the child goes through this evaluation once for 1-link patterns. If instead the target state is one 3-link linking theory, the child goes through this evaluation once for 1-link patterns, composes reliable 1-link patterns into possible 3-link patterns, and then goes through this evaluation again for the generated 3-link patterns.

Evaluating individual verbs. How then can a child evaluate whether an individual verb obeys a particular linking theory? We first describe the evaluation process for the 1-link patterns comprising 1-link theories and the individual links of 3-link theories; we then describe the evaluation process for the 3-link patterns comprising 3-link theories.

Table 2 demonstrates the evaluation process for 1-link patterns when assessing individual verbs. If the link goes from thematic role to syntactic position (e.g., AGENT $\rightarrow$ *subject*, as in the top part of Table 2), we consider all the links that link from that thematic role (i.e., we also include AGENT $\rightarrow$ *object* and AGENT $\rightarrow$ *oblique*). N is the set of link instances involving that thematic role (e.g., AGENT), and we count how many obey the link under consideration (AGENT $\rightarrow$ *subject*); if this number is greater than or equal to $N - \frac{N}{ln(N)}$, the link under consideration is reliable. In Table 2, we see that this leads to a three-year-old child considering the AGENT $\rightarrow$ *subject* link reliable for *use* and *break*, but not for *belong*.

Similarly, if the link goes from syntactic position to thematic role (e.g., *subject* $\rightarrow$ AGENT, as in the bottom part of Table 2), we consider all the links that link from that syntactic position (i.e., we also include *subject* $\rightarrow$ PATIENT and *subject* $\rightarrow$ OTHER). N is the set of link instances involving that syntactic position (e.g., *subject*), and we count how many obey the link under consideration (*subject* $\rightarrow$ AGENT); if this number is greater than or equal to $N - \frac{N}{ln(N)}$, the link under consideration is reliable. In Table 2, we see that this leads to a three-year-old child considering the *subject* $\rightarrow$ AGENT link reliable for *use*, but not for *break* and *belong*.

Table 2: Link instances across the syntactic positions of *subject*, *object* and *oblique object* for the different thematic representations of UTAH, given example verbs and link instance counts from the <3yrs corpus. Instances compatible with the link under consideration are **bolded**. Tolerance Principle analysis is shown, involving the total link instances N, the number of instances that obey the link under consideration, and the tolerance threshold $N - \frac{N}{ln(N)}$. If the instances obeying the link are greater than or equal to the tolerance threshold, the link is perceived as reliable.

	use			*break*			*belong*		
AGENT $\rightarrow$ *subj*	*subj*	*obj*	*obliq*	*subj*	*obj*	*obliq*	*subj*	*obj*	*obliq*
AGENT instances	**46**	0	0	**24**	0	0	**2**	0	12
TolP	N	*# obey*	*thresh*	N	*# obey*	*thresh*	N	*# obey*	*thresh*
	46	46	33	24	24	16	14	2	8
Reliable?	Yes			Yes			No		
subj $\rightarrow$ AGENT	AGENT	PATIENT	OTHER	AGENT	PATIENT	OTHER	AGENT	PATIENT	OTHER
subj instances	**46**	0	0	**24**	46	0	**2**	25	0
TolP	N	*# obey*	*thresh*	N	*# obey*	*thresh*	N	*# obey*	*thresh*
	46	46	33	70	24	53	27	2	18
Reliable?	Yes			No			No		

Table 3 demonstrates the evaluation process for 3-link patterns when assessing individual verbs. Here, all link instances that obey the 3-link pattern are collapsed together, as are all the link instances that don't. That is, with a 3-link pattern in hand, the fine-grained details of where a specific thematic role (e.g., AGENT) appears don't matter. Instead, the child is considering whether each link instance (e.g., AGENT or PATIENT or OTHER) appears where it's expected to according to the 3-link theory (e.g., *subject* or *object* or *oblique*). N is then the entire set of link instances for the verb, and we count how many obey the 3-link pattern; if this number is greater than or equal to $N - \frac{N}{ln(N)}$, the 3-link pattern under consideration is reliable. In Table 3, we see that this leads to a three-year-old child considering UTAH's 3-link pattern reliable for *use*, but not for *break* and *belong*.

Table 3: Link instances across the syntactic positions of *subject*, *object* and *oblique object* for the different thematic representations of UTAH, given example verbs and link instance counts from the <3yrs corpus. Instances compatible with UTAH's 3-link pattern are **bolded**. Tolerance Principle analysis is shown, involving total link instances N, the number of instances that obey that 3-link pattern, and the tolerance threshold $N - \frac{N}{ln(N)}$. If the instances obeying the 3-link pattern are greater than or equal to the tolerance threshold, the 3-link pattern is perceived as reliable.

	use			*break*			*belong*		
	subj	*obj*	*obliq*	*subj*	*obj*	*obliq*	*subj*	*obj*	*obliq*
AGENT	**46**	0	0	**24**	0	0	**2**	0	12
PATIENT	0	**57**	1	46	**31**	0	25	**0**	0
OTHER	0	1	**4**	0	0	**1**	0	0	**12**
TolP	N	*# obey*	*thresh*	N	*# obey*	*thresh*	N	*# obey*	*thresh*
	109	107	86	102	56	80	51	14	39
Reliable?	Yes			No			No		

4   Results

4.1   1-link patterns: UTAH vs. rUTAH

Tables 4 and 5 show Tolerance Principle analysis for individual links over the verb types in English children's input. We can see that the results are identical for all three ages, with the reliable links summarized in Table 6. For the UTAH-based links, two of the three appropriate role-to-position links are reliable and one of the appropriate position-to-role links is reliable. This contrasts with the rUTAH-based links, where all three appropriate role-to-position links are reliable and all three appropriate position-to-role links are reliable.

What does this mean for a child trying to derive either three 1-link theories or generate one 3-link pattern for further evaluation? At least four options seem plausible. The first and most conservative option is that the bidirectional links of a linking theory can only be successfully derived if both corresponding unidirectional links were perceived as reliable (e.g., AGENT ↔ *subject* is

Table 4: Tolerance Principle analysis of whether the individual links that use UTAH thematic representations and UTAH syntactic positions are reliable for English children's verbs at different ages. Verbs with 5 or more relevant link instances in the corpus are included in the analysis. The total number of verbs involving the link under consideration is shown (N), along with the number of verbs that obey the individual link and the tolerance threshold ($N - \frac{N}{ln(N)}$). If the link is perceived as reliable for that age, the row is **bolded**.

age	role to position			N	# obey	thresh	position to role			N	# obey	thresh
<3	**AGENT**	$\rightarrow$	***subject***	**217**	**215**	**177**	*subject*	$\rightarrow$	AGENT	230	103	188
	PATIENT	$\rightarrow$	*object*	230	96	188	***object***	$\rightarrow$	**PATIENT**	**206**	**197**	**168**
	OTHER	$\rightarrow$	***oblique***	**162**	**147**	**131**	*oblique*	$\rightarrow$	OTHER	171	128	138
<4	**AGENT**	$\rightarrow$	***subject***	**245**	**242**	**201**	*subject*	$\rightarrow$	AGENT	259	123	213
	PATIENT	$\rightarrow$	*object*	259	110	213	***object***	$\rightarrow$	**PATIENT**	**238**	**224**	**195**
	OTHER	$\rightarrow$	***oblique***	**186**	**171**	**151**	*oblique*	$\rightarrow$	OTHER	200	145	163
<5	**AGENT**	$\rightarrow$	***subject***	**254**	**252**	**209**	*subject*	$\rightarrow$	AGENT	271	130	223
	PATIENT	$\rightarrow$	*object*	271	115	223	***object***	$\rightarrow$	**PATIENT**	**248**	**234**	**204**
	OTHER	$\rightarrow$	***oblique***	**197**	**179**	**160**	*oblique*	$\rightarrow$	OTHER	211	151	172

Table 5: Tolerance Principle analysis of whether the individual links that use rUTAH thematic representations and rUTAH syntactic positions are reliable for English children's verbs at different ages. Verbs with 5 or more relevant link instances in the corpus are included in the analysis. The total number of verbs involving the link under consideration is shown (N), along with the number of verbs that obey the individual link and the tolerance threshold ($N - \frac{N}{ln(N)}$). We note that all links are above the tolerance threshold and so perceived as reliable.

age	role to position			N	# obey	thresh	position to role			N	# obey	thresh
<3	FIRST	$\rightarrow$	*first-syn*	230	226	188	*first-syn*	$\rightarrow$	FIRST	230	227	188
	SECOND	$\rightarrow$	*second-syn*	207	198	169	*second-syn*	$\rightarrow$	SECOND	208	197	170
	THIRD	$\rightarrow$	*third-syn*	170	160	137	*third-syn*	$\rightarrow$	THIRD	170	162	137
<4	FIRST	$\rightarrow$	*first-syn*	259	254	213	*first-syn*	$\rightarrow$	FIRST	259	254	213
	SECOND	$\rightarrow$	*second-syn*	239	232	196	*second-syn*	$\rightarrow$	SECOND	239	231	196
	THIRD	$\rightarrow$	*third-syn*	200	188	163	*third-syn*	$\rightarrow$	THIRD	200	191	163
<5	FIRST	$\rightarrow$	*first-syn*	271	266	223	*first-syn*	$\rightarrow$	FIRST	271	266	223
	SECOND	$\rightarrow$	*second-syn*	248	240	204	*second-syn*	$\rightarrow$	SECOND	248	239	204
	THIRD	$\rightarrow$	*third-syn*	211	197	172	*third-syn*	$\rightarrow$	THIRD	211	200	172

derived only if both AGENT $\rightarrow$ *subject* and *subject* $\rightarrow$ AGENT were reliable). Under this approach of linking theory derivation, only rUTAH would enable the child to derive the appropriate three 1-link theories and generate the appropriate 3-link pattern for further evaluation. This is because UTAH only ever has reliable links in one direction for each thematic role and syntactic position, and so the child couldn't derive the appropriate 1-link theories or generate the appropriate 3-link pattern for UTAH.

The second option builds on the intuition that linking theories are expectations about the positional preferences of thematic roles (i.e., links from roles to positions), rather than the thematic

Table 6: The individual links that would be perceived by an English child using the Tolerance Principle as reliable, considering the verb types for the <3yrs, <4yrs, and <5yrs child-directed speech data.

system	role to position links			position to role links		
UTAH	AGENT	→	*subject*			
				object	→	PATIENT
	OTHER	→	*oblique*			
rUTAH	FIRST	→	*first-syn*	*first-syn*	→	FIRST
	SECOND	→	*second-syn*	*second-syn*	→	SECOND
	THIRD	→	*third-syn*	*third-syn*	→	THIRD

role preferences of syntactic positions (i.e., links from positions to roles). Under this view, a child can derive the appropriate theory or derive the appropriate multi-link pattern if there's a unidirectional link from the appropriate thematic role to the appropriate syntactic position. This leads to the same results as before: UTAH has only two of the three appropriate unidirectional links from role to position, while rUTAH has all three. So, as before, a child couldn't derive all three UTAH 1-link theories or the 3-link UTAH pattern for further evaluation. In contrast, she could do this for the three rUTAH 1-link theories and the 3-link rUTAH pattern.

The third option takes the opposite view, and builds on the intuition that linking theories are expectations about the the thematic role preferences of syntactic positions (i.e., links from positions to roles). Under this view, a child can derive the appropriate theory or derive the appropriate multi-link pattern if there's a unidirectional link from the appropriate syntactic position to the appropriate thematic role. This again leads to the same results as before: UTAH has only one of the three appropriate unidirectional links from position to role, while rUTAH has all three. So, as before, a child couldn't derive all three UTAH 1-link theories or the 3-link UTAH pattern for further evaluation. In contrast, she could do this for the three rUTAH 1-link theories and the 3-link rUTAH pattern.

The fourth and most liberal option is that the child considers any reliable unidirectional link between a thematic role and a syntactic position sufficient for deriving the corresponding 1-link theory or including the appropriate link in a 3-link pattern for further evaluation. Using this option, a child would be able to derive all three appropriate 1-link theories for both UTAH and rUTAH, and generate appropriate 3-link UTAH and rUTAH patterns for further evaluation. This is because at least one unidirectional link is viewed as reliable for each connection between a UTAH thematic role and UTAH syntactic position (i.e., AGENT ↔ *subject* has AGENT → *subject*; PATIENT ↔ *object* has *object* → PATIENT; OTHER ↔ *oblique* has OTHER → *oblique*). And, as before, rUTAH has reliable unidirectional links in both directions for all appropriate links.

While we don't know for certain how children derive 1-link theories or the individual links that serve as building blocks for 3-link patterns, working through the plausible options highlights the learning assumptions needed to support derivation of different linking theory representations. In particular, if UTAH's linking theory is the correct target knowledge, then there's only one way a

child could derive either the three 1-link theories or the building blocks for the 3-link pattern: she must use the relatively liberal option where only a single unidirectional link in either direction (role to position or position to role) is required to be reliable. In contrast, if rUTAH's linking theory is the correct target knowledge, a child could derive the three 1-link theories or the building blocks for the 3-link pattern in a number of ways – that is, the acquisition process for deriving rUTAH is more robust. This is presumably because rUTAH redefines many exceptions to the UTAH primary linking pattern as paradigmatic cases of the rUTAH primary linking pattern, and so it's easier for a child to view the rUTAH links as reliable from the input.

4.2   One 3-link theory: UTAH vs. rUTAH

But what if we believe the target state is one 3-link theory? The previous analysis suggests that a child trying to derive UTAH may have a harder time generating the appropriate 3-link pattern, compared with a child trying to derive rUTAH. However, let's suppose that the child has successfully generated the appropriate 3-link pattern for UTAH or rUTAH, and now needs to evaluate whether the 3-link pattern is reliable enough to derive the corresponding 3-link theory.

Recall from the evaluation process for 3-link patterns discussed in section 3.3.2 that Tolerance Principle analysis is done through the filter of the 3-link pattern: a link instance either obeys the 3-link pattern or it doesn't. This brings up the question how to count instances of a 3-link pattern. Consider this use of the verb *pet*: *Lily pets the kitties*. If each link is considered an instance of the 3-link pattern, this use counts as two instances that obey the UTAH 3-link pattern: one for AGENT ↔ *subject* and one for PATIENT ↔ *object*. In contrast, if each verb use is considered an instance of the 3-link pattern, this use counts as a single instance that obeys the UTAH pattern, as all thematic roles are in their expected positions. Because it's unclear *a priori* which one a child would select, we show the results of both approaches (link-based and verb-use-based) to counting 3-link linking instances in Table 7 below.

Table 7: Tolerance Principle analysis of whether the 3-link patterns for UTAH and rUTAH are perceived as reliable from the verb usage in English children's input at different ages. The total number of verbs with 5 or more link-based instances or verb-use-based instances in the corpus is shown, along with the number of verbs required to meet the tolerance threshold and the number of verbs obeying the appropriate 3-link pattern for both UTAH and rUTAH. If the 3-link pattern is viewed as reliable for that age, the number obeying the 3-link pattern is **bolded**.

	age	total (N)	thresh	UTAH # obey	rUTAH # obey
	<3yrs	231	189	126	**229**
link-based	<4yrs	260	214	136	**253**
	<5yrs	275	227	142	**264**
	<3yrs	224	183	97	**220**
verb-use-based	<4yrs	255	209	108	**248**
	<5yrs	267	220	114	**255**

This analysis again highlights the impact that the choice of UTAH or rUTAH can have. Put simply, a child trying to derive one 3-link theory for UTAH won't be able to infer that the 3-link pattern is in fact reliable for all English verbs, no matter what age the child is and no matter how linking pattern instances are counted (by individual link or by verb use). The child's input before age three, four, and five never surpasses the required number of pattern-matching verbs to support this inference. In contrast, a child trying to derive one 3-link theory for rUTAH will always be able to infer that the 3-link pattern is reliable and therefore this pattern is good to have as 3-link linking theory. This is because there are very few exceptional verbs when viewing the input through a rUTAH lens.

As mentioned, the low number of exceptions to rUTAH is expected, given that rUTAH generally minimizes the number of exceptions to the predicted 3-link linking pattern. But what is novel, at least to our knowledge, is the discovery that the 3-link pattern of UTAH is not derivable from child-directed speech because of the number of verbs that are exceptions on the surface.[7] We discuss the implications of these findings for both language acquisition and syntactic theory next.

5   Discussion

We investigated how the linking theories specified by UTAH and rUTAH could be derived from English children's input, thereby providing an existence proof for derived approaches to linking theory development. More specifically, we specified an acquisition theory relying on a combination of linguistic knowledge and general-purpose learning mechanisms, evaluated it with respect to realistic samples of children's input, and discovered notable acquisition differences between UTAH and rUTAH. We first discuss the different components of the proposed acquisition theory in more detail, speculating on their domain-specificity and innateness. We then turn to the implications of our findings for syntactic theory, as well as possible future investigations.

5.1   The components necessary for deriving linking theories

One major conclusion from our investigation is that there are a relatively large number of components necessary for deriving linking theories, whether they are a set of 1-link theories or a single 3-link theory. Based on the acquisition theory specified here, we propose that any theory of how children derive linking theories will require the components in Table 8.

For each of these components, we can ask whether they are likely to be domain-specific (to language) or domain-general, and whether they are likely to be innately-specified or derived during the acquisition process. Our goal is to determine if any of the components are likely to be simultaneously language-specific and innate, as this component type figures most prominently in the debates between innate and derived approaches to linking theories (as well as many other aspects of language acquisition).

---

[7]This can, of course, be mediated through various mechanisms, such as innate knowledge of the complete complex theory that UTAH encodes, or prior knowledge of (movement-based) derivations that turn these exceptions into paradigmatic cases of the predicted pattern.

Table 8: Proposed components that are required to derive linking theories, along with their likely categorization according to current knowledge. Components that might currently be considered domain-specific and innate are **bolded**.

component	domain-general or specific	derived or innate
thematic roles	domain-general	innate
syntactic positions	domain-specific	derived
a bias to look for links	**domain-specific**	**either**
a bias to look for 1-to-1 links	**either**	**either**
a bias for a single reliable link	domain-specific	derived
the ability to track links	domain-general	innate
a procedure to generate linking patterns	**either**	**innate**
a procedure to evaluate linking patterns	domain-general	innate

Thematic roles are based on non-linguistic concepts of event participants. Because of this, they are likely to be domain-general (though they may contribute to language differently than other cognitive domains) and innate. Syntactic positions, in contrast, are likely domain-specific and, at least in their final form, derived from prior language experience. We note that we remain agnostic as to whether innate, domain-specific knowledge is required to derive these syntactic positions.

The bias to look for links between roles and positions appears to be domain-specific, as we know of no equivalent in other domains. One possibility is that this bias is simply innate. Another possibility is that this bias is a specific instantiation of a more general bias to look for correlations between active representations in any single cognitive domain (e.g., active representations in the visual domain, the spatial domain, the social cognition domain, the language domain, etc.). The question then is how to formulate that bias in such a way as to yield the links we want (e.g., between thematic roles and syntactic positions), while not yielding links that we don't want (i.e., between thematic roles and anything else active during language processing). Such fine-tuning likely requires innate knowledge, though the status of that innate knowledge (i.e., whether it's domain-specific or domain-general) is currently unknown.

In the acquisition approach proposed here, there was a bias to look for a 1-to-1 mapping between roles and positions. That is, the child only considered links that involved a single syntactic position or thematic role (e.g., AGENT → *subject*), rather than allowing disjunctive options that involve multiple syntactic positions or multiple thematic roles (e.g., AGENT or PATIENT → *subject*; PATIENT → *subject* or *object*). This constraint might be thought of as similar to the mutual exclusivity bias young children often show during early word learning (Markman & Wachtel, 1988; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992), where they assume each word refers to a distinct referent. It remains an open question what the origins of this 1-to-1 bias are (Clark, 1988; Markman & Wachtel, 1988; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, Wasow, & Hansen, 2003; Frank, Goodman, & Tenenbaum, 2009), in particular whether they are innate and/or language-specific. Therefore we list this bias as "either" with respect to both domain-specific vs. domain-general and derived vs. innate.

Another necessary component is that the child must assume there is only a single reliable link per role or position (i.e., AGENT is linked to only one of the available options, *subject* is linked to only one of the available options, etc.). This domain-specific knowledge derives directly from the Tolerance Principle (as discussed in section 3.3.1), which we take to be domain-general because of its reliance on item storage and retrieval, irrespective of what cognitive domain the item comes from. The Tolerance Principle is likely innate, as it's unclear how a child would learn to optimize item retrieval with respect to item access time.

The ability to track links is likely derived from the innate domain-general ability to track frequencies (Saffran, Aslin, & Newport, 1996; Xu & Tenenbaum, 2007; Smith & Yu, 2008; Denison, Reed, & Xu, 2011; Denison, Bonawitz, Gopnik, & Griffiths, 2013; Stahl, Romberg, Roseberry, Golinkoff, & Hirsh-Pasek, 2014; Yurovsky, Case, & Frank, 2017). This ability is then applied to linking patterns as cognitive objects.

The procedures that we postulated to generate linking patterns were all domain-specific because they only apply to linking patterns. However, it's currently unknown if general-purpose mechanisms of explicit hypothesis generation (see Perfors, 2012 for discussion) would suffice to generate a set of reasonable linking patterns. If so, these procedures could be an example of a domain-general procedure applied to the domain-specific hypothesis space of linking theories. Therefore we list it as "either". At our current level of understanding, the hypothesis generation procedures would also likely need to be innate, as it's unclear how to break this sort of hypothesis generation down into learnable components.

Finally, our evaluation procedure was the Tolerance Principle, which as we previously noted, is likely domain-general and innate.

Taken together, there are three components that are potentially both domain-specific and innate, given our current level of understanding: the bias to look for links between roles and positions, the bias to look for 1-to-1 links, and the linking pattern generation procedures. The remaining components are likely to be either domain-general or derived or both. It's always possible that future work may find a way to reduce the number of domain-specific and innate components to zero. For now, our proposal for how to derive linking theories seems to potentially require three.

Given that this component list rests on the specific implementation we propose here for deriving linking theories, it's reasonable to wonder if the overall complexity of the system could have been simplified by making different specific choices for each component (other thematic systems, other syntactic systems, other linking pattern generation and evaluation procedures, etc). There are surely other choices for each of the components in our acquisition theory, but we don't believe that different choices would substantially lessen the complexity of the system. This is because we attempted to choose the simplest available options that are both cognitively plausible and theoretically motivated.

In particular, we tested both fixed and relative thematic systems. We chose thematic systems and syntactic systems that only have three roles. We explored how to generate and evaluate both a set of 1-link theories and a single 3-link theory compatible with current UTAH and rUTAH specifications. In short, it's not obvious how a substantial amount of complexity could be removed from the system – the linking problem seems to simply be a problem with a certain amount of inherent complexity. This is likely why the dominant linking theories in the syntactic literature both

appear to contain the same amount of complexity, but shift that complexity between movement operations in UTAH and a relativized hierarchy of roles and positions in rUTAH.

As mentioned in section 2, we tested UTAH and rUTAH because we believe these are ideal case studies for exploring the bounds of the learning problem associated with the acquisition of linking theories. Therefore, we also believe that the components listed above should extend to the acquisition of any linking theory that can be stated with enough specificity, even if the precise implementation of the acquisition process differs from the one proposed here. If there are linking theories that diverge substantially from the UTAH/rUTAH systems in form or content, the next step to evaluate them with respect to acquisition will be to formulate those theories in enough detail such that we can apply the acquisition approach demonstrated here. This involves the implementation of the components in Table 8 and evaluation on child-directed speech data like those contained in the CHILDES Treebank.

5.2   Consequences for UTAH and rUTAH as adult linking theories

What do our results mean for adult knowledge of the linking theory? We turn first to UTAH, as it's the theory that fared the worst in our acquisition evaluations. Our first finding is that the complex form of UTAH (the single 3-link theory) is difficult to generate as a 3-link pattern because the input that English children receive doesn't support all three links in both directions. Generating it as a 3-link pattern to evaluate requires the child to adopt the most liberal hypothesis generation procedure that we considered: assume a bidirectional link between thematic role and syntactic position if a unidirectional link is reliable enough in either direction.

Our second finding is that, even assuming the 3-link theory version of UTAH can be generated as a 3-link pattern, it isn't derivable from children's input as a language-wide linking theory. This is because the surface forms of so many constructions are apparent exceptions to the 3-link linking theory. This in itself may not be surprising, as the innate form of UTAH leverages the existence of movement in derivational theories of grammar to reanalyze these exceptions as paradigmatic cases of the linking pattern. Crucially, this reanalysis is only available when the 3-link linking theory of UTAH is available *prior* to the acquisition process we explored here. If UTAH is instead derived during the acquisition process, then it's not obvious how the child would know when movement is occurring because no default linking theory is yet available that would require movement in order to match surface forms that deviate from that theory. That is, the child would need some other unequivocal marker that movement has occurred in order to realize that the surface form is not the original form. To the best of our knowledge, there doesn't appear to be an unequivocal marker for movement in these constructions, at least in English. For example, unlike A'-dependencies, A-dependencies don't have an obvious gap site or inversion (e.g., as in subject-auxiliary inversion) to signify movement.

In principle, a second option to make UTAH derivable would be to freely allow movement for any analysis, even without direct evidence of movement. This would allow the child to reanalyze any exception as fitting the hypothesized linking pattern. But, of course, that means all data could then be evidence in favor of all linking theories – whenever a surface pattern doesn't conform to the hypothesized linking theory, a movement reanalysis would allow it to conform. That is, this approach of freely allowing movement would make all observable data ambiguous with respect to

all possible linking theories. It would therefore be surprising that children end up with the same linking theories after acquisition from their input. To combat this acquisition problem, we would need prior constraints on the types of movements available. To the best of our knowledge, these constraints would basically re-instantiate UTAH – just as movement constraints instead of linking patterns (i.e., only certain movements are allowed, and those movements are exactly the ones that UTAH posits to explain deviations).

Based on this, we tentatively conclude that the complex form of UTAH as typically proposed (a single 3-link theory) is unlikely to be derivable, at least given the acquisition approach investigated here. That said, it's potentially derivable as a set of three 1-link theories (AGENT → *subject*, *object* → PATIENT, OTHER → *oblique object*), assuming a reliable unidirectional link in either direction is enough to establish a link between thematic role and syntactic position. We leave it to future work to explore the consequences of decomposing UTAH into the three reliable unidirectional links uncovered here.[8]

In contrast with UTAH, rUTAH – whether as a single 3-link theory or a set of 1-link theories – appears to be easily derivable from English child-directed input. This in itself aligns with our intuitions, given that (i) rUTAH was specifically designed to eliminate exceptions (and to eliminate the need for movement), and (ii) the Tolerance Principle is attuned to the number of exceptions to a potential rule. Taken together, our results suggest that a fixed approach to roles and positions (as in UTAH) is likely to lead to too many exceptions for a child using the Tolerance Principle to derive the linking theory specified by UTAH. The only linking theories that are derivable under the acquisition approach pursued here are those that seek to minimize exceptions to the theory, and rUTAH is one example of this that leverages a relativized approach to roles and positions.

5.3 Future investigations

While we have provided an existence proof for a derived approach to linking theory development using the Tolerance Principle, there of course remain several interesting questions. First, this approach was evaluated over American English input to children. Will it also yield the same results when evaluated over other languages and dialects? Perhaps more interestingly, if there are languages where UTAH and rUTAH don't appear to be empirically adequate (for whatever reason), we could ask whether the approach here fails to derive these linking theories. For instance, if this approach succeeds at deriving rUTAH for languages where rUTAH is empirically adequate and fails at deriving rUTAH for languages where rUTAH is empirically inadequate, that would bolster support for the derived acquisition approach we proposed here.

Another interesting question concerns making the inference process more realistic. Here, we assumed the child could keep track of all the relevant link pattern counts and make a single decision based on the Tolerance Principle at the end. That is, we assumed a batch learning process. However, we know that children's acquisition is incremental, with them processing information as they encounter it. So, it would be useful to see if the Tolerance Principle process we proposed

---

[8]We note that UTAH in its typical syntactic formulation is a tightly integrated system of innate links and an innate movement operation. So, our finding that UTAH is unlikely to be derivable is only relevant when we're interested in exploring the complexity inherent in deriving linking theories.

here would yield the same results if children's input were processed incrementally (e.g., see Wang and Mintz (2008) for translating a batch inference process for syntactic categorization to an incremental inference process). The process itself would be fundamentally the same: the child would track the relevant link pattern counts from her input. So, for example, at one time, the child might have encountered 30 uses of a verb, where 24 of them obey the link pattern under consideration. Later on, she might have encountered 50 more uses, where 35 of them obey the link pattern under consideration – this would lead to 30+50=80 uses total, with 24+35=59 obeying the link pattern under consideration. The child's decision about link pattern reliability could be made as often as she liked – for example, after the first time point in the above example where 30 verb uses had been encountered and also after the second time point where 80 verb uses had been encountered. In the example above, assessing reliability at the first time point would cause the child to think the link is reliable ($30-\frac{30}{ln(30)}$=22, and 24 uses obey the link pattern); however, assessing reliability at the second time point would cause the child to think the link is unreliable ($80-\frac{80}{ln(80)}$=62, and only 59 uses obey the link pattern). So, a child's assessment might well fluctuate over time. In this way, at any point in development, we could see if the modeled child would view a given link pattern as reliable, and so use it to derive a linking theory. This kind of incremental inference would allow us to observe the trajectory of linking theory derivation, rather than a snapshot at the end of three, four, or five years old.[9]

A third interesting avenue is to explore other learning approaches for deriving linking theories from realistic children's input. As mentioned, we provided an existence proof using the Tolerance Principle as the core learning principle, but there are many other learning approaches that may be cognitively plausible (e.g., Bayesian inference (Perfors, Tenenbaum, Griffiths, & Xu, 2011), variational learning (Yang, 2002, 2004, 2012)). Future work can thus see if these other learning approaches are able to derive UTAH and rUTAH from children's input and, if so, whether rUTAH retains its learnability advantage.

6   Conclusion

We built concrete acquisition theories for UTAH and rUTAH that assumed a derived – rather than innate – approach to the development of linking theory knowledge. Our goal was to explore the complexity of the acquisition problem created by linking theories, and to do so using well-specified linking theories that occupy relatively distinct positions within the hypothesis space of possible linking theories. We leveraged a conceptual acquisition framework that specified key aspects of the child's acquisition task: the initial state, data intake, inference mechanism, and target knowledge state. The initial state involved minimal domain-specific, prior knowledge and incorporated cognitively-plausible learning abilities; the data intake was based on linguistically-annotated realistic child-directed input from the CHILDES Treebank (Pearl & Sprouse, 2013, 2019) and

---

[9]We note that this particular incremental approach also requires the child to have some sort of stopping mechanism for assessing links, unless we think children continuously reevaluate the links that underlie linking theories; in that vase, it would only be the reliability of their data that prevents constant fluctuation. A sample stopping mechanism might be something like the child noticing a set amount of time has passed without the link reliability changing (either becoming reliable or ceasing to be reliable).

empirically-based estimates of data quantity; the inference mechanism relied on Yang (2005, 2016)'s cognitively-motivated Tolerance Principle; and the target linking knowledge states were specified by UTAH and rUTAH. Using this framework, we found a striking difference between UTAH and rUTAH. UTAH is difficult to derive from English children's input, whether as a single 3-link theory or a set of three 1-link theories. In contrast, rUTAH is easy to derive as either a single 3-link theory or a set of three 1-link theories. Moreover, these results hold for English children's input at ages three, four, and five – it doesn't matter the age (in this age range) that children attempt to derive these linking theories. This suggests a learnability advantage for rUTAH over UTAH under derived approaches to linking theory knowledge development, based on the acquisition theories specified here.

Beyond our concrete results, our acquisition framework also highlights the components necessary for deriving linking theories, some of which may be both domain-specific and innate given our current understanding of child language acquisition. These components include a bias to look for links between thematic roles and syntactic positions, a bias to look for 1-to-1 links, and the procedure for generating linking patterns to evaluate. An interesting open question is whether a way can be found to derive these components from other, more fundamental, components. Finally, our results suggest that the Tolerance Principle is a useful, cognitively-grounded evaluation procedure for deriving linking theories from children's input. It is our hope that these results will spur future research both into the evaluation-via-acquisition of other linking theories, and into the syntactic consequences of the learnability of UTAH and rUTAH.

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.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

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.

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

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, *15*(2), 317–335.

Davis, K. F., Parker, K. P., & Montgomery, G. L. (2004). Sleep in infants and young children: Part one: normal sleep. *Journal of Pediatric Health Care*, *18*(2), 65–71.

Denison, S., Bonawitz, E., Gopnik, A., & Griffiths, T. L. (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition*, *126*(2), 285–300.

Denison, S., Reed, C., & Xu, F. (2011). The emergence of probabilistic reasoning in very young infants. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33).

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

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

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

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.

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125–155.

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

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young american children*. Baltimore, MD: P.H. Brookes.

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

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

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

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

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

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

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. (2017). An argument from acquisition: Comparing English metrical stress representations by how learnable they are from child-directed speech. *Language*

Acquisition, *24*, 307–342.

Pearl, L., & Sprouse, J. (2013). 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. (2015). Computational modeling for language acquisition: A tutorial with syntactic islands. *Journal of Speech, Language, and Hearing Research*, *58*, 740–753.

Pearl, L., & Sprouse, J. (2019). *Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition.* University of California, Irvine and University of Connecticut, Storrs. Retrieved from https://ling.auf.net/lingbuzz/003913

Perfors, A. (2012). Bayesian models of cognition: What's built in after all? *Philosophy Compass*, *7*(2), 127–138.

Perfors, A., Tenenbaum, J., Griffiths, T., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321.

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

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

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.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

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

Stahl, A. E., Romberg, A. R., Roseberry, S., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Infants segment continuous events using transitional probabilities. *Child Development*, *85*(5), 1821–1826.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental Psychology*, *39*(4), 706.

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

Wang, H., & Mintz, T. (2008). A dynamic learning model for categorizing words using frames. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings of the 32nd annual Boston University Conference on Language Development [BUCLD 32]* (pp. 525–536). Somerville, MA: Cascadilla Press.

Xu, F., & Tenenbaum, J. (2007). Word Learning as Bayesian Inference. *Psychological Review*, *114*(2), 245-272.

Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford, UK: Oxford University Press.

Yang, C. (2004). Universal grammar, statistics or both? *Trends in Cognitive Science*, *8*(10),

451–456.

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

Yang, C. (2012). Computational models of syntactic acquisition. *WIREs Cognitive Science*, *3*, 205–213.

Yang, C. (2015). Negative knowledge from positive evidence. *Language*, *91*(4), 938–953.

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

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.

Yurovsky, D., Case, S., & Frank, M. C. (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. *Psychological Science*, *28*(1), 132–140.

A Calculating the number of verb links children hear

At the individual verb level, a child using the Tolerance Principle would consider the verb links that obey and disobey the pattern under consideration (1-link or 3-link). This means that to apply the Tolerance Principle appropriately, we need to extrapolate from our corpus input sample to the true count of verb link instances for each individual verb. If we assume our corpus distribution of verb link instances reflects the true distribution that children hear, we simply need to estimate the true number of link instances children hear. We counted the individual verb link instances based on the utterances in our corpus sample that American English children hear at different ages (in particular, by age three, four, and five). So, if we can calculate the total utterances children hear by these ages, we can calculate the necessary multiplier for the verb link instance counts in our corpus. Then, we can multiply any individual verb's link instances by this multiplier and apply the Tolerance Principle.

To estimate the total number of utterances children hear by different ages, we draw on Hart and Risley (1995, 2003), who find that professional class parents spoke an average of 487 utterances per hour to their children ages 13-36 months. We also draw on Davis et al. (2004), which provides average total daily sleep hours for children; we subtract sleeping hours from total hours per day (24) to calculate the waking hours during which children hear input from their caretakers.

We also assume that children need certain linguistic knowledge and abilities in place before they can reliably extract the syntactic positions that the arguments appear in. In particular, they need to be able to segment the speech stream (potentially available at 7 months: Thiessen & Saffran, 2003), identify the meaning of word forms appearing in certain positions (potentially available at 6 months: Bergelson & Swingley, 2012), and recognize enough syntactic structure to identify syntactic positions (potentially available at 28 months: L. Naigles, 1990; L. G. Naigles & Kako, 1993; Scott & Fisher, 2009; Yuan & Fisher, 2009). Given this, we calculate children's waking hours starting at 28 months, when they would hear utterances and potentially be able to extract the relevant verb argument information (i.e., syntactic position and thematic role of the argument).

Based on Davis et al. (2004), children sleep for approximately 13 hours/day at 2 (24-35 months), 12 hours/day at 3 (36-47 months), and 11.5 hours/day at 4 (48-59 months). Therefore, we can cal-

culate their waking hours and total utterances heard, as in Table 9.

Table 9: Calculating the total utterances children hear by ages three (36 months), four (48 months), and five (60 months), for the purposes of learning linking theories. These calculations are based on waking hours per day (waking), total waking hours, and children hearing 487 utterances per waking hour. Cumulative utterances heard by age three ($<$3yrs = $<$36 months), four ($<$4yrs = $<$48 months), and five ($<$5yrs = $<$60 months) are shown.

age	age range	waking	total waking hours	total utt	cumulative utt
$<$3yrs	28 - 35 months	11	11 hrs/day * 365 days/yr * 8/12 = 2676.67	1303537	1303537
$<$4yrs	36 - 47 months	12	12 hrs/day * 365 days/yr = 4380	2133060	3436597
$<$5yrs	48 - 59 months	12.5	12.5 hrs/day * 365 days/yr = 4562.5	2221938	5658535

With these total utterance estimates, we can then extrapolate from our corpus samples by multiplying the individual verb utterance counts by an appropriate constant. This constant is calculated for each dataset in Table 10. As mentioned in the beginning, individual verb links were derived from the utterances in our corpus; so, this same multiplier can be used to estimate the true counts of verb links that children of different ages would have heard in their input (i.e., corpus count * multiplier = true count). We then apply the Tolerance Principle to these counts.

Table 10: Calculating the multiplier constant for each dataset, based on the number of utterances in the corpus sample and the number of utterances children would have heard by that age.

dataset	# utt	# utt heard	multiplier
$<$3yrs	39772	1303537	1303537/39772 = 32.775
$<$4yrs	50737	3436597	3436957/50737 = 67.741
$<$5yrs	56461	5658535	5658535/56461 = 100.220