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

Lisa Pearl & Jon Sprouse

Abstract

We present an argument from acquisition for linking theories, and explore to what extent two prominent linking theories in the syntactic literature – UTAH and rUTAH – can be learned from the data English children encounter. We leverage a conceptual acquisition framework involving explicit hypothesis generation and evaluation, the Tolerance Principle as a cognitively-motivated mechanism for hypothesis evaluation, and realistic English child-directed input. We find that UTAH – unlike rUTAH – is neither easily generated as an explicit hypothesis nor learnable from child-directed speech in its traditional form. We discuss the implications of these results for both syntactic and acquisition theories.

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

1 Introduction

The linking problem is a fundamental component of verb learning - children must learn how to link the thematic roles specified by a verb's lexical semantics to the syntactic argument positions specified by that verb's syntactic frame. 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 INSTRUMENT/SOURCE/GOAL-like roles tend to appear in oblique syntactic positions such as *indirect object* or *object of PP*. 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. Here we attempt to relax this link between UTAH/rUTAH and the innate approach, and instead ask what it would look like to attempt to derive UTAH and rUTAH from more general learning mechanisms applied to the input children encounter. In particular, we explore to what extent the Tolerance Principle, a cognitively plausible decision criterion, could be used to learn UTAH and rUTAH from realistic samples of English child-directed speech data.

It might seem counterintuitive to explore how two nativist linking theories could be learned. However, though UTAH and rUTAH are typically assumed to be innate by syntacticians exploring them, there is nothing inherent in either linking theory that requires them to be innate (we will review UTAH and rUTAH in detail in section 2). Because most theories seek to minimize innate, domain-specific knowledge, it would be useful to know how necessary the innateness assumption truly is for UTAH and rUTAH. Moreover, there is a growing body of empirical work in theoretical syntax suggesting that, while UTAH and rUTAH appear empirically adequate for the vast majority of accusative languages, there are some languages where UTAH and rUTAH do not appear as adequate. This raises questions about the universality of UTAH and rUTAH (Wood, 2015; Kastner, 2016; Myler, 2016). If UTAH and rUTAH are not universal, but still believed to be operative in at least some languages, this suggests that they may be learned from the input of those languages. So, we can use acquisition evidence to help refine linking theories – i.e., an argument from acquisition (Pearl, 2017; Pearl, Ho, & Detrano, 2017) for UTAH and rUTAH.

From a developmental perspective, there is also relatively little modeling research on the specific mechanisms underlying the acquisition of linking theories. UTAH and rUTAH make ideal case studies for exploring the acquisition of linking theories because (i) they are well-defined theories, (ii) they represent relatively cognitively plausible linking theories, and (iii) they represent relatively distinct locations in the space of possible linking theories (particularly along the absolute-relativized dimension, as discussed in more detail in section 2).

To concretely investigate the acquisition of UTAH and rUTAH, we leverage several conceptual and empirical tools from the developmental modeling literature: (i) the ideas of an implicit hypothesis space, explicit hypothesis generation, and explicit hypothesis evaluation (Yang, 2004, 2005; Perfors, 2012), (ii) Yang's (2005, 2016) Tolerance Principle as a way to evaluate explicit hypotheses, and (iii) a corpus of syntactically and thematically annotated child-directed speech (the CHILDES Treebank: Pearl & Sprouse, 2013, 2018) which serves as realistic child input. In the sections that follow, we first review the details of UTAH and rUTAH. We then explore how the components of UTAH and rUTAH define an implicit hypothesis space, and how two cycles of explicit hypothesis generation and explicit hypothesis evaluation, once over all verb tokens and once over verbs as lexical units, can be used to potentially learn UTAH and rUTAH. We then review the Tolerance Principle, and how it can be used as a psychologically-motivated decision criterion for explicit hypothesis evaluation. We next discuss the data that we use as input for explicit hypothesis generation and evaluation, directed at children under age 3, under age 4, and under age 5. We then present the results of the two cycles of hypothesis generation and evaluation for both UTAH and rUTAH across all three age groups, as well as an additional analysis investigating whether the results would be different if we decomposed UTAH into its constituent mappings instead of assuming one complex mapping.

Our results complement the emerging empirical and theoretical debates surrounding UTAH and rUTAH in potentially interesting ways. First, our results suggest that a child using the Tolerance Principle to learn linking theories from English child-directed speech data has two empirical advantages for learning rUTAH instead of UTAH. First, a rUTAH-learning child will have a significantly easier time explicitly generating complex hypotheses for solving the linking problem. Second, rUTAH is the only complex linking pattern that can be successfully learned from these realistic child-directed input using the Tolerance Principle. We discuss the implications of our results for both syntactic theory and acquisition theory.

2 Linking theories: a thematic role system, a set of syntactic positions, and a linking principle

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 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. Making these components explicit will be beneficial when we discuss the procedures for explicit hypothesis generation and explicit hypothesis evaluation.

2.1 UTAH

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 thematic (macro)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) The relationship between Baker's (1997) three 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

Baker's (1997) formulation of UTAH similarly posits three syntactic positions, which are defined by specific syntactic features (again, with much debate about which features are relevant for defining syntactic positions, and even whether these positions can be suitably defined). For convenience, here we will call these *subject*, *object*, and *oblique* (such as the object of a prepositional phrase), and use italics to indicate these are theory-specific labels; however, we note that we are using these terms descriptively, and do not intend to imply that subject, object, or oblique are (or are not) theoretical primitives. Finally, 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.

With these three components in place, it is straightforward to see how Baker's UTAH implementation accounts 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 derivation, UTAH is indeed respected.

2.2 rUTAH

Whereas the mapping between thematic roles and syntactic positions in UTAH is *absolute* in the sense that each thematic role will map to the same syntactic position in every construction, one of the defining features of rUTAH is that the mapping between thematic roles and syntactic position 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 will call these FIRST, SECOND, and THIRD here. 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*. Finally, 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 highest syntactic position, the SECOND thematic role maps to the next highest syntactic position, and so on.

One of the potentially interesting features 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 is a paradigm example of the rUTAH mapping: the one and only thematic role, patient, is the FIRST in the sentence, and it is mapped to the *subject* position, which is the one and only (and therefore highest) syntactic position in the sentence. There is no need for a movement operation (or, indeed, even a derivational grammar). This sentence simply is 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 given grammatical rule.

2.3 UTAH and rUTAH as case studies for modeling the acquisition of linking theories

We have chosen to 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 such fine detail that they help to make the scope of the acquisition problem 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 an implicit 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). Every acquisition theory must additionally include a procedure for generating explicit hypotheses about which link to evaluate, whether these hypotheses are about basic links like AGENT $\rightarrow$ *subject* or the complex patterns that UTAH and rUTAH ultimately specify. Moreover, every acquisition theory must specify a procedure for evaluating those hypotheses relative to the data. The detail with which UTAH and rUTAH have been specified show us that, while different acquisition theories may make different choices in the details of each component (different roles, different positions, different explicit hypothesis generation or evaluation procedures), the overall complexity in terms of the number of components of the acquisition theory 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 cognitively plausible simplifying assumptions about the number of thematic roles and syntactic positions. These simplifying assumptions are plausible from the perspective of language development because in order to map thematic roles onto syntactic positions, children are likely to either (i) make a small number of categories of thematic roles corresponding to roles (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 both to theorists exploring more complex systems and to theorists exploring less complex systems.

Third, UTAH and rUTAH are fairly far apart in the linking theory hypothesis space to the extent that absolute systems and relative systems are categorical opposites, and to the extent that UTAH and rUTAH are pure instantiations of these systems (UTAH is absolute 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 absolute 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 are both well-specified in the literature, cognitively plausible, and are likely to return useful information about whether the relevant linking knowledge can be derived from the input using an absolute vs. a relative system approach.

3 Conceptual tools: The implicit hypothesis space, explicit hypothesis generation, and explicit hypothesis evaluation

A useful framework for thinking about the acquisition of linking theories involves conceptual tools from developmental modelers (Yang, 2002, 2004; Perfors, 2012). The first component of this framework is an implicit hypothesis space (sometimes called a latent hypothesis space) (Perfors, 2012): this is the space of possible hypotheses licensed by the the problem under consideration. Importantly, the implicit hypothesis space is constrained. For example, under a parametric approach to grammars, a grammar is a combination of parameter values; so, the parameters and the values they can take define the implicit hypothesis space of all possible grammars (e.g., n binary-valued parameters define an implicit hypothesis space of 2^n grammars). The implicit hypothesis space for UTAH/rUTAH is constrained by the the assumptions that we make about the form of the linking theory; this involves the number of thematic roles, the number of syntactic positions, the direction of the links licensed by the mapping, and any constraints we impose on the sets of links that can co-occur in a linking pattern.

UTAH and rUTAH, as specified in the syntactic literature, appear to assume the following: (i) there are 3 roles and 3 positions, (ii) links can go in either or both directions, and (iii) roles and positions can only participate in one link at a time. This last constraint has several implications. First, no unidirectional links with overlapping roles or positions are allowed (e.g., an AGENT $\rightarrow$ *subject* link can't exist simultaneously with an AGENT $\rightarrow$ *object* link, because the AGENT role is overlapping). Second, multiple roles can't map to the same position (e.g., AGENT and PATIENT can't both map to *subject* via the same unidirectional link, such as AGENT OR PATIENT $\rightarrow$ *subject*). Relatedly, multiple positions can't map to the same role (e.g., *subject* and *object* can't both map to AGENT via the same unidirectional link, such as *subject or object* $\rightarrow$ AGENT). This leads to

1155 possible linking theories, with the theories containing anywhere from 0 to 3 links in either direction (role-to-position or position-to-role).

The second component of this framework is a cycle of explicit hypothesis generation and explicit hypothesis evaluation. The idea that children construct explicit hypotheses from the implicit hypothesis space in order to evaluate these hypotheses is familiar to developmental linguists working within the linguistic parameters approach. Linguistic parameters (and the values these parameters take) define the implicit space of possible grammars, but a child doesn't necessarily consider all these grammars explicitly for any given data point; instead, she can generate a grammar as an explicit combination of parameter values and then evaluate that grammar with respect to the data at hand (Yang, 2002, 2004). In the case of the primary linking pattern that we consider here, what we need is a way for children to generate either the UTAH or rUTAH linking theory, and then evaluate it against the verbs of their language. The acquisition process we propose for linking theories involves two cycles of explicit hypothesis generation and evaluation: one over all verb tokens together, and one over verbs as a lexical type.

In the first cycle, explicit hypothesis generation is handled by an assumption available *a priori*: the learner will first focus on the basic unidirectional links defined by the 3 roles and 3 positions in UTAH and rUTAH. This yields 9 unidirectional links from role to a single position (e.g, AGENT $\rightarrow$ *subject*), and 9 unidirectional links from position to a single role (e.g., *subject* $\rightarrow$ AGENT). Explicit hypothesis evaluation is then handled by the Tolerance Principle (Yang, 2005, 2016), which evaluates these 18 explicitly hypothesized links by the instances from all the verbs collectively that English children encounter before ages 3, 4, and 5. This is because this cycle of explicit hypothesis generation and evaluation represents a child trying to identify reliable basic links in verb usage in general.

In the second cycle, we explore three ways to generate explicit hypotheses of complex linking theories from the reliable unidirectional links of the first cycle. Hypothesis evaluation of a complex theory is again handled by the Tolerance Principle, but this time it's applied to verb types (i.e., the lexical entries for each individual verb), rather than the collective verb tokens. This is because the overall linking patterns given by the complex theories of UTAH and rUTAH are expectations that hold for individual verbs: there are verbs that follow the pattern, and verbs that are exceptions.

The evaluation of a linking pattern is itself a two step process. First, a child must evaluate if each verb follows the hypothesized linking pattern or not (i.e., if it's a pattern-obeying verb). This can be done by applying the Tolerance Principle to the individual usage tokens of that verb in the child's input data. Second, the child must evaluate if the linking pattern is a reliable pattern for all the verbs collectively; this can be done by applying the Tolerance Principle to determine if there are sufficient pattern-obeying verbs. Sections 6 and 7 discuss these two cycles in detail, along with the implications of the results of these cycles for the learnability of both UTAH and rUTAH.

4 The Tolerance Principle as a decision criterion

The Tolerance Principle is a formal approach for determining when a child would choose to adopt a "rule" or default pattern to account for a set of items (Yang, 2005, 2016). 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, and determines how many exceptions a rule can "tolerate" in the data before it's not worthwhile to have that rule at all.

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 et al., 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) specifies this 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 in 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 the default. For example, a rule that could apply to 100 items allows only 21 exceptions (21%), and thus requires 79 items that match the rule to be adopted. A rule that could apply to 1000 items allows only 144 exceptions (14.4%), and therefore requires 856 items that match the rule to be adopted. A rule that could apply to 10000 items allows only 1085 exceptions (10.85%), and therefore requires 8915 items that match the rule to be adopted. This has the practical effect of allowing only one option to be the default rule (i.e., this disallows two or more "defaults" 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 hypothesis evaluation in both cycles of learning linking theories.

5 The child-directed speech data used for inference

The child-directed speech data we use (summarized in Table 1) come from the CHILDES Treebank (Pearl & Sprouse, 2013, 2018), which contains realistic samples of speech directed at children between one and five years old, annotated with linguistic and non-linguistic information. This portion of the CHILDES Treebank involved around 140,000 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. Here, we're interested in the syntactic information corresponding to the syntactic positions of *subject*, *object*, and *oblique object*, and the thematic information corresponding to the thematic roles assumed by UTAH and rUTAH (AGENT, FIRST, etc). We divided these ≈140K utterances into age ranges based on the age of the child the speech was directed at: less than 3 years of age (<3yrs), less than 4 years of age (<4yrs), and less than 5 years of age (<5yrs). We then constructed datasets representing the input to a child of a particular age. 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 minimize data sparseness problems with respect to the set of N items the Tolerance Principle could be applied to, we restrict our analyses – and therefore the child's intake for acquisition – 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 threw the ball*), and one utterance with arguments in *subject*, *object*, and *oblique object* position (*He threw me the penguin*). 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 linking pattern, we will refer to an argument use of this kind as a "linking pattern instance" and we only include verbs with 5 or more linking pattern instances in our analyses.

6 Cycle 1: Basic unidirectional links

In section 3, we outlined a two cycle process for using the Tolerance Principle to learn the complex UTAH and rUTAH linking theory hypotheses from the implicit hypothesis space. The first cycle of that process relies on an *a priori* bias to focus on the 18 basic, unidirectional links that can be explicitly generated from the implicit hypothesis space: 9 links from role to position (such as AGENT $\rightarrow$ *subject*) and 9 links from position to role (such as *subject* $\rightarrow$ AGENT). A child could then use the Tolerance Principle to evaluate the reliability of each of these 18 links – i.e., whether the link is regular enough to be a dominant link.

In this cycle, the first step is for the child to track how often each thematic role appears in each syntactic position across all verb uses (this is in fact what's already encoded in the derived input files in the publicly available code base at https://github.com/lisapearl/linking-problem-code). Table 2 presents this information for the three age groups and the two linking systems (UTAH and rUTAH). Armed with these frequencies, the child can use the Tolerance Principle to evaluate each

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 linking pattern 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

link. More specifically, for each potential link, the Tolerance Principle can indicate how many items must follow that link pattern in order for the link to be viewed as a default link.

For example, when assessing the AGENT → *subject* link for UTAH, the Tolerance Principle will indicate how many AGENTs must appear in *subject* position, based on the size of the entire set of AGENTs. If the number of AGENTs appearing in *subject* position is above the Tolerance Principle threshold, the AGENT → *subject* link will be viewed by the child as a default link. The calculation itself requires reading across the rows in Table 2 for the role-to-position links, and down the columns for the position-to-role links. For example, to determine N for the UTAH AGENT → *subject* link in the <3yrs data, the total number of AGENTs is summed over the syntactic positions: 14464 + 11 + 19 = 14494. In contrast, to determine N for the UTAH *subject* → AGENT link in the <3yrs data, the total number of *subject*s is summed over the thematic roles: 14464 + 3739 + 156 = 18359. The number of required items following the link ($N - \frac{N}{ln(N)}$) is then calculated from N in each case.

We report the results of those calculations in Tables 3-8: the total number of items potentially following the link, the required number of items that must follow the link according to the Tolerance Principle, and the actual number items observed to follow the link. If the actual number observed is equal to or greater than the required number, the Tolerance Principle supports establishing this link as the default. We have **bolded** the links that would be viewed as default links in the these tables.

Table 2: Linking pattern instances across the syntactic positions of *subject*, *object* and *oblique object* for the different thematic representations, given the data available to three-, four-, and five-year-olds. Instances compatible with the primary linking pattern are **bolded**.

		<3yrs			<4yrs			<5yrs		
		subject	*object*	*oblique*	*subject*	*object*	*oblique*	*subject*	*object*	*oblique*
UTAH	AGENT	**14464**	11	19	**19324**	16	23	**22048**	19	25
	PATIENT	3739	**16105**	1068	4619	**21126**	1466	5074	**24013**	1677
	OTHER	156	287	**3404**	173	358	**4423**	177	383	**4939**
rUTAH	FIRST	**18015**	271	27	**23733**	298	33	**26899**	307	36
	SECOND	264	**15832**	100	287	**20799**	154	296	**23665**	180
	THIRD	27	139	**4287**	33	210	**5620**	36	237	**6313**

Table 3: Tolerance Principle analysis for each basic link under the UTAH system for <3yrs input. Links that exceed the Tolerance Principle threshold ($\geq N - \frac{N}{ln(N)}$) are **bolded**.

role to position			total (N)	required	actual	position to role			total (N)	required	actual
AGENT	→	***subject***	**14494**	**12982**	**14464**	*subject*	→	AGENT	18359	16490	14464
AGENT	→	*object*	14494	12982	11	*object*	→	AGENT	16403	14713	11
AGENT	→	*oblique*	14494	12982	19	*oblique*	→	AGENT	4491	3957	19
PATIENT	→	*subject*	20912	18900	3739	*subject*	→	PATIENT	18359	16490	3739
PATIENT	→	*object*	20912	18900	16105	***object***	→	**PATIENT**	**16403**	**14713**	**16105**
PATIENT	→	*oblique*	20912	18900	1068	*oblique*	→	PATIENT	4491	3957	1068
OTHER	→	*subject*	3847	3381	156	*subject*	→	OTHER	18359	16490	156
OTHER	→	*object*	3847	3381	287	*object*	→	OTHER	16403	14713	287
OTHER	→	***oblique***	**3847**	**3381**	**3404**	*oblique*	→	OTHER	4491	3957	3404

Table 4: Tolerance Principle analysis for each basic link under the rUTAH system for <3yrs input. Links that exceed the Tolerance Principle threshold ($\geq N - \frac{N}{ln(N)}$) are **bolded**.

role to position			total (N)	required	actual	position to role			total (N)	required	actual
FIRST	→	***subject***	**18313**	**16448**	**18015**	***subject***	→	**FIRST**	**18306**	**16441**	**18015**
FIRST	→	*object*	18313	16448	271	*object*	→	FIRST	16242	14567	271
FIRST	→	*oblique*	18313	16448	27	*oblique*	→	FIRST	4414	3889	27
SECOND	→	*subject*	16196	14526	264	*subject*	→	SECOND	18306	16441	264
SECOND	→	***object***	**16196**	**14526**	**15832**	***object***	→	**SECOND**	**16242**	**14567**	**15832**
SECOND	→	*oblique*	16196	14526	100	*oblique*	→	SECOND	4414	3889	100
THIRD	→	*subject*	4453	3923	27	*subject*	→	THIRD	18306	16441	27
THIRD	→	*object*	4453	3923	139	*object*	→	THIRD	16242	14567	139
THIRD	→	***oblique***	**4453**	**3923**	**4287**	***oblique***	→	**THIRD**	**4414**	**3889**	**4287**

Table 5: Tolerance Principle analysis for each basic link under the UTAH system for <4yrs input. Links that exceed the Tolerance Principle threshold ($\geq N - \frac{N}{ln(N)}$) are **bolded**.

role to position			total (N)	required	actual	position to role			total (N)	required	actual
AGENT	→	***subject***	**19363**	**17402**	**19324**	*subject*	→	AGENT	24116	21727	19324
AGENT	→	*object*	19363	17402	16	*object*	→	AGENT	21500	19436	16
AGENT	→	*oblique*	19363	17402	23	*oblique*	→	AGENT	5912	5232	23
PATIENT	→	*subject*	27211	24638	4619	*subject*	→	PATIENT	24116	21727	4619
PATIENT	→	*object*	27211	24638	21126	***object***	→	**PATIENT**	**21500**	**19436**	**21126**
PATIENT	→	*oblique*	27211	24638	1466	*oblique*	→	PATIENT	5912	5232	1466
OTHER	→	*subject*	4954	2372	173	*subject*	→	OTHER	24116	21727	173
OTHER	→	*object*	4954	2372	358	*object*	→	OTHER	21500	19436	358
OTHER	→	***oblique***	**4954**	**2372**	**4423**	*oblique*	→	OTHER	5912	5232	4423

Table 6: Tolerance Principle analysis for each basic link under the rUTAH system for <4yrs input. Links that exceed the Tolerance Principle threshold ($\geq N - \frac{N}{ln(N)}$) are **bolded**.

role to position			total (N)	required	actual	position to role			total (N)	required	actual
FIRST	→	***subject***	**24064**	**21861**	**23733**	***subject***	→	**FIRST**	**24053**	**21679**	**23733**
FIRST	→	*object*	24064	21861	298	*object*	→	FIRST	21307	18900	298
FIRST	→	*oblique*	24064	21861	233	*oblique*	→	FIRST	5807	5317	233
SECOND	→	*subject*	21240	19109	287	*subject*	→	SECOND	24053	21679	287
SECOND	→	***object***	**21240**	**19109**	**20799**	***object***	→	**SECOND**	**21307**	**18900**	**20799**
SECOND	→	*oblique*	21240	19109	154	*oblique*	→	SECOND	5807	5317	154
THIRD	→	*subject*	5863	5188	33	*subject*	→	THIRD	24053	21679	33
THIRD	→	*object*	5863	5188	210	*object*	→	THIRD	21307	18900	210
THIRD	→	***oblique***	**5863**	**5188**	**5620**	***oblique***	→	**THIRD**	**5807**	**5317**	**5620**

Table 7: Tolerance Principle analysis for each basic link under the UTAH system for <5yrs input. Links that exceed the Tolerance Principle threshold ($\geq N - \frac{N}{ln(N)}$) are **bolded**.

role to position			total (N)	required	actual	position to role			total (N)	required	actual
AGENT	→	***subject***	**22092**	**19884**	**22048**	*subject*	→	AGENT	27299	24627	22048
AGENT	→	*object*	22092	19884	19	*object*	→	AGENT	24415	21999	19
AGENT	→	*oblique*	22092	19884	25	*oblique*	→	AGENT	6641	5786	25
PATIENT	→	*subject*	30764	27788	5074	*subject*	→	PATIENT	27299	24627	5074
PATIENT	→	*object*	30764	27788	24013	***object***	→	**PATIENT**	**24415**	**21999**	**24013**
PATIENT	→	*oblique*	30764	27788	1677	*oblique*	→	PATIENT	6641	5786	1677
OTHER	→	*subject*	5499	4861	177	*subject*	→	OTHER	27299	24627	177
OTHER	→	*object*	5499	4861	383	*object*	→	OTHER	24415	21999	383
OTHER	→	***oblique***	**5499**	**4861**	**4939**	*oblique*	→	OTHER	6641	5786	4939

Table 8: Tolerance Principle analysis for each basic link under the rUTAH system for <5yrs input. Links that exceed the Tolerance Principle threshold ($\geq N - \frac{N}{ln(N)}$) are **bolded**.

role to position			total (N)	required	actual	position to role			total (N)	required	actual
FIRST	$\rightarrow$	***subject***	**27242**	**24575**	**26899**	***subject***	$\rightarrow$	**FIRST**	**27231**	**24565**	**26899**
FIRST	$\rightarrow$	*object*	27242	24575	307	*object*	$\rightarrow$	FIRST	24209	21811	307
FIRST	$\rightarrow$	*oblique*	27242	24575	36	*oblique*	$\rightarrow$	FIRST	6529	5786	36
SECOND	$\rightarrow$	*subject*	24141	21749	296	*subject*	$\rightarrow$	SECOND	27231	24565	296
SECOND	$\rightarrow$	***object***	**24141**	**21749**	**23665**	***object***	$\rightarrow$	**SECOND**	**24209**	**21811**	**23665**
SECOND	$\rightarrow$	*oblique*	24141	21749	180	*oblique*	$\rightarrow$	SECOND	6529	5786	180
THIRD	$\rightarrow$	*subject*	6586	5837	36	*subject*	$\rightarrow$	THIRD	27231	24565	36
THIRD	$\rightarrow$	*object*	6586	5837	237	*object*	$\rightarrow$	THIRD	24209	21811	237
THIRD	$\rightarrow$	***oblique***	**6586**	**5837**	**6313**	***oblique***	$\rightarrow$	**THIRD**	**6529**	**5786**	**6313**

We can make three observations on the basis of these results. First, the results are the same for all three age groups. That is, the links that emerge as defaults for UTAH for the <3, <4, and <5 age ranges are the same, and the links that emerge as defaults for rUTAH for the <3, <4, and <5 age ranges are the same. So, the quality of the speech directed at children of different ages doesn't seem to impact the learnability of individual links. Second, within the absolute UTAH-based system of roles and positions, there are only three links that are reliable enough to be viewed as default according to the Tolerance Principle, and these are listed in Table 9. Interestingly, none of those links can be composed into a bidirectional link. Finally, within the relative rUTAH-based system of roles and positions, there are six links that are reliable enough according to the Tolerance Principle. Those six links potentially combine to form three bidirectional links, as listed in Table 9. We note that these are precisely the bidirectional links that form the complex linking pattern for rUTAH. Taken together these results suggest that the complete rUTAH linking theory can potentially be composed from individual links that are sufficiently reliable, according to the Tolerance Principle analysis. In contrast, the complete UTAH linking theory will require additional assumptions, which we turn to in the next section.

7 Cycle 2: The complex linking patterns

7.1 Explicitly generating the complex pattern for UTAH vs. rUTAH

The first cycle of explicit hypothesis generation and evaluation identified which of the 18 basic unidirectional links are learnable as default links, based on the Tolerance Principle and children's exposure to all verb use tokens. The goal of the second cycle is to generate the complex linking patterns of UTAH and rUTAH as explicit hypotheses, based on the default unidirectional links. Then, a child could evaluate those explicit complex hypotheses over the set of verbs in her input.

In terms of generating these explicit complex hypotheses,we can envision at least three possibilities that seem plausible. The first, and most conservative, possibility is that a complex hypothesis can only consist of links between a thematic role and syntactic position if the evaluation in cycle

Table 9: The basic links that would be evaluated by a child using the Tolerance Principle as default links, considering all verb use tokens collectively 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 → *subject*	*subject* → FIRST
	SECOND → *object*	*object* → SECOND
	THIRD → *oblique*	*oblique* → THIRD

1 indicated a default bidirectional link between the two. That is, if a thematic role T has a default syntactic position S *and* that syntactic position S has the thematic role T as its default, only then can a complex linking hypothesis include a link between T and S. Under this approach of explicit hypothesis generation, only rUTAH would enable the child to explicitly generate the correct complex linking hypothesis. UTAH only ever has links in one direction for each thematic role and syntactic position, and so the child could not explicitly generate the correct complex linking hypothesis for UTAH.

The second possibility builds on the intuition that linking patterns are an expectation about thematic roles and what their positional preferences are, rather than about syntactic positions and their preferences for thematic roles. Under this view, all that matters is establishing a link from role to position. That is, the child will only include links from thematic role T to syntactic position S when explicitly generating complex linking hypotheses. Under this approach, two-thirds of the correct UTAH linking hypothesis could be generated: AGENT → *subject* and OTHER → *oblique object*. However, no link would be included for PATIENT. So, as before, the complete correct UTAH linking hypothesis would not be explicitly generated – instead, only a partial pattern would be. In contrast, the correct rUTAH linking hypothesis would be explicitly generated, as all three thematic roles were assigned to default syntactic positions via the evaluation of default links in cycle 1.

The third, and most liberal, possbility is that the child considers any link between a thematic role T and a syntactic position S as sufficient for including that link in the complex linking hypothesis. That is, as long as thematic role T has a default syntactic position S *or* syntactic position S has a default thematic role T, the child's hypothesis includes a link between T and S. Under this approach, the correct complex linking hypothesis for both UTAH and rUTAH would be generated.

While we don't know how children generate explicit complex linking hypotheses, the exercise above of working three plausible possibilities is instructive. It shows us that certain relations hold between the representations and the learning assumptions needed to support the acquisition of those representations, provided we believe in the acquisition process captured by the first cycle of explicit hypothesis generation and evaluation. In particular, if we ultimately believe that UTAH is the correct target knowledge state, then the generation mechanism for complex linking hypotheses must be relatively liberal, as in possibility 3. In contrast, the complex linking hypothesis for rUTAH is relatively robust to the complex hypothesis generation mechanism – 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 the rUTAH links are easier to learn as defaults from the input.

7.2 Evaluating explicit complex hypotheses

Once children have generated the complex linking pattern of UTAH or rUTAH as an explicit hypothesis, they can then evaluate it to see if it's the default pattern for verbs of their language. We can again assess this using the Tolerance Principle, based on the number of verbs in the input. More specifically, under the assumption that a complex linking pattern is a hypothesis about how individual verbs behave, a child needs to determine how many verbs follow the primary pattern and how many don't. Given the total number of verbs V (i.e., lexical entries for verbs) in a sample of children's input, the Tolerance Principle offers a threshold for how many exceptions to a "complex linking pattern rule" could exist without the child abandoning the rule altogether: $\frac{V}{ln(V)}$; so this requires that $V - \frac{V}{ln(V)}$ verbs follow this rule. If the number of verbs following the complex linking pattern in children's input is equal to or greater than this threshold, then those data support the existence of a "complex linking pattern rule"; this corresponds to English children developing an expectation of the complex linking pattern as the default for verbs of the language.

Given the corpus data available, we can determine the number of verbs that must follow the complex linking pattern to support the child's generalization that the complex linking pattern is the default pattern. So, we simply have to calculate from these corpus data how many verbs at each age actually *do* follow the UTAH or rUTAH complex linking patterns.

How then do we tell if an individual verb follows the complex linking pattern or not? Importantly, this evaluation is done through the filter of the complex linking pattern: it is a binary distinction where an instance of a thematic role in a syntactic position either is or isn't compatible with the complex linking pattern. That is, with an explicit complex linking hypothesis in hand, the fine-grained details of where a specific thematic role (e.g., AGENT or FIRST) appears doesn't matter. Instead, the child is evaluating if a thematic role (e.g., AGENT or FIRST) appears in the syntactic position it's expected to appear in according to the explicit complex linking pattern hypothesis (e.g., *subject*).

This brings to a light the question how to count instances of a complex linking pattern. Consider this use of the verb *pet*: *Lily pets the kitties*. If each link is considered an instance of the complex linking pattern, this use counts as two instances that obey the UTAH complex linking pattern: one for AGENT↔*subject* and one for PATIENT↔*object*. In contrast, if each verb use is considered an instance of the complex linking pattern, this use counts as a single instance that obeys the UTAH pattern, as all thematic roles are in their expected positions. We consider both approaches (link-based and verb-use-based) to counting complex linking pattern instances below, as it's unclear *a priori* which one a child would select.

Any individual verb $v_i \in V$ will then have some number of instances of complex linking patterns. If we view the Tolerance Principle as a general cognitive principle for how children make decisions, then we can also apply it at the individual verb level to determine if a verb has a default pattern or not. (This is similar to how this principle has been proposed to be used in metrical phonology to determine if a word form has a regular stress pattern or not: Legate & Yang, 2013; Pearl et al., 2017; Pearl, 2017). In particular, we can apply the Tolerance Principle to determine if an individual verb is a "complex linking pattern" verb by looking at the verb's linking pattern instances I_{v_i}. The Tolerance Principle would predict that the verb v_i is a complex linking pattern verb if there are fewer than $\frac{I_{v_i}}{ln(I_{v_i})}$ exceptions to the pattern for verb v_i; this means there are at least $I_{v_i} - \frac{I_{v_i}}{ln(I_{v_i})}$) instances that follow the complex linking pattern.

If so, v_i is a complex linking pattern verb; if not, it isn't. Table 10 shows examples of applying the Tolerance Principle this way for individual verbs, given the link-based instances in the $<$3yrs data (which English children would have encountered by age three).

Table 10: An example individual-verb Tolerance Principle analysis for three verbs: *use*, *break*, and *belong*. The total number of link-based verb instances is shown, along with the number of pattern matching uses required by the Tolerance Principle and the actual number for that verb, given either UTAH and rUTAH. If the complex pattern is supported for that verb, the actual number is **bolded**.

verb	total (N)	required	UTAH actual	rUTAH actual
use	109	86	**107**	**107**
break	102	80	56	**102**
belong	51	39	14	29

As we see in Table 10, the inference that the child using the Tolerance Principle would make can vary, depending on the individual verb and the thematic representation used. For *use*, a child using the Tolerance Principle to evaluate either the UTAH or rUTAH complex linking pattern would find sufficient instances following the pattern (107>86); so the child would infer that *use* is a complex linking pattern verb under either theory. However, for *break*, the inference differs by complex linking theory: a child evaluating UTAH wouldn't find sufficient pattern-following instances (56<80), while a child evaluating rUTAH would (102>80). So, a child evaluating UTAH would not infer *break* is a complex-pattern-following verb, while a child evaluating rUTAH would. For *belong*, a child evaluating either UTAH or rUTAH wouldn't find sufficient pattern-following instances (14<39, 29<39). So, a child relying on the Tolerance Principle would infer that *belong* is an exception for either theory.

Once this inference is made for each of the verbs in English children's input, we can determine how many verbs are exceptions with respect to the complex linking pattern. If there aren't too many exceptional verbs, children should be able to infer that the complex linking pattern being evaluated is the default for English verbs. That is, children would be able to derive the knowledge that this complex linking pattern is the default pattern for all verbs, and so they should subsequently expect verbs to obey this linking pattern. In contrast, if there are too many exceptional verbs, children would not be able to derive the knowledge that the complex linking pattern is the default. So, they should not (yet) expect verbs to obey this linking pattern. Tables 11 and 12 show the results of

Tolerance Principle analysis on the verbs available to English children before ages three (<3yrs), four (<4yrs), and five (<5yrs), using either a link-based or verb-use-based approach to counting complex linking pattern instances.

Table 11: Tolerance Principle analysis of whether the complex linking pattern for UTAH and rUTAH could be inferred as the default pattern from the verb usage in children's input at different ages. The total number of verbs with 5 or more link-based complex linking pattern instances is shown, along with the number of pattern-matching verbs required by the Tolerance Principle and the actual number for both UTAH and rUTAH. If the complex linking pattern is supported for that age, the actual number is **bolded**.

age	total (N)	required	UTAH actual	rUTAH actual
<3yrs	231	189	169	**229**
<4yrs	260	214	194	**258**
<5yrs	275	227	205	**273**

Table 12: Tolerance Principle analysis of whether the complex linking patterns for UTAH and rUTAH could be inferred as the default pattern from the verb usage in children's input at different ages. The total number of verbs with 5 or more verb-use-based complex linking pattern instances is shown, along with the number of pattern-matching verbs required by the Tolerance Principle and the actual number for both UTAH and rUTAH. If the complex linking pattern is supported for that age, the actual number is **bolded**.

age	total (N)	required	UTAH actual	rUTAH actual
<3yrs	224	183	147	**222**
<4yrs	255	209	139	**253**
<5yrs	267	220	178	**265**

This analysis again highlights the impact that the choice of UTAH or rUTAH can have. Put simply, a child evaluating UTAH won't be able to infer that the complex linking pattern is in fact the default linking pattern that should be expected for all 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 evaluating rUTAH will always be able to infer that the complex linking pattern is the default linking pattern. 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 not unexpected, given that rUTAH generally minimizes the number of exceptions to the predicted linking pattern. But what is novel, at least to our knowledge, is the discovery that the complex linking pattern of UTAH is not learnable from child-directed speech because of the number of verbs that are exceptions on the surface. 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. We discuss the implications of this for both syntactic theory and language acquisition in section 9. But first, we provide one additional learnability analysis that involves breaking a complex linking theory into its constituent pieces.

8 Decomposing the linking theory: A single cycle focusing on basic links

For the previous acquisition analyses, we have been following the syntactic literature in assuming that the complex linking patterns postulated by UTAH and rUTAH are the correct adult target state. However, our results suggest that the UTAH complex linking pattern may not be learnable from child-directed input without making additional assumptions about other knowledge available *a priori*. Before exploring what those assumptions might be, we want to explore one other logical possibility: perhaps the correct linking theory is not a complex pattern, but instead a list of individual basic links. One interesting consequence of deciding to view the linking theory as a list of individual basic links is that we can reduce the acquisition process to a single cycle of explicit hypothesis generation (postulating the 18 basic links) and evaluation. Furthermore, this evaluation can be calculated over individual verbs, eliminating the original first cycle's approach of looking at all verb usage tokens together to identify reliable links that could be part of a complex linking theory. Of course, the potential downside of this approach is that it breaks the linking pattern into separate pieces, potentially complicating the grammatical theory (which in turn may have empirical consequences for syntactic theory that are beyond the scope of this paper).

We can assess this possibility with the same evaluation approach as before, using the Tolerance Principle to evaluate whether each of the 18 basic links between thematic role and syntactic position would be viewed as the default. Importantly, this evaluation occurs occurring over verbs to determine if each verb is an exception to the basic link or not. So, like the original first cycle, it evaluates basic links individually; like the original second cycle, it operates at the individual verb level, rather than collective verb token level. Table 13 reports the results of this acquisition analysis for the 18 basic UTAH links, and Table 14 reports the results of this analysis for the 18 basic rUTAH links.

The results of this analysis are identical to the results of first cycle in section 6. When evaluating over individual verbs rather than collective verb use tokens, the same three basic links emerge for UTAH and the same six links leading to three bidirectional links emerge for rUTAH. As with the original first cycle analysis, the results don't depend on the age group the input is directed at. What this means in practice is that the rUTAH linking theory is learnable as either a complex linking pattern or as a list of six unidirectional (forming three bidirectional) links. In contrast, the UTAH linking theory isn't learnable as a complex theory, though it *is* learnable as a system of three unidirectional links: AGENT $\rightarrow$ *subject*, *object* $\rightarrow$ PATIENT, and OTHER $\rightarrow$ *oblique*.

Table 13: Tolerance Principle analysis of whether the individual links that constitute UTAH could be inferred as the default patterns from the verb usage in children's input at different ages. Verbs with 5 or more relevant linking pattern usages are included in the analysis. The total number of verbs is shown, along with the number of pattern-matching verbs required by the Tolerance Principle and the actual number observed. If the link is supported for that age, the row is **bolded**.

age	role to position			total	required	actual	position to role			total	required	actual
<3	**AGENT**	**→**	***subject***	**158**	**127**	**157**	*subject*	→	AGENT	204	166	138
	PATIENT	→	*object*	214	175	134	***object***	**→**	**PATIENT**	**156**	**126**	**152**
	OTHER	**→**	***oblique***	**93**	**73**	**87**	*oblique*	→	OTHER	104	82	81
<4	**AGENT**	**→**	***subject***	**185**	**150**	**184**	*subject*	→	AGENT	229	187	152
	PATIENT	→	*object*	245	201	159	***object***	**→**	**PATIENT**	**183**	**148**	**178**
	OTHER	**→**	***oblique***	**101**	**80**	**95**	*oblique*	→	OTHER	115	91	89
<5	**AGENT**	**→**	***subject***	**294**	**158**	**193**	*subject*	→	AGENT	239	196	160
	PATIENT	→	*object*	257	211	166	***object***	**→**	**PATIENT**	**197**	**160**	**192**
	OTHER	**→**	***oblique***	**107**	**85**	**100**	*oblique*	→	OTHER	119	95	91

Table 14: Tolerance Principle analysis of whether the individual links that constitute rUTAH could be inferred as the default patterns from the verb usage in children's input at different ages. Verbs with 5 or more relevant linking pattern usages are included in the analysis. The total number of verbs is shown, along with the number of pattern-matching verbs required by the Tolerance Principle and the actual number observed. All of the links surpass the threshold set by the Tolerance Principle, so they aren't explicitly marked.

age	role to position			total	required	actual	position to role			total	required	actual
<3	FIRST	→	*subject*	204	165	202	*subject*	→	FIRST	204	165	202
	SECOND	→	*object*	155	124	154	*object*	→	SECOND	156	125	155
	THIRD	→	*oblique*	104	82	101	*oblique*	→	THIRD	104	82	102
<4	FIRST	→	*subject*	229	187	227	*subject*	→	FIRST	229	187	227
	SECOND	→	*object*	182	148	181	*object*	→	SECOND	182	148	181
	THIRD	→	*oblique*	115	91	111	*oblique*	→	THIRD	114	90	111
<5	FIRST	→	*subject*	239	196	237	*subject*	→	FIRST	239	196	237
	SECOND	→	*object*	197	160	196	*object*	→	SECOND	197	160	196
	THIRD	→	*oblique*	118	94	114	*oblique*	→	THIRD	117	95	114

9 Discussion

9.1 The components necessary for learning linking theories

We have set aside the typical claim from innate-linking approaches that the complex linking patterns for UTAH and rUTAH are innately specified, and asked if they could instead be learned from child-directed input in combination with more general learning mechanisms. This can help provide a concrete acquisition theory to accompany the derived-linking approach to learning complex linking theories. One major conclusion from our investigation is that there are a relatively large number of components necessary for learning linking theories, even after setting aside the complex patterns of UTAH and rUTAH. Based on the acquisition theory specified here, we propose that any theory of how children learn linking theories will require the components in Table 15.

Table 15: Proposed components that are required to learn 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 default link	domain-specific	derived
the ability to track links	domain-general	innate
a procedure to generate hypotheses	**either**	**innate**
a procedure to evaluate hypotheses	domain-general	innate

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. The goal is to determine if any of the components are likely to be simultaneously language-specific and innate, as these are the components that figure most prominently in the debates between innate and derived approaches to linking theories (and many other aspects of language acquisition).

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 the syntactic structures that underlie these positions require innate, domain-specific knowledge.

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 that is 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 followed here, there was a bias to look for a 1-to-1 mapping between roles and positions. That is, the child must only consider links that involve a single syntactic position or thematic role (e.g., AGENT $\rightarrow$ *subject*), rather than allowing disjunctive options that involve multiple syntactic positions or multiple thematic roles (e.g., AGENT or PATIENT $\rightarrow$ *subject*; PATIENT $\rightarrow$ *subject* or *object*). This constraint may be thought of as a sort of mapping akin 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 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 bias is that the child must assume there is only a single default link per role or position (i.e., AGENT defaults to only one of the available options, *subject* defaults to only one of the available options, etc.). This domain-specific knowledge derives directly from the Tolerance Principle (as discussed in section 4), 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), applied to the domain-specific knowledge of linking patterns as cognitive objects.

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

Given that this component list rests on the specific implementation we propose here for learning 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 hypothesis generation and evaluation procedures, etc). There are surely other choices for each of the components in our learning theories, 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 cognitively plausible and theoretically motivated.

In particular, we tested both absolute and relative thematic systems. We chose thematic systems and syntactic systems that only have three roles. We explored how to generate explicit complex linking theories via two cycles of hypothesis generation, one evaluating 18 basic links and one that evaluated the complex linking theory constructed one of three plausible ways. We also explored the generation and evaluation of decomposed linking patterns that didn't rely on aggregation via a complex linking theory. 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 15 and evaluation on the the child-directed speech data like those contained in the CHILDES Treebank.

9.2 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 (with all three links) is difficult to generate as an explicit hypothesis because the input that English children receive doesn't support all three links in both directions. Generating it as a hypothesis requires adopting the most liberal hypothesis generation procedure that we considered, which posits a bidirectional link between thematic role and syntactic position if a unidirectional link exists in either direction.

Our second finding is that, even assuming the complex form of UTAH can be generated as an explicit hypothesis, it isn't learnable 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 complex 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 complex linking theory of UTAH is available prior to the acquisition process we explored here. If UTAH is instead learned during that 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 (i.e., A-movement) has occurred in order to realize that the surface form is not the original form specified by the linking theory. 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 learnable 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 is unlikely to be learnable, at least given the acquisition approach investigated here. That said, it's potentially learnable as three separate links (AGENT → *subject*, *object* → PATIENT, OTHER → *oblqiue object*). We leave it to future syntactic work to explore the consequences of decomposing UTAH into the three links uncovered here.

In contrast with UTAH, rUTAH appears to be both easily generated by all of the hypothesis generation procedures that we explored here and easily learnable from child-directed input. This in itself may not be particularly surprising, given (i) that 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 an absolute approach to roles and positions (as in UTAH) is likely to lead to too many exceptions for a child using the Tolerance Principle to infer the complex linking theory specified by UTAH. The only complex linking theories that are learnable under plausible acquisition assumptions are those that seek to minimize exceptions to the theory. One example of this is the relative approach to roles and positions leveraged by rUTAH.

For theoreticians who believe that UTAH or rUTAH must be learned (rather than innate), and who are looking to choose between UTAH and rUTAH as the target state for adult grammars, our results lend support against UTAH as the correct target state for adult grammars. This is because our results suggest that (i) rUTAH has learnability advantages over UTAH, and (ii) UTAH can't be learned in its complex form. We note that recent work in the theoretical syntax literature suggests that UTAH may not be operative in all languages (Wood, 2015; Kastner, 2016; Myler, 2016).

This would then indicate that UTAH may not be universal, which in turn suggests UTAH must be learned in the languages it's operative in because there are other target state options. Because of the difficulty in learning UTAH from the data children encounter, UTAH may not *be* learnable and so is therefore unlikely to be the correct target state for adult grammars.

10 Conclusion

Our goal here was to build concrete acquisition theories for UTAH and rUTAH that assumed minimal domain-specific, prior knowledge, and incorporated cognitively-plausible learning mechanisms. We leveraged a conceptual acquisition framework involving explicit hypothesis generation and evaluation, Yang (2005, 2016)'s Tolerance Principle as the cognitively-motivated mechanism for hypothesis evaluation, and linguistically-annotated realistic child-directed input from the CHILDES Treebank (Pearl & Sprouse, 2013, 2018). Using this framework, we found a striking difference between UTAH and rUTAH. UTAH isn't easily generated as an explicit hypothesis in its complex form, and isn't learnable from child-directed speech in its complex form (though it is in fact learnable as three basic links between thematic role and syntactic position). In contrast, rUTAH is easily generated as an explicit hypothesis, and is learnable from child-directed speech. Moreover, these results hold for English children's input at ages three, four, and five.

Our acquisition framework also highlights components necessary for learning linking theories, some of which may be both domain-specific and innate given our current level of 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 explicit hypothesis generation procedure. An interesting open question is whether a way can be found to derive these components from other components. Finally, our results suggest that the Tolerance Principle is a useful, cognitively-grounded evaluation procedure for learning linking theories. 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.

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.

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.

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.

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.

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

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

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. (2018). *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.

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.*

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.

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.

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. (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.

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