Computational cognitive modeling for syntactic acquisition: Approaches that integrate information from multiple places

Lisa Pearl
University of California, Irvine

Abstract

Computational cognitive modeling is a tool we can use to evaluate theories of syntactic acquisition. Here, I review several models implementing theories that integrate information from both linguistic and non-linguistic sources to learn different types of syntactic knowledge. Some of these models additionally consider the impact of factors coming from children's developing non-linguistic cognition. I discuss some existing child behavioral work that can inspire future model-building, and conclude by considering more specifically how to build better models of syntactic acquisition.

1 Introduction

1.1 About computational cognitive modeling for syntactic acquisition

One tool we can use to understand how syntactic acquisition works is computational cognitive modeling. The *computational* part refers to implementing an idea (that is, a theory) very precisely, typically using mathematical techniques that are carried out on computers. The *cognitive* part refers to what the implemented ideas are about, which is some part of human cognition. The *modeling* part refers to the theory itself, which captures (i.e., *models*) some aspect of cognition (here: syntactic acquisition). With this tool of computational cognitive modeling, we can then make a theory about syntactic acquisition concrete enough to evaluate, because the computational cognitive model allows us to generate predictions about children's syntactic behavior that can be evaluated. That is, when we have a computational cognitive model for syntactic acquisition, we have a theory about syntactic acquisition that's implemented precisely enough to evaluate against empirical data.

Importantly, the computational cognitive model serves as a "proof of concept" for a theory. When the model generates predictions that match human behavior (e.g., children's syntactic behavior), this is proof there's at least one way the theory could explain human behavior – which is

the way the theory was implemented in the computational cognitive model. See Pearl (2014, in press) for more detailed discussion about how to interpret computational cognitive model success (and failure).

1.2 Implementing a theory in a computational cognitive model

When we have a theory of syntactic acquisition, how do we implement it in a computational cognitive model? Implementing the model involves several key aspects. First, the model needs to encode relevant prior knowledge and learning abilities the child is supposed to have at this stage of development. This knowledge and these abilities are often assumed implicitly by the acquisition theory. For instance, a syntactic acquisition theory might assume prior knowledge of individual words in the language and the ability to segment speech reliably from the input. Second, the model needs to learn from realistic input. For instance, a model meant to capture syntactic acquisition behavior that occurs at age four should ideally learn from input that children encounter by age four. Third, the model needs to output predictions that connect in some interpretable way to children's behavior. For instance, a model might predict if a child at age four would treat two verbs as being syntactically the same (i.e., appearing in the same syntactic contexts and having the same interpretations of their arguments). Fourth, the model needs to encode learning, which is how the modeled child uses the information from the input to update hypotheses about syntax. Learning is typically the main component specified by the acquisition theory. For instance, a model might attend to the distribution of certain features of the input viewed as relevant (e.g., animacy of verb arguments, syntactic contexts a verb appears in), and then use probabilistic inference to group verbs together that seem similar enough with respect to those relevant features.

So, to sum up, implementing an acquisition theory in a computational cognitive model involves encoding the acquisition theory assumptions (i.e., the prior knowledge assumed, the learning abilities assumed, and how learning proceeds), learning from realistic input estimates, and generating interpretable output that can be evaluated against empirical data from children. This is an approach that the models I'll review below have taken for investigating syntactic acquisition.

1.3 Road map

Below I'll focus on computational cognitive models of syntactic acquisition that integrate information from multiple places, including both linguistic and non-linguistic sources of information. That is, the syntactic acquisition theories implemented by these models assume that syntactic learning proceeds by children attending to information from these different sources, rather than solely syntactic sources. Why discuss this kind of model? To me, these models seem more realistic because children are surrounded by many different types of information and have many different learning goals simultaneously. That is, children aren't ever only learning about syntax; instead, they're learning about syntax and about who's likely to give them a hug and about how to ask for more milk, among many other things. So, non-syntactic sources of information may be particularly salient in any given moment while children are learning about syntax; if these sources of information happen to be helpful for learning about syntax, then children may very well be able to harness those sources to do so.

Moreover, children are likely impacted by non-linguistic factors during acquisition. For instance, cognitive limitations on memory, attention, and executive control can affect how children perceive the information in their input, how they update their internal hypotheses, and how they generate their observable syntactic behavior. In addition, children likely rely on non-linguistic learning mechanisms to update their internal hypotheses, such as probabilistic inference. In fact, all the models of syntactic acquisition that I'll review below rely on probabilistic inference, and so already incorporate this non-linguistic component into their theories of syntactic acquisition.[1]

Here, as mentioned, I'll focus on syntactic acquisition models that also integrate information from non-syntactic sources. I should note that these are selected case studies in syntactic acquisition modeling from my own work, rather than capturing the full range of computational cognitive models that implement this type of syntactic acquisition theory. I'll first review three case studies, whose acquisition theories incorporate conceptual information such as the animacy of an event participant, participant event roles more generally, and components of lexical meaning, as well as incorporating non-linguistic cognitive limitations affecting both input perception and hypothesis updating. These case studies involve the acquisition of syntactic knowledge about linking theories, the passive, and pronoun interpretation.

I'll then briefly review some existing child behavioral work that we can take inspiration from when it comes to building better computational cognitive models of syntactic acquisition. I'll also discuss more specifically how we can think about building better models, and how we can incorporate the insights from both the behavioral work reviewed and current modeling work. I'll conclude with a few other ideas for building better models of syntactic acquisition in the future.

2 Some modeling case studies in syntactic acquisition

For each of the modeling case studies below, I'll first describe the syntactic knowledge children are trying to acquire. I'll then describe the relevant aspects of the acquisition theory implemented in the computational cognitive model, including which information sources are used and how those sources are used to update the modeled child's hypotheses. I'll explicitly highlight which information sources are non-syntactic, as relevant. I'll also describe the input to the model, and how the model's output is evaluated against empirical data from children's behavior.

2.1 Linking theories

The syntactic knowledge. One type of syntactic knowledge is how to interpret a verb's arguments in context. For instance, consider this sentence: *The little girl blicked the kitten on the stairs.* Even if we don't know what *blick* means, we still prefer to interpret this sentence as the little girl doing something (blicking) to the kitten, and that event happening on the stairs. The reason we as adults prefer this interpretation is because we have *linking theories* that link the thematic roles specified by a verb's lexical semantics (e.g., AGENT, PATIENT, LOCATION) to the syntactic

---

[1]See Pearl (in press) for discussion of many other examples of syntactic acquisition models that rely on probabilistic inference, statistical learning, or otherwise "counting things", even if those models learn only from syntactic information.

argument positions specified by that verb's syntactic frame (e.g., subject, direct object, other object such as object of a preposition). 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 (like here with *blick*).

Verbs can be grouped together into classes where the verbs in a class behave the same way with respect to the links between syntactic positions and thematic roles. That is, solving the linking problem (i.e., acquiring linking theories for the verbs of the language) involves learning how to link syntactic positions and thematic roles for different verbs; verb classes are collections of verbs that behave the same way for linking. For example, verbs with "subject-raising" behavior like *appear* and *seem* allow their subject to not have a thematic role. So, in *Lindy seemed/appeared to hug the kitten*, *Lindy* isn't a "seemer" or an "appearer", but rather a kitten-hugger. As another example, verbs with "unaccusative" behavior like *fall* and *break* have a PATIENT in the subject position. So, in *The toy kitten fell/broke*, falling or breaking is happening to the toy kitten. As a third example, verbs with passivizable behavior like *hug* and *break* allow their subject to be a PATIENT in the passive construction, while verbs like *appear*, *seem*, and *fall* don't. That is, *The toy kitten was hugged/broken by Lindy*, with hugging or breaking happening to the toy kitten, is acceptable. In contrast, *The toy kitten was seemed/appeared/fallen by Lindy*, with seeming, appearing, or falling happening to the toy kitten, isn't acceptable.

These examples demonstrate that a verb class can involve many linking behaviors. Here, one verb class involving *fall* might be characterized as +unaccusative and -passivizable; another verb class involving *break* might be characterized as +unaccusative and +passivizable; a third verb class involving *seem* and *appear* might be characterized as +subject-raising and -passivizable. To learn what verbs belong together in a class, children must implicitly develop the linking theory for that verb class. This is why acquiring verb classes can be used as a measure of linking theory development. In short, if a child (and therefore a modeled child) can cluster verbs together into classes that behave the same linking-wise, then the child (real or modeled) can be said to have developed the relevant linking theory knowledge that leads to those verb classes.

The acquisition theory implemented in the model.   Pearl and Sprouse (2019) proposed that children can cluster verbs into appropriate verb classes by paying attention to several pieces of information associated with verbs in their input; argument animacy, syntactic context, and link distribution. To see a concrete example of each, let's consider two of the utterances involving *break* from our examples: the unaccusative *The toy kitten broke* and the passive *The toy kitten was broken by Lindy*. First, the animacy of the verb's arguments matters. For instance, a child would notice that *The toy kitten* is inanimate. Second, the syntactic contexts that a verb appears in matters. So, a child would notice that *break* appeared in an unaccusative context of the form *Noun-Phrase Verb* and a passive context *Noun-Phrase was Verb Preposition Noun-Phrase*. Third, the distribution of links between thematic roles and syntactic positions matters. Here, a child would notice that *break* has the following links in the two utterances above: two instances of PATIENT in subject position (from *The toy kitten* in both utterances) and one instance of AGENT in the prepositional phrase position (from *Lindy* in the passive utterance).

With this information[2], children would then create verb classes by using Bayesian inference, a type of probabilistic learning shown to accord with a variety of developmental patterns across cognition (see Pearl, 2021 for a brief review). When using Bayesian inference, a learner updates hypotheses by balancing prior knowledge or biases against fit to the observed data. For learning verb classes, Pearl and Sprouse (2019) built in a standard type of prior knowledge for learning classes of any kind, which is that fewer classes are preferred. The fit to the observed data is about the child's input: here, if the modeled child assumes a certain set of verb classes, is the information observed in the input about argument animacy, syntactic context, and link distribution more probable? A verb class hypothesis that causes the observed information to be more probable is a better fit than a hypothesis that causes the observed information to be less probable.

Information integrated. The acquisition theory implemented in the model involves integrating several types of information: (i) animacy (non-linguistic), (ii) syntactic contexts (syntactic), and (iii) links between thematic roles (semantic) and syntactic positions (syntactic). These information sources are combined using the non-linguistic learning mechanism of Bayesian inference.

Model input. To generate predictions about verb classes that English-learning children would have, the model learned from verb uses in English child-directed speech samples. Pearl & Sprouse estimated how many verb uses children at different ages (three, four, and five) would encounter, and implemented models that learned from these same quantities. So, for instance, the three-year-old modeled child learned from the amount of verb uses a three-year-old English-learning child would encounter, distributed according to the samples of speech directed to English-learning children up to age three.

Model output and evaluation. To evaluate a modeled child, Pearl & Sprouse compared the verb classes predicted by the modeled child against verb classes that children of the appropriate age seem to have. More specifically, Pearl & Sprouse used 12 types of syntactic or interpretation behavior surveyed from a large collection of child behavioral studies in order to identify verb classes that three-, four-, and five-year-old English children have. These behaviors included subject-raising, unaccusative, and passivizable, among others. From these verb behaviors at ages three to five, Pearl & Sprouse derived age-specific verb classes that a modeled child should attempt to match when it learns from the same data that three-, four-, or five-year-olds learn from. In particular, verbs in the same class are treated the same by children of that age (i.e., the verbs either have or don't have a specific syntactic or interpretation behavior, such as being passivizable). So, the modeled child of that age should cluster those verbs together if it's learned the way children of that age learn.

Pearl & Sprouse found that their modeled three-, four-, and five-year-olds were able to generate verb classes that matched English-learning children's verb classes fairly well. This success supports the acquisition theory implemented in the model, and suggests that children may indeed be learning from these different information types when developing the linking theory knowledge that

---

[2]Pearl & Sprouse's theory also assumed children could potentially have additional biases about how to interpret the link distribution in their input. See Pearl and Sprouse (2019) for details.

leads to their observable verb classes. More specifically, the way English-learning children cluster verbs together during syntactic acquisition aligns with them learning not just from syntactic information (e.g., syntactic contexts), but also non-syntactic information (e.g., animacy and thematic roles).

2.2 Passives

The syntactic knowledge. As mentioned above, the passive structure in English allows the subject to be a PATIENT. For instance, in *The toy kitten was broken by Lindy*, the *The toy kitten* is the one being broken. So, this sentence seems to have a structure more like *The toy kitten was broken* $_{The\ toy\ kitten}$ *by Lindy*, where $_{The\ toy\ kitten}$ marks the position where *The toy kitten* is understood (as the object of *break*).

Children then need to learn that this interpretation is possible, which involves understanding where the element in the subject position is understood (in this case, a position where it can serve as PATIENT). Importantly, not all verbs passivize: recall that *The toy kitten was fallen* isn't acceptable to English speakers (i.e., *fall* doesn't passivize). So, a key learning problem is to learn *which* verbs in English can passivize (i.e., which verbs allow the passive structure and related interpretation with the subject as PATIENT).

Interestingly, there seems to be significant variation in English for when children realize certain verbs are passivizable. Some verbs, such as *hug*, are recognized as young as age three while others, such as *love*, appear delayed till after age five. Moreover, verb meaning (i.e., the lexical semantics) seems to matter. For instance, *hug* is an observable action, and *love* is not; *love* is a psych subject-experiencer verb where the subject experiences the psychological state described (love), while *hug* isn't a psychological verb at all. These and other lexical semantic features have been proposed to impact when English-learning children learn that specific verbs are passivizable. (See Nguyen & Pearl, 2021 for a review of the acquisition trajectory and proposed lexical semantic features.)

In addition, the syntactic feature of transitivity has been proposed as a key indicator that a verb is likely passivizable in English (Levin, 1993). A transitive syntactic context has a subject and direct object, as in *Lindy broke the toy kitten*, with *Lindy* as the subject and *the toy kitten* as the direct object. So, verbs that allow a transitive context, like *break*, are likely to be passivizable in English.

The acquisition theory implemented in the model. Nguyen and Pearl (2019) proposed that children decide whether a verb is passivizable on the basis of (i) several of its lexical semantic features (like being observable or a psych subject-experiencer verb) and potentially the syntactic feature of transitivity, as well as (ii) the input about how often verbs with those features allow passivization. Information about a verb's features are integrated via Bayesian inference.

As before, Bayesian inference balances prior knowledge or biases against fit to the observed data. Here, the prior captures how easy (or difficult) it is for children to deploy their knowledge of the passive in the moment, which can be impacted by immature cognitive development. That is, even if a child knows a specific verb is passivizable, she might not be able to access the passive structure appropriately in the moment after hearing the verb in the passive. So, she might not use

her syntactic knowledge of the passive structure for that verb instance.

The fit to the observed data is again about the child's input. In particular, the modeled child assumes passivization is based on a verb's features and the frequencies of those features in passive forms. Is the information observed in the input about how often verbs with certain features passivize more or less probable? If the verbs in the input are more probable, then there's a good fit to the observed data.

Importantly, the modeled child can heed or ignore any given feature when deciding if a particular verb is passivizable. So, for instance, a five-year-old might ignore whether a verb is an observable action, and instead key into whether it encodes a psychological state. The acquisition theory implemented in the model of Nguyen & Pearl explored theories of selective learning for the English passive (i.e., selectively ignoring available information when deciding if a verb is passivizable).

Information integrated. The information integrated via Bayesian inference is the selected features of a verb (syntactic and lexical semantic), whatever those happen to be. Notably, these features will be the ones children attend to for all the verbs of the language (rather than a feature set for each verb or type of verb). So, the acquisition theory assumes both syntactic and non-syntactic information is relevant. These information sources are then combined using the non-linguistic learning mechanism of Bayesian inference.

Model input. The model learned from verb uses in English child-directed speech samples, both passive uses like *The toy kitten was broken* and active uses like *The toy kitten broke*.

Model output and evaluation. To evaluate a modeled learner attending to some set of features, Nguyen & Pearl looked at the age when children have been observed to correctly interpret or produce the passive of a verb more than half the time in previous child behavioral experiments. They called this age the *age of acquisition* (**AoA**) for the passive of that verb, and Nguyen & Pearl used the AoA of 30 verbs as a model target. They focused on age five, and therefore split the 30 verbs into verbs whose AoA was five or younger versus verbs whose AoA was older.

The modeled learner predicts a specific verb is either passivizable or not at a certain age, on the basis of its input. So, the modeled five-year-old learned from the distribution of verb input that English-learning five-year-olds encounter and predicted which verbs would be passivizable. Nguyen & Pearl found that a modeled five-year-old who ignored many of the available features was able to match the behavior of English-learning five-year-olds, and passivize the subset of verbs whose AoA was five or younger. This modeled child instead focused on the syntactic feature of transitivity and a single lexical semantic feature.[3] These modeling results thus suggest that English five-year-old passivization behavior can be captured if five-year-olds selectively attend to syntactic and lexical semantic features in their input.

---

[3]This lexical semantic feature was psych object-experiencer, where the object of the verb experiences the psychological state. An example is *annoy*: In *The non-stop crying annoyed Lisa*, the object *Lisa* is experiencing the psychological state of being annoyed.

2.3   Pronoun interpretation

The syntactic knowledge.   Consider this English sentence: *Lisa sang to the triplets and then* PRONOUN *took a nap.* How we interpret PRONOUN depends on several factors. One is agreement information: If the pronoun is the singular *she*, we look for a singular antecedent like *Lisa*; if the pronoun is the plural *they*, we look for a plural antecedent like *the triplets*. Another factor is our discourse-level knowledge about the lexical items that connect the two clauses together, such as *and then*. In languages like Spanish, the equivalent to *and then* biases the interpretation towards the subject *Lisa* rather than the object *the triplets*. Another factor in languages like Spanish is whether the pronoun is overt (i.e., pronounced) or not; Spanish is a language that allows the pronoun not to be pronounced. When it's not pronounced, the subject (e.g., *Lisa*) tends to be favored as the pronoun's antecedent. (See Pearl and Forsythe (2022) for a brief overview of these factors in pronoun interpretation.) Children need to learn how to interpret pronouns of their language in context, taking these factors (and others) into account the way adult speakers of their language do.

The acquisition theory implemented in the model.   Pearl & Forsythe (Forsythe & Pearl, 2020; Pearl & Forsythe, 2022) proposed that Spanish-learning children decide how to interpret a pronoun in context by potentially considering information from their input about agreement, lexical connective items, and whether the pronoun is overt. These information sources are integrated via Bayesian inference.

As before, Bayesian inference balances prior knowledge or biases against fit to the observed data. Here, the prior encodes how often a pronoun preferred a particular antecedent in children's input, irrespective of any other useful information about how to interpret that pronoun. The fit to the observed data is about how often each information type occurs in children's input when a pronoun has a particular interpretation. If certain information (e.g., singular agreement information) almost always occurs when a pronoun's antecedent is interpreted a certain way (e.g., a singular antecedent), then using that highly-reliable information to interpret the pronoun is a good fit.

In addition, Pearl & Forsythe included two cognitive limitation factors in their model. One factor was whether the information in the input was represented accurately or not, with the idea that cognitive limitations could cause the available information to be represented inaccurately. If that happened, the information children would use to decide a pronoun's interpretation (via Bayesian inference) would be inaccurate. The second factor was whether the represented information, whether accurate or inaccurate from each source, was always used or instead ignored in the moment. That is, cognitive limitations could cause a child to temporarily ignore information that could be useful for pronoun interpretation (i.e., not deploy it). If that happened, some information would be missing when children use Bayesian inference to decide a pronoun's interpretation. Thus, the theory implemented in the model of Pearl & Forsythe considered certain theories for children's pronoun interpretation behavior that involve cognitive limitations.

Information integrated.   The information integrated via Bayesian inference is linguistic: agreement information (morphology), the lexical connectives between clauses (lexical), and whether the pronoun is pronounced (syntactic/phonological). These information sources are then combined using the non-linguistic learning mechanism of Bayesian inference. The way the infor-

mation is combined can be mediated by non-linguistic factors arising from cognitive limitations: mis-representing the information from the input and/or not deploying select information during Bayesian inference.

Model input.   The modeled child learned from pronoun uses in Spanish speech samples involving children. These pronoun uses involved two clauses and had the pronoun as the subject of the second clause (e.g., *[Lisa sang to the triplets]$_{clause_1}$ and then [*PRONOUN *took a nap]$_{clause_2}$.*)

Model output and evaluation.   Pearl & Forsythe evaluated modeled children that attended to this set of linguistic features and potentially had cognitive limitations impacting information representation and/or deployment. The modeled children generated predictions for how to interpret pronouns that Spanish-learning children ages three to five had interpreted in different experimental contexts involving information about agreement, lexical connectives, and whether the pronoun was pronounced.

Pearl & Forysthe found that modeled three-, four-, and five-year-olds were able to best match the interpretation preferences of actual three-, four-, and five-year-olds when either of the two cognitive limitations (but not both) were active. That is, children's interpretation behavior could be captured by integrating information from agreement, lexical connectives, and whether the pronoun was pronounced as long as children either (i) always mis-perceived information from these sources in the input, leading to inaccurate information, or (ii) often ignored accurate information from these sources when deciding how to interpret a pronoun in the moment. Importantly, children's behavior wasn't captured as well if the modeled child had both limitations (inaccurate information often ignored) or neither limitation (accurate information never ignored). These modeling results thus offer specific explanations about how cognitive limitations could impact children's pronoun interpretation preferences, if children rely on these linguistic information sources.

3   Some experimental work to take inspiration from

I want to briefly now turn to some work from child behavioral experiments that can provide some inspiration for other factors we might want to consider (or consider further) for syntactic acquisition. The first set of experiments involves cognitive limitations, while the second involves knowledge about pragmatics and the world more generally.

3.1   Cognitive limitations

The model of Forsythe & Pearl highlighted one effect that cognitive limitations could have on children's acquisition (syntactic or otherwise): children have adult-like knowledge but can't deploy it effectively in the moment. Several child behavioral experiments demonstrate this effect for syntactic acquisition, including Gerard, Lidz, Zuckerman, and Pinto (2018), Ud Deen et al. (2018), and Liter, Grolla, and Lidz (2022).

In Gerard et al. (2018), four- and five-year-old English-learning children were asked to interpret utterances with unpronounced subject pronouns in the second clause, like *Dora washed Diego*

before eating a red apple. An adult-like interpretation is that Dora is the one eating a red apple, so the syntactic representation is something like this: *Dora washed Diego before* PRONOUN$_{Dora}$ *eating a red apple.* Children were asked to interpret this kind of utterance in tasks that were either more or less cognitively-demanding. A more cognitively-demanding task might involve children having to hold additional information in mind and also evaluate whether the utterance itself is true; a less cognitively-demanding task would involve children simply indicating their interpretation by coloring a picture of the appropriate interpretation (i.e., Dora eating the apple, rather than Diego). When children had to do the more cognitively-demanding task – and so use up more cognitive resources on something besides interpreting the unpronounced pronoun – they gave more non-adult-like interpretations (e.g., Diego eating the apple). In contrast, when children did the less cognitively-demanding task – and so focused more cognitive resources on interpreting the unpronounced pronoun – they gave more adult-like interpretations (e.g., Dora eating the apple). One way to interpret these results is that four- and five-year-olds have adult-like knowledge of how to interpret these unpronounced pronouns, but can't always deploy that knowledge in the moment when their cognitive resources are being used up by other things. This idea aligns broadly with the Forsythe & Pearl modeled children who can't accurately deploy their information about pronoun interpretation in the moment.

Another example comes from Ud Deen et al. (2018) on children's interpretation of the passive. English-learning four-year-olds correctly interpreted passives like *Elephant was surprised by Monkey* more often when the utterance was simply repeated. One interpretation of this finding is that children can adjust their mistaken expectations about the thematic role associated with the subject (i.e., that Elephant isn't the surprise-causer but instead the surprise-experiencer) when they hear the sentence again because they know they made a mistake the first time. That is, children can inhibit the incorrect thematic role assignment of Elephant because they know it's not going to be correct. However, the first time children hear the utterance, they don't know this and so they make an incorrect assignment (e.g., of Elephant as surprise-causer), which is hard for them to adjust afterwards. In other words, children have adult-like knowledge about how to interpret the passive, but can't deploy it effectively when their cognitive inhibition ability isn't strong enough. So, more broadly, domain-general cognitive factors like immature cognitive inhibition impact children's ability to deploy their knowledge of the passive.

A third example comes from Liter et al. (2022), and also involves immature cognitive inhibition, this time impacting children's production of questions involving *wh*-words like *where*. More specifically, English-learning children will sometimes produce "medial *wh*" questions that seem to duplicate the *wh*-word, with an extra copy appearing in the middle, such as *Where do you think where they were walking?* Liter et al. (2022) found that children's production of medial-*wh* questions correlated with a measure of their cognitive inhibition abilities. One way to interpret this is that children do in fact know that English doesn't allow medial *wh*, but simply lack the cognitive control sometimes to inhibit the extra *wh*-word from being produced in the moment. As with the passive example above, this result highlights that acquisition theories (and therefore the computational cognitive models we build to explain children's behavior) need to consider the non-linguistic systems controlling cognitive inhibition in children.

3.2   Pragmatics and world knowledge

Other sources of information children could harness involve knowledge about how speakers use their language (i.e., pragmatic knowledge) and knowledge about the world more generally. We already have behavioral evidence that children can rely on these information sources during syntactic acquisition, such as when learning to interpret pronouns (e.g., Wykes, 1981; Song & Fisher, 2005, 2007; Pyykkönen, Matthews, & Järvikivi, 2010; Hartshorne, Nappa, & Snedeker, 2015, among others).

As one example of pragmatic knowledge with pronouns, let's consider the sentence *Lisa sang to Lindy and then she took a nap.* The pronoun *she* could refer to either *Lisa* or *Lindy*, but adults know that speakers like to have clauses refer to the same topic (Asher & Lascarides, 2003). This leads to a "first-mention bias", where the element first mentioned (e.g., the subject *Lisa*) is the topic and listeners prefer a subsequent pronoun to refer to that first-mentioned element (Crawley, Stevenson, & Kleinman, 1990; Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Järvikivi, van Gompel, Hyönä, & Bertram, 2005). English-learning children ages three to five also seem to have this pragmatic knowledge, leading to a first-mention bias in a variety of contexts (Song & Fisher, 2005, 2007; Pyykkönen et al., 2010; Hartshorne et al., 2015).

As one example of world knowledge with pronouns, let's consider this sentence pair: *Jane needed Susan's pencil. She gave it her.* Knowledge about how the world works allows listeners to pick situationally-appropriate interpretations (e.g. Hobbs, 1979; Kehler, Kertz, Rohde, & Elman, 2008). Here, if Jane needs a pencil, she can't already have one, so she can't be the one to give a pencil away. That means that the one doing the giving (referred to by *She* in the second sentence) must not be Jane, and instead is probably the other mentioned person *Susan*. Similarly, if Jane needs a pencil, she's likely to be the one getting a pencil from someone else, i.e., the recipient of giving indicated by *her*. So, world knowledge allows listeners to interpret *She* as *Susan* and *her* as *Jane*. English-learning five-year-olds seem able to complete this chain of reasoning and correctly interpret the second sentence (Wykes, 1981).

These are just select examples of pragmatic and world knowledge impacting pronoun interpretation, which of course is simply one aspect of syntactic knowledge. More generally, these examples suggest that future syntactic acquisition theories (and the computational cognitive models implementing them) should consider these information sources.

4   Moving forward

Computational cognitive modeling is a tool that complements other techniques for investigating language development, providing insight into aspects of language acquisition that can be difficult to investigate otherwise. For instance, the models reviewed here investigated how children might learn certain syntactic knowledge from their input (verb constructions like subject-raising, unaccusatives, and passives) and why child behavior may differ from adult behavior for certain syntactic elements (pronoun interpretation). In general, I think questions of *how* acquisition works and *why* children behave as they do are much easier to investigate with modeling.

Informative models typically build on data collected with other techniques. Model input is

based on estimates of the information children encounter in their language interactions. Model learning mechanisms are based on ideas of what abilities and learning biases children demonstrate at certain ages. Model output is based on data collected from children (or that can be collected in the future), so that the model can explain children's observed linguistic behavior.

As we move forward, a basic goal is to build "better" models – that is, models that capture more of the relevant aspects of the acquisition process so that we can better link children's input to their observable behavior. When we have these better models, we then have better explanations – as implemented in the models – for why acquisition (syntactic or otherwise) proceeds the way it does. So, how do we build better models?

4.1   Building better models

To build a computational cognitive model of language acquisition, we need to be very precise about the acquisition process the model is implementing. One concrete proposal for the relevant components of the acquisition process is in Figure 1, adapted from Pearl (in press). This proposal specifies components both external and internal to the child during the acquisition process, and is meant to capture the iterative process of acquisition unfolding over time.

External components are observable. We can observe the input signal available to children (e.g., the child language interactions they experience). For example, consider a version of our utterance from before: *"Lisa sang to the triplets and then she took a power nap."* The input signal is the physical signal in the world, such as auditory components like pitch, timbre, and loudness of the utterance. The input can also include other aspects of the environment, such as who said the utterance, where they said it, when they said it, and what people or objects were in the environment at the time.

We can also observe children's behavior at any stage of development, either through naturalistic productions and behavior or clever experimental designs that elicit productions or behavior. In the example utterance above, we can observe who the child thinks *she* refers to, Lisa or the triplets. One way to do this is to present the child with two pictures, one of Lisa napping and one of the triplets napping, and ask the child to point to the picture the utterance describes.

The internal components of the acquisition process involve several pieces. The first piece concerns the information the child is able to perceive in the input signal. In particular, *perceptual encoding* involves extracting information from the input signal to create the *perceptual intake*. Perceptual encoding draws on the child's developing knowledge and systems to extract information. For instance, in our example utterance, the child may be able to perceive syllables (e.g., /li/, /sə/, /sɛŋ/, etc.), words (e.g., *Lisa*, *sang*, etc.), syntactic structure (e.g., [$_{IP}$ Lisa [$_{VP}$ sang [$_{PP}$ to [$_{NP}$ the triplets]]]]), pronoun interpretations (*she=Lisa*), as well as the event participants (Lisa, the triplets) and properties of the events described (singing, napping), among many other types of information. What children can perceive depends on what they know about their language (e.g., developing linguistic knowledge: *Lisa*, *the triplets*, and *she* are words), what they know about the world (e.g., developing non-linguistic knowledge: who's likely to take a power nap), and how well they can extract information of different kinds (e.g., developing linguistic systems: speech segmentation, syntactic parsing, pronoun interpretation biases; developing non-linguistic systems: memory, cognitive inhibition). Notably, extracting information from the input signal involves ig-
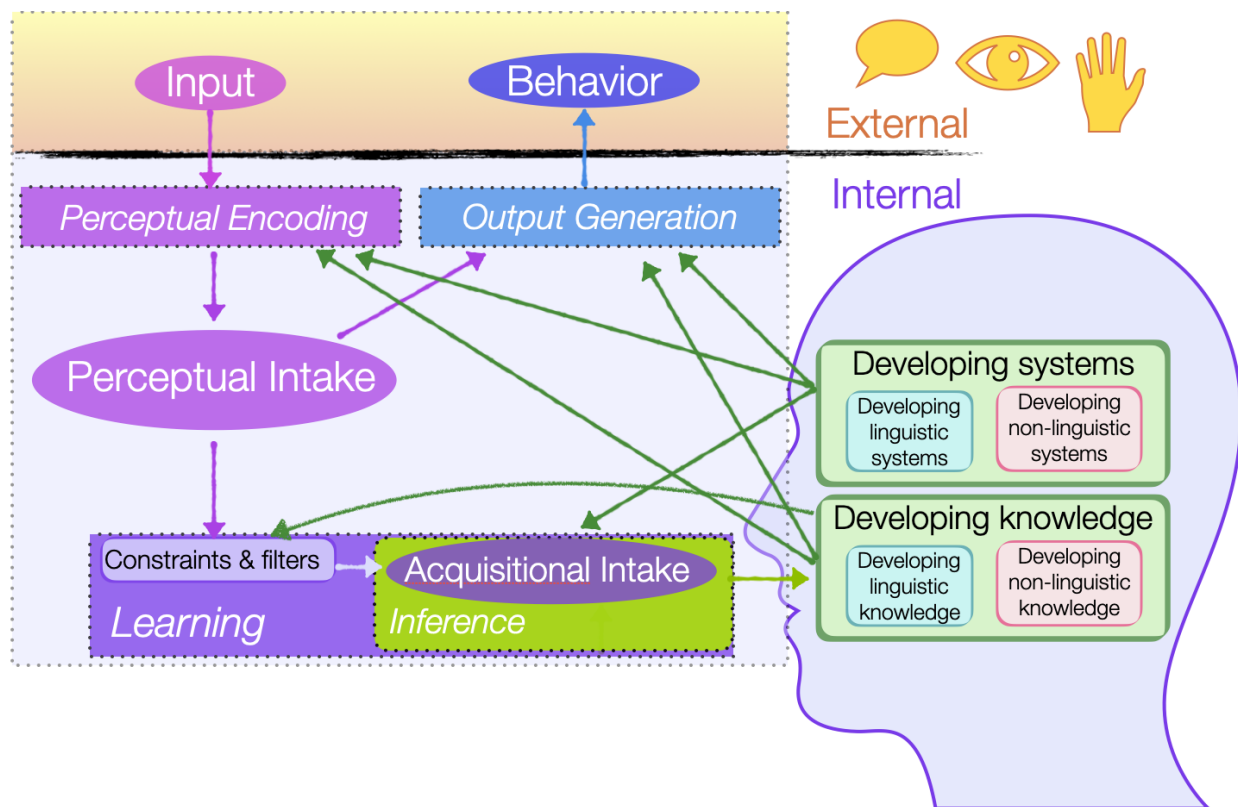
Figure 1: Proposal for the relevant components of the acquisition process that a computational cognitive model of language acquisition should consider. External components (input and behavior) are observable. Internal components aren't observable, and include perceptually encoding information from the input signal (yielding the perceptual intake), generating output from the encoded information (yielding observable behavior), and learning from the encoded information (using constraints & filters to yield the acquisitional intake, and doing inference over that intake). The developing systems and developing knowledge (both linguistic and non-linguistic) impact all internal components, while the learning component updates the developing knowledge.

noring information present (e.g., where the utterance was spoken) and adding information not explicitly present (e.g., where the words are, how a pronoun is interpreted). What children ignore and add depends on their developing knowledge and developing systems.

The second internal piece concerns how children generate their observable behavior. For this, children rely on the information they've been able to perceptually encode (the perceptual intake) and their developing systems and knowledge. In particular, children apply their production systems to the perceptual intake in order to generate behavior like speaking (which relies on linguistic systems and non-linguistic systems involved in utterance generation). In our example utterance, a child might say "Lisa's the one napping". Children can also respond non-verbally (e.g., look at a picture that encodes a scene described by the utterance, which relies on non-linguistic systems like motor control, attention, and decision-making). In our example utterance, a child might look at the

picture of Lisa napping.

The last internal piece concerns learning, which is how the child's developing knowledge (both linguistic and non-linguistic) is updated over time. As with the other internal pieces, the child's developing systems and knowledge impact this piece. In particular, learning occurs over the part of the perceptual intake the child deems relevant to learn from: this is the *acquisitional intake*. The acquisitional intake is typically *not* all of the perceptual intake. That is, it's not everything the child is able to encode. Instead, depending on what the child is trying to learn, what's relevant is likely some subset of the perceptual intake. For instance, in our example utterance, the fact that the pronoun *she* is singular may be in the acquisitional intake, while the fact that *she* is a separate word from *took* may not.

The child's developing knowledge can filter the perceptual intake down to the relevant information by providing both constraints on possible hypotheses (i.e., what options are worth considering) and attentional filters (i.e., what in the information signal to pay attention to). For instance, in our pronoun interpretation example, a linguistic constraint may limit the possible hypotheses for *she*'s antecedent to noun phrases, and so the number feature is relevant for choosing among different noun phrases; a non-linguistic constraint may limit potential antecedents to animate participants who are capable of power napping. An attentional filter may focus the child on the pronoun's interpretation, rather than other aspects of the utterance, because of uncertainty about how to interpret pronouns more generally at the child's current stage of development.

Inference then operates over the acquisitional intake, and typically involves non-linguistic abilities like probabilistic inference, statistical learning, or hypothesis testing. The result of this inference can be used to update the developing knowledge – potentially both linguistic knowledge and non-linguistic knowledge. For instance, in our pronoun interpretation example, the child might update her hypotheses about how likely it is that *she*'s antecedent is singular (linguistic knowledge) and how likely adults like Lisa are to take power naps (non-linguistic knowledge).

With this proposal in hand for relevant components of a computational cognitive model of acquisition, we can now think about some of the ideas we might want to incorporate into future models of syntactic acquisition. I'll briefly discuss some ideas for incorporating non-syntactic components and simultaneous acquisition of different knowledge aspects.

4.2   Incorporating non-syntactic components into acquisition models

Prior behavioral work has found that children are sensitive to animacy when learning aspects of syntax (e.g., see Becker, 2015). Pearl and Sprouse (2019) used animacy in their model of linking theory acquisition, allowing the animacy of a verb's arguments to be part of the acquisitional intake that children learned from.

Prior behavioral work has also found that children can use both pragmatic and world knowledge to help them choose between potential interpretations of pronouns (e.g., Wykes, 1981; Song & Fisher, 2005, 2007; Pyykkönen et al., 2010; Hartshorne et al., 2015). Some recent computational cognitive modeling work has investigated how children choose between potential interpretations of utterances like *Every horse didn't jump*, which can either mean "No horses jumped" or "Not all horses jumped" (Savinelli, Scontras, & Pearl, 2017, 2018; Scontras & Pearl, 2021). The modeled children in these studies incorporated both pragmatic knowledge about what speakers think the

topic of conversation is and world knowledge about the event described (e.g., how likely horses are to jump) into the perceptual intake. Notably, differences in children's ability to adjust their expectations about the pragmatics and world of the experiment – due to immature non-linguistic systems – can explain children's observed non-adult-like behavior, according to these models.

More generally, prior behavioral work (Gerard et al., 2018; Ud Deen et al., 2018; Liter et al., 2022) has noted the impact of immature non-linguistic systems (e.g., cognitive inhibition) in children's deployment of their knowledge – that is, how children generate their observed behavior in experimental contexts. So, I think it's useful for future computational cognitive models to consider the impact of these developing non-linguistic systems when accounting for children's behavior (i.e., the output generation process).

Moreover, these developing non-linguistic systems may also impact several other pieces of the acquisition process: (i) perceptual encoding, leading to a perceptual intake that captures immature representations of information in the input, (ii) constraints & filters, leading to an acquisitional intake that's inaccurate, and (iii) inference, leading to learning that's non-adult-like. The exact way developing non-linguistic systems impact these pieces depends on what system is developing and how that system is proposed to contribute to the acquisition process. While this is certainly non-trivial to specify for any given non-linguistic system and model piece, the more we can do it, the better we're able to capture the acquisition process in children and link their input to their observable behavior with a concrete acquisition theory encoded in a model.

4.3   Thinking about simultaneous acquisition

Another interesting consideration is simultaneous acquisition, where multiple types of knowledge are learned simultaneously. In the case studies discussed here, the acquisition of linking theories from Pearl and Sprouse (2019) was an example of this. More specifically, when learning how to cluster verbs together into classes whose linking theories were similar, the modeled child effectively learned about many different verb constructions simultaneously (e.g., which verbs are subject-raising, which verbs are unaccusative, which verbs are passivizable, etc.). The key insight is that the modeled child's objective was broad – learn about verbs that "behave" similarly with respect to certain types of information in the acquisitional intake (argument animacy, syntactic contexts, links between thematic roles and syntactic positions), instead of learning about which verbs allow a specific syntactic behavior (e.g., subject-raising). In other words, the specific syntactic knowledge about which constructions any given verb allows is a by-product of trying to learn something else about that verb, namely which other verbs it behaves similarly to (i.e., which class it belongs to) and what the behavior of that verb class is.

I think this may be a more realistic approach to syntactic acquisition (and acquisition more generally), with children trying to learn about their language more broadly and picking up specific linguistic knowledge along the way as part of that broader learning goal. What this means modeling-wise is that the modeled child's objective – what hypotheses are being considered – would be adjusted. For instance, instead of *explicitly* learning if a verb is subject-raising, can children's observable behavior about which verbs are subject-raising be captured by a modeled child learning about verb classes more generally and *implicitly* learning which verbs are subject-raising? This approach worked well for Pearl and Sprouse (2019).

15

Another example of simultaneous syntactic acquisition from my own research (Pearl & Sprouse, 2013; Bates & Pearl, 2019; Dickson, Pearl, & Futrell, 2022; Pearl & Bates, in press) is the acquisition of knowledge about syntactic islands where children, for example, learn that *Who did Lily think the kitten for __who was cute?* isn't a good *wh*-question. Here, the modeled child's objective is to learn in general how to represent *wh*-dependencies like those in *wh*-questions, rather than learning how good a specific *wh*-dependency is (or isn't). By learning to do this, modeled children learn to have adult-like preferences about how good different *wh*-dependencies are (Pearl & Sprouse, 2013; Bates & Pearl, 2019; Pearl & Bates, in press), especially if the modeled children are trying to represent *wh*-dependencies in an "efficient" way (Dickson et al., 2022) that makes processing future *wh*-dependencies easier.

A related approach gaining momentum in syntactic acquisition modeling involves simply learning to predict the next word, with the modeled children implicitly learning whatever knowledge is necessary to make that next word highly probable (and therefore easier to process). Along the way, several models of this type implicitly learn a variety of syntactic knowledge, including knowledge about syntactic islands (e.g., Wilcox, Levy, Morita, & Futrell, 2018; Futrell et al., 2019; Chaves, 2020; Warstadt et al., 2020; Wilcox, Futrell, & Levy, 2021).

5   Conclusion

Here I hope to have shown how computational cognitive modeling can inform our understanding of syntactic acquisition by implementing theories of acquisition precisely enough to evaluate against empirical data from children. I reviewed some previous models that consider information from non-syntactic sources and the impact of non-linguistic cognitive development on syntactic acquisition. I also highlighted some behavioral work that notes the role of other information sources children use and specific cognitive limitations children have during syntactic acquisition. I then discussed how we might build future models that incorporate these insights and so provide better explanations of syntactic acquisition. With this information in mind, I believe we can create, evaluate, and refine better theories of syntactic acquisition through computational cognitive modeling.

References

Arnold, J., Eisenband, J., Brown-Schmidt, S., & Trueswell, J. C. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*, B13-B26.

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Bates, A., & Pearl, L. (2019). *What do you think that happens?* A quantitative and cognitive modeling analysis of linguistic evidence across socioeconomic status for learning syntactic islands. In Megan M. Brown and Brady Dailey (Ed.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 42–56). Somerville, MA: Cascadilla Press.

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

Chaves, R. P. (2020). What Don't RNN Language Models Learn About Filler-Gap Dependencies? *Proceedings of the Society for Computation in Linguistics*, *3*(1), 20–30.

Crawley, R. A., Stevenson, R., & Kleinman, D. (1990). The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research*, *19*(4), 245-264.

Dickson, N., Pearl, L., & Futrell, R. (2022). Learning constraints on *wh*-dependencies by learning how to efficiently represent *wh*-dependencies: A developmental modeling investigation with Fragment Grammars. *Proceedings of the Society for Computation in Linguistics*, *5*(1), 220–224.

Forsythe, H., & Pearl, L. (2020). Immature representation or immature deployment? Modeling child pronoun resolution. In *Proceedings of the Society for Computation in Linguistics 3* (pp. 22–26).

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*.

Gerard, J., Lidz, J., Zuckerman, S., & Pinto, M. (2018). The acquisition of adjunct control is colored by the task. *Glossa: A Journal of General Linguistics*, *3*(1), 1-22. Retrieved from https://doi.org/10.5334/gjgl.547 doi: 10.5334/gjgl.547

Hartshorne, J. K., Nappa, R., & Snedeker, J. (2015). Development of the First-Mention Bias. *Journal of Child Language*, *42*(2).

Hobbs, J. R. (1979). Coherence and coreference. *Cognitive Science*, *3*, 67-90.

Järvikivi, J., van Gompel, R., Hyönä, J., & Bertram, R. (2005). Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science*, *16*(4), 260–264.

Kehler, A., Kertz, L., Rohde, H., & Elman, J. L. (2008). Coherence and coreference revisited. *Journal of Semantics*, *25*(1), 1–44.

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

Liter, A., Grolla, E., & Lidz, J. (2022). Cognitive inhibition explains children's production of medial *wh*-phrases. *Language Acquisition*. Retrieved from https://doi.org/10.1080/10489223.2021.2023813 doi: 10.1080/10489223.2021.2023813

Nguyen, E., & Pearl, L. (2019). Using Developmental Modeling to Specify Learning and Representation of the Passive in English Children. In M. M. Brown & B. Dailey (Eds.), *Proceedings of the Boston University Conference on Language Development 43* (pp. 469–482). Somerville, MA: Cascadilla Press.

Nguyen, E., & Pearl, L. (2021). The link between lexical semantic features and children's comprehension of English verbal *be*-passives. *Language Acquisition*, *28*(4), 433–450.

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

Pearl, L. (2021). Theory and predictions for the development of morphology and syntax: A Universal Grammar+ statistics approach. *Journal of Child Language*, *48*(5), 907–936.

Pearl, L. (in press). Modeling syntactic acquisition. In J. Sprouse (Ed.), *Oxford handbook of experimental syntax.* Oxford University Press.

Pearl, L., & Bates, A. (in press). A new way to identify if variation in children's input could be developmentally meaningful: Using computational cognitive modeling to assess input across socio-economic status for syntactic islands. *Journal of Child Language.* Retrieved from https://ling.auf.net/lingbuzz/004687

Pearl, L., & Forsythe, H. (2022). Inaccurate representations, inaccurate deployment, or both? Using computational cognitive modeling to investigate the development of pronoun interpretation in Spanish. *lingbuzz.* Retrieved from https://ling.auf.net/lingbuzz/006141

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

Pearl, L., & Sprouse, J. (2019). Comparing solutions to the linking problem using an integrated quantitative framework of language acquisition. *Language*, *95*(4), 583–611. Retrieved from https://ling.auf.net/lingbuzz/003913

Pyykkönen, P., Matthews, D., & Järvikivi, J. (2010). Three-year-olds are sensitive to semantic prominence during online language comprehension: A visual world study of pronoun resolution. *Language and Cognitive Processes*, *25*(1), 115-129.

Savinelli, K., Scontras, G., & Pearl, L. (2017). Modeling scope ambiguity resolution as pragmatic inference: Formalizing differences in child and adult behavior. In *Proceedings of the 39th annual meeting of the Cognitive Science Society.* London, UK: Cognitive Science Society.

Savinelli, K., Scontras, G., & Pearl, L. (2018). Exactly two things to learn from modeling scope ambiguity resolution: Developmental continuity and numeral semantics. In *Proceedings of the 8th workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)* (pp. 67–75).

Scontras, G., & Pearl, L. S. (2021). When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: A journal of General Linguistics*, *6*(1). doi: https://doi.org/10.16995/glossa.5724

Song, H.-J., & Fisher, C. (2005). Who's "She?": Discourse prominence influences preschooler's comprehension of pronouns. *Journal of Memory and Language*, *52*(1), 29-57.

Song, H.-J., & Fisher, C. (2007). Discourse prominence effects on 2.5-year-old children's interpretation of pronouns. *Lingua*, *117*(11), 1959-1987.

Ud Deen, K., Bondoc, I., Camp, A., Estioca, S., Hwang, H., Shin, G.-H., . . . Zhong, J. (2018). Repetition brings success: Revealing knowledge of the passive voice. In *Proceedings of the 42nd annual boston university conference on language development* (Vol. 42, pp. 200–213). Cascadilla Press.

Warstadt, A., Parrish, A., Liu, H., Mohananey, A., Peng, W., Wang, S.-F., & Bowman, S. R. (2020). BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics*, *8*, 377–392.

Wilcox, E., Futrell, R., & Levy, R. (2021). Using computational models to test syntactic learnability. *https://ling.auf.net/lingbuzz/006327*.

Wilcox, E., Levy, R., Morita, T., & Futrell, R. (2018). What do RNN language models learn about filler-gap dependencies? In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyz-*

ing and Interpreting Neural Networks for NLP. Association for Computational Linguistics.

Wykes, T. (1981). Inference and Children's Comprehension of Pronouns. *Journal of Experimental and Child Psychology*, *32*, 264-278.