Bootstrapping into Filler-Gap: An Acquisition Story

Marten van Schijndel  Micha Elsner
The Ohio State University
{vanschm,melsner}@ling.ohio-state.edu

Abstract

Analyses of filler-gap dependencies usually involve complex syntactic rules or heuristics; however recent results suggest that filler-gap comprehension begins earlier than seemingly simpler constructions such as ditransitives or passives. Therefore, this work models filler-gap acquisition as a byproduct of learning word orderings (e.g. SVO vs OSV), which must be done at a very young age anyway in order to extract meaning from language. Specifically, this model, trained on part-of-speech tags, represents the preferred locations of semantic roles relative to a verb as Gaussian mixtures over real numbers.

This approach learns role assignment in filler-gap constructions in a manner consistent with current developmental findings and is extremely robust to initialization variance. Additionally, this model is shown to be able to account for a characteristic error made by learners during this period (e.g. A and B gorped interpreted as A gorped B).

1 Introduction

The phenomenon of filler-gap, where the argument of a predicate appears outside its canonical position in the phrase structure (e.g. [the apple] that the boy ate $t_i$, or [what] did the boy eat $t_i$), has long been an object of study for syntacticians (Ross, 1967) due to its apparent processing complexity. Such complexity is due, in part, to the arbitrary length of the dependency between a filler and its gap (e.g. [the apple], that Mary said the boy ate $t_i$).

Recent studies indicate that comprehension of filler-gap constructions begins around 15 months (Seidl et al., 2003; Gagliardi et al., 2014). This finding raises the question of how such a complex phenomenon could be acquired so early since children at that age do not yet have a very advanced grasp of language (e.g. ditransitives do not seem to be generalized until at least 31 months; Goldberg et al. 2004, Bello 2012). This work shows that filler-gap comprehension in English may be acquired through learning word orderings rather than relying on hierarchical syntactic knowledge.

This work describes a cognitive model of the developmental timecourse of filler-gap comprehension with the goal of setting a lower bound on the modeling assumptions necessary for an ideal learner to display filler-gap comprehension. In particular, the model described in this paper takes chunked child-directed speech as input and learns orderings over semantic roles. These orderings then permit the model to successfully resolve filler-gap dependencies. Further, the model presented here is also shown to initially reflect an idiosyncratic role assignment error observed in development (e.g. A and B kradded interpreted as A kradded B; Gertner and Fisher, 2012), though after training, the model is able to avoid the error. As such, this work may be said to model a learner from 15 months to between 25 and 30 months.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{timeline.png}
\caption{The developmental timeline of subject (Wh-S) and object (Wh-O) wh-clause extraction comprehension suggested by experimental results (Seidl et al., 2003; Gagliardi et al., 2014). Parentheses indicate weak comprehension. The final row shows the timeline of 1-1 role bias errors (Naigles, 1990; Gertner and Fisher, 2012). Missing nodes denote a lack of studies.}
\end{figure}

\footnote{This model does not explicitly learn gap positions, but rather assigns thematic roles to arguments based on where those arguments are expected to manifest. This approach to filler-gap comprehension is supported by findings that show people do not actually link fillers to gap positions but instead link the filler to a verb with missing arguments (Pickering and Barry, 1991)}
2 Background

The developmental timeline during which children acquire the ability to process filler-gap constructions is not well-understood. Language comprehension precedes production, and the developmental literature on the acquisition of filler-gap constructions is sparsely populated due to difficulties in designing experiments to test filler-gap comprehension in preverbal infants. Older studies typically looked at verbal children and the mistakes they make to gain insight into the acquisition process (de Villiers and Roeder, 1995).

Recent studies, however, indicate that filler-gap comprehension likely begins earlier than production (Seidl et al., 2003; Gagliardi and Lidz, 2010; Gagliardi et al., 2014). Therefore, studies of verbal children are probably actually testing the acquisition of production mechanisms (planning, motor skills, greater facility with lexical access, etc.) rather than the acquisition of filler-gap. Note that these may be related since filler-gap could introduce greater processing load which could overwhelm the child’s fragile production capacity (Phillips, 2010).

Seidl et al. (2003) showed that children are able to process wh-extractions from subject position (e.g. [who t ate pie]) as young as 15 months while similar extractions from object position (e.g. [what t did the boy eat t]) remain unparsable until around 20 months of age. This line of investigation has been reopened and expanded by Gagliardi et al. (2014) whose results suggest that the experimental methodology employed by Seidl et al. (2003) was flawed in that it presumed infants have ideal performance mechanisms. By providing more trials of each condition and controlling for the pragmatic felicity of test statements, Gagliardi et al. (2014) provide evidence that 15-month old infants can process wh-extractions from both subject and object positions. Object extractions are more difficult to comprehend than subject extractions, however, perhaps due to additional processing load in object extractions (Gibson, 1998; Phillips, 2010). Similarly, Gagliardi and Lidz (2010) show that relativized extractions with a wh-relativizer (e.g. find [the boy], who t ate the apple) are easier to comprehend than relativized extractions with that as the relativizer (e.g. find [the boy], that t ate the apple).

Yuan et al. (2012) demonstrate that 19-month olds use their knowledge of nouns to learn both verbs and their associated argument structure. In their study, infants were shown video of a person talking on a phone using a nonce verb with either one or two nouns (e.g. Mary kranded Susan). Under the assumption that infants look longer at things that correspond to their understanding of a prompt, the infants were then shown two images that potentially depicted the described action – one picture where two actors acted independently (reflecting an intransitive proposition) and one picture where one actor acted on the other (reflecting a transitive proposition). Even though the infants had no extralinguistic knowledge about the verb, they consistently treated the verb as transitive if two nouns were present and intransitive if only one noun was present.

Similarly, Gertner and Fisher (2012) show that intransitive phrases with conjoined subjects (e.g. John and Mary gorped) are given a transitive interpretation (i.e. John gorped Mary) at 21 months (henceforth termed ‘1-1 role bias’), though this effect is no longer present at 25 months (Naigles, 1990). This finding suggests both that learners will ignore canonical structure in favor of using all possible arguments and that children have a bias to assign a unique semantic role to each argument. It is important to note, however, that cross-linguistically children do not seem to generalize beyond two arguments until after at least 31 months of age (Goldberg et al., 2004; Bello, 2012), so a predicate occurring with three nouns would still likely be interpreted as merely transitive rather than ditransitive.

Computational modeling provides a way to test the computational level of processing (Marr, 1982). That is, given the input (child-directed speech, adult-directed speech, and environmental experiences), it is possible to probe the computational processes that result in the observed output. However, previous computational models of grammar induction (Klein and Manning, 2004), including infant grammar induction (Kwiatkowski et al., 2012), have not addressed filler-gap comprehension. The closest work to that presented here is the work on BabySRL (Connor et al., 2008; Connor et al., 2009; Connor et al., 2010). BabySRL is a computational model of semantic role acquisition using a similar set of assumptions to the current work. BabySRL learns weights over ordering constraints (e.g. preverbal, second noun, etc.) to acquire semantic role labelling while still exhibiting 1-1 role bias. However, no analysis has evaluated the abil-

\[2\text{Since the wh-phrase is in the same (or a very similar) position as the original subject when the wh-phrase takes subject position, it is not clear that these constructions are true extractions (Culicover, 2013), however, this paper will continue to refer to them as such for ease of exposition.}\]

\[3\text{There were two actors in each image to avoid biasing the infants to look at the image with more actors.}\]

\[4\text{As one reviewer notes, Joshi et al. (1990) and subsequent work show that filler-gap phenomena can be formally captured by mildly context-sensitive grammar formalisms; these have the virtue of scaling up to adult grammar, but due to their complexity, do not seem to have been described as models of early acquisition.}\]
Table 1: An example of a chunked sentence (Susan said John gave the girl a red book) with the sentence positions labelled. Nominal heads of noun chunks are in bold.

<table>
<thead>
<tr>
<th>Susan</th>
<th>said</th>
<th>John</th>
<th>gave</th>
<th>girl</th>
<th>book</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

3 Assumptions

The present work restricts itself to acquiring filler-gap comprehension in English. The model presented here learns a single, non-recursive ordering for the semantic roles in each sentence relative to the verb since several studies have suggested that early child grammars may consist of simple linear grammars that are dictated by semantic roles (Diesel and Tomasello, 2001; Jackendoff and Wittenberg, in press). This work assumes learners can already identify nouns and verbs, which is supported by Shi et al. (1999) who show that children at an extremely young age can distinguish between content and function words and by Waxman and Booth (2001) who show that children can distinguish between different types of content words. Further, since Waxman and Booth (2001) demonstrate that, by 14 months, children are able to distinguish nouns from modifiers, this work assumes learners can already chunk nouns and access the nominal head. To handle recursion, this work assumes that children treat the final verb in each sentence as the main verb (implicitly assuming sentence segmentation), which ideally assigns roles to each of the nouns in the sentence.

Due to the findings of Yuan et al. (2012), this work adopts a ‘syntactic bootstrapping’ theory of acquisition (Gleitman, 1990), where structural properties (e.g. number of nouns) inform the learner about semantic properties of a predicate (e.g. how many semantic roles it confers). Since infants infer the number of semantic roles, this work further assumes they already have expectations about where these roles tend to be realized in sentences, if they appear. These positions may correspond to different semantic roles for different predicates (e.g. the subject of run and of melt); however, the role for predicates with a single argument is usually assigned to the noun that precedes the verb while a second argument is usually assigned after the verb. The semantic properties of these roles may be learned lexically for each predicate, but that is beyond the scope of this work. Therefore, this work uses syntactic and semantic roles interchangeably (e.g. subject and agent).

<table>
<thead>
<tr>
<th></th>
<th>µ</th>
<th>σ</th>
<th>π</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_{SC}$</td>
<td>-1</td>
<td>0.5</td>
<td>.999</td>
</tr>
<tr>
<td>$G_{SN}$</td>
<td>-1</td>
<td>3</td>
<td>.001</td>
</tr>
<tr>
<td>$G_{OC}$</td>
<td>1</td>
<td>0.5</td>
<td>.999</td>
</tr>
<tr>
<td>$G_{ON}$</td>
<td>1</td>
<td>3</td>
<td>.001</td>
</tr>
<tr>
<td>$\Phi$</td>
<td></td>
<td></td>
<td>.00001</td>
</tr>
</tbody>
</table>

Table 2: Initial values for the mean (µ), standard deviation (σ), and prior (π) of each Gaussian as well as the skip penalty (Φ) used in this paper.

Finally, following the finding by Gertner and Fisher (2012) that children interpret intransitives with conjoined subjects as transitives, this work assumes that semantic roles have a one-to-one correspondence with nouns in a sentence (similarly used as a soft constraint in the semantic role labelling work of Titov and Klementiev, 2012).
Thus, the initial model conditions (see Figure 2) are most likely to realize an SVO ordering, although it is possible to obtain SOV (by sampling a negative number from the blue curve) or even OSV (by also sampling the red curve very close to 0). The model is most likely to hypothesize a preverbal object when it has already assigned the subject role to something and, in addition, there is no postverbal noun competing for the object label. In other words, the model infers that an object extraction may have occurred if there is a ‘missing’ postverbal argument.

Finally, the probability of a given sequence is the product of the label probabilities for the component argument positions (e.g. $G_{SC}$ generating an argument at position -2, etc). Since many sentences have more than two nouns, the model is allowed to skip nouns by multiplying a penalty term ($\Phi$) into the product for each skipped noun; the cost is set at 0.00001 for this study, though see Section 7 for a discussion of the constraints on this parameter. See Table 2 for initialization parameters and Figure 2 for a visual representation of the initial expectations of the model.

This work uses a model with 2-component mixtures for both subjects and objects (termed the symmetric model). This formulation achieves the best fit to the training data according to the Bayesian Information Criterion (BIC). However, follow-up experiments find that the non-canonical subject Gaussian only improves the likelihood of the data by erroneously modeling postverbal nouns in imperative statements. The lack of a canonical subject in English imperatives allows the model to improve the likelihood of the data by using the non-canonical subject Gaussian to capture fictitious postverbal arguments. When imperatives are filtered out of the training corpus, the symmetric model obtains a worse BIC fit than a model that lacks the non-canonical subject Gaussian. Therefore, if one makes the assumption that imperatives are prosodically-marked for learners (e.g. the learner is the implicit subject), the best model is one that lacks a non-canonical subject. The remainder of this paper assumes a symmetric model to demonstrate what happens if such an assumption is not made; for the evaluations described in this paper, the results are similar in either case.

This model differs from other non-recursive computational models of grammar induction (e.g. Goldwater and Griffiths, 2007) since it is not based on Hidden Markov Models. Instead, it determines the best ordering for the sentence as a whole. This approach bears some similarity to a Generalized Mallows model (Chen et al., 2009), but the current formulation was chosen due to being independently posited as cognitively plausible (Boersma, 1997).

Figure 2 (Right) shows the converged, final state of the model. The model expects the first argument (usually agent) to be assigned preverbally and expects the second (say, patient) to be assigned postverbally; however, there is now a larger chance that the second argument will appear preverbally.

5 Evaluation

The model in this work is trained using transcribed child-directed speech (CDS) from the BabySRL portions (Connor et al., 2008) of CHILDES (MacWhinney, 2000). Chunking is performed us-
Table 3: Overall accuracy on the Eve and Adam sections of the BabySRL corpus. Bottom rows reflect accuracy when non-agent roles are collapsed into a single role. Note that improvements are numerically slight since filler-gap is relatively rare (Schuler, 2011). *p << .01

<table>
<thead>
<tr>
<th></th>
<th>Eve (n = 4820)</th>
<th>Adam (n = 4461)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Initial</td>
<td>.54</td>
<td>.64</td>
</tr>
<tr>
<td>Trained</td>
<td>.52</td>
<td>.69</td>
</tr>
<tr>
<td>Initialc</td>
<td>.56</td>
<td>.66</td>
</tr>
<tr>
<td>Trainedc</td>
<td>.54</td>
<td>.71</td>
</tr>
</tbody>
</table>

Table 4: (Above) Filters to extract filler-gap constructions: A) the subject and verb are not adjacent, B) the object precedes the verb. (Below) Filler-gap accuracy on the Eve and Adam sections of the BabySRL corpus when non-agent roles are collapsed into a single role. *p < .01

<table>
<thead>
<tr>
<th></th>
<th>Eve (n = 1345)</th>
<th>Adam (n = 1287)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Subject Extraction filter: S x V ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object Extraction filter: O ... V ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>.53</td>
<td>.57</td>
</tr>
<tr>
<td>Trained</td>
<td>.55</td>
<td>.67</td>
</tr>
</tbody>
</table>

This is included for transparency, though the initial parameters have very little bearing on the final results as stated in Section 7, so the danger of overfitting to development data is very slight.
Table 5: (Left) Subject-extraction accuracy and object-extraction accuracy and (Right) Wh-relative accuracy and that-relative accuracy; calculated over the Eve and Adam sections of the BabySRL corpus with non-agent roles collapsed into a single role. \( p = .02 \) *\( p << .01 \)

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Eve Subj (n = 691)</td>
<td>Obj (n = 654)</td>
<td>Initial _</td>
<td>.66</td>
<td>.83</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>Trained _</td>
<td>.64</td>
<td>.84</td>
<td>.72*</td>
<td>.45</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>Adam Subj (n = 886)</td>
<td>Obj (n = 1050)</td>
<td>Initial _</td>
<td>.69</td>
<td>.81</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>Trained _</td>
<td>.66</td>
<td>.81</td>
<td>.73</td>
<td>.44</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>Eve Wh- (n = 689)</td>
<td>That (n = 125)</td>
<td>Initial _</td>
<td>.63</td>
<td>.45</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>Trained _</td>
<td>.73</td>
<td>.75</td>
<td>.74*</td>
<td>.44</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td>Adam Wh- (n = 748)</td>
<td>That (n = 189)</td>
<td>Initial _</td>
<td>.50</td>
<td>.37</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>Trained _</td>
<td>.61</td>
<td>.65</td>
<td>.63*</td>
<td>.47</td>
<td>.56</td>
</tr>
</tbody>
</table>

6 Comparison to BabySRL

The acquisition of semantic role labelling (SRL) by the BabySRL model (Connor et al., 2008; Connor et al., 2009; Connor et al., 2010) bears many similarities to the current work and is, to our knowledge, the only comparable line of inquiry to the current one. The primary function of BabySRL is to model the acquisition of semantic role labelling while making an idiosyncratic error which infants also make (Gertner and Fisher, 2012), the 1-1 role bias error (John and Mary gorped Mary) interpreted as John gorped Mary. Similar to the model presented in this paper, BabySRL is based on simple ordering features such as argument position relative to the verb and argument position relative to the other arguments.

This section will demonstrate that the model in this paper initially reflects 1-1 role bias comparably to BabySRL, though it progresses beyond this bias after training. \(^{10}\) Further, the model in this paper is able to reflect the concurrent acquisition of filler-gap whereas BabySRL does not seem well-suited to such a task. Finally, BabySRL performs undesirably in intransitive settings whereas the model in this paper does not.

Connor et al. (2008) demonstrate that a supervised perceptron classifier, based on positional features and trained on the silver role label annotations of the BabySRL corpus, manifests 1-1 role bias errors. Follow-up studies show that supervision may be lessened (Connor et al., 2009) or removed (Connor et al., 2010) and BabySRL will still reflect a substantial 1-1 role bias.

Connor et al. (2008) and Connor et al. (2009) run direct analyses of how frequently their models make 1-1 role bias errors. A comparable evaluation may be run on the current model by generating 1000 sentences with a structure of NNV and reporting how many times the model chooses a subject-first labelling (see Table 6). \(^{11}\)

\(^{9}\) Though performance is slightly worse when arguments are not collapsed, all the same patterns emerge.

\(^{10}\) All evaluations in this section are preceded by training on the chunked Eve corpus.

\(^{11}\) While Table 6 analyses erroneous labellings of NNV structure, the ‘Obj’ column of Table 5 (Left)
Error rate

<table>
<thead>
<tr>
<th>Initial</th>
<th>Trained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial (given 2 args)</td>
<td>Trained (given 2 args)</td>
</tr>
<tr>
<td>2008 arg-arg position</td>
<td>2008 arg-verb position</td>
</tr>
<tr>
<td>2009 arg-arg position</td>
<td>2009 arg-verb position</td>
</tr>
<tr>
<td>Initial (given 2 args)</td>
<td>Trained (given 2 args)</td>
</tr>
</tbody>
</table>

Table 6: 1-1 role bias error in this model compared to the models of Connor et al. (2008) and Connor et al. (2009). That is, how frequently each model labelled an NNV sentence SOV. Since the Connor et al. models are perceptron-based, they require both arguments to be labelled. The model presented in this paper does not share this restriction, so the raw error rate for this model is presented in the first two lines; the error rate once this additional restriction is imposed is given in the second two lines.

The results of Connor et al. (2008) and Connor et al. (2009) depend on whether BabySRL uses argument-argument relative position as a feature or argument-verb relative position as a feature (there is no combined model). Further, the model presented here from Connor et al. (2009) has a unique argument constraint, similar to the model in this paper, in order to make comparison as direct as possible.

The 1-1 role bias error rate (before training) of the model presented in this paper is comparable to that of Connor et al. (2008) and Connor et al. (2009), which shows that the current model provides comparable developmental modeling benefits to the BabySRL models. Further, similar to real children (see Figure 1) the model presented in this paper develops beyond this error by the end of its training, whereas the BabySRL models still make this error after training.

Connor et al. (2010) look at how frequently their model correctly labels the agent in transitive and intransitive sentences with unknown verbs (to demonstrate that it exhibits an agent-first bias). This evaluation can be replicated for the current study by generating 1,000 sentences with the transitive setting and 1,000 sentences with the intransitive setting. The difference in transitive settings stems from increased lexicalization, as is apparent from their results alone; the model presented here initially performs close to their weakly lexicalized model, though training impedes agent-predicate relative position as well as argument-predicate relative position and so is closest to the model presented in this paper.

For the intransitive case, however, whereas the model presented in this paper is generally able to successfully label the lone noun as the subject, the model of Connor et al. (2010) chooses to label lone nouns as objects about 40% of the time. This likely stems from their model’s reliance on argument-argument relative position as a feature; when there is no additional argument to use for reference, the model’s accuracy decreases. This is borne out by their model (not shown in Table 7) that omits the argument-argument relative position feature and solely relies on verb-argument position, which achieves up to 70% accuracy in intransitive settings. Even in that case, however, BabySRL still chooses to label lone nouns as objects 30% of the time. The fact that intransitive sentences are more common than transitive sentences in both the Eve and Adam sections of the BabySRL corpus suggests that learners should be more likely to assign...
correct roles in an intransitive setting, which is not reflected in the BabySRL results.

The overall reason for the different results between the current work and BabySRL is that BabySRL relies on positional features that measure the relative position of two individual elements (e.g. where a given noun is relative to the verb). Since the model in this paper operates over global orderings, it implicitly takes into account the positions of other nouns as it models argument position relative to the verb; object and subject are in competition as labels for preverbal nouns, so a preverbal object is usually only assigned once a subject has already been detected.

Further, while BabySRL consistently reflects 1-1 role bias (corresponding to a pre 25-month old learner), it also learns to productively label five roles, which developmental studies have shown does not take place until at least 31 months (Goldberg et al., 2004; Bello, 2012). Finally, it does not seem likely that BabySRL could be easily extended to capture filler-gap acquisition. The argument-verb position features impede acquisition of filler-gap by classifying preverbal arguments as agents, and the argument-argument position features inhibit accurate labelling in intransitive settings and result in an agent-first bias which would tend to label extracted objects as agents. In fact, these observations suggest that any linear classifier which relies on positioning features will have difficulties modeling filler-gap acquisition.

In sum, the unlexicalized model presented in this paper is able to achieve greater labelling accuracy than the lexicalized BabySRL models in intransitive settings, though this model does perform slightly worse in the less common transitive setting. Further, the unsupervised model in this paper initially reflects developmental 1-1 role bias as well as the supervised BabySRL models, and it is able to progress beyond this bias. Finally, unlike BabySRL, the model presented here provides a cognitive model of the acquisition of filler-gap comprehension, which BabySRL does not seem well-suited to model.

7 Discussion

This paper has presented a simple cognitive model of filler-gap acquisition, which is able to capture several findings from developmental psychology. Training significantly improves role labelling in the case of object-extractions, which improves the overall accuracy of the model. This boost is accompanied by a slight decrease in labelling accuracy in subject-extraction settings. The asymmetric ease of subject versus object comprehension is well-documented in both children and adults (Gibson, 1998), and while training improves the model’s ability to process object-extractions, there is still a gap between object-extraction and subject-extraction comprehension even after training.

Further, the model exhibits better comprehension of wh-relatives than that-relatives similar to children (Gagliardi and Lidz, 2010). This could also be an area where a lexicalized model could do better. As Gagliardi and Lidz (2010) point out, whereas wh-relatives such as who or which always signify a filler-gap construction, that can occur for many different reasons (demonstrative, determiner, complementizer, etc) and so is a much weaker filler-gap cue. A lexical model could potentially pick up on clues which could indicate when that is a relativizer or simply improve on its comprehension of wh-relatives even more.

It is interesting to note that the current model does not make use of that as a cue at all and yet is still slower at acquiring that-relatives than wh-relatives. This fact suggests that the findings of Gagliardi and Lidz (2010) may be partially explained by a frequency effect: perhaps the input to children is simply biased such that wh-relatives are much more common than that-relatives (as shown in Table 5).

This model also initially reflects the 1-1 role bias observed in children (Gertner and Fisher, 2012) as well as previous models (Connor et al., 2008; Connor et al., 2009; Connor et al., 2010) without sacrificing accuracy in canonical intransitive settings.

Finally, this model is extremely robust to different initializations. The canonical Gaussian expectations can begin far from the verb (±3) or close to the verb (±0.1), and the standard deviations of the distributions and the skip-penalty can vary widely; the model always converges to give comparable results to those presented here. The only constraint on the initial parameters is that the probability of the extracted object occurring preverbally must exceed the skip-penalty (i.e. extraction must be possible). In short, this paper describes a simple, robust cognitive model of the development of a learner between 15 months until somewhere between 25- and 30-months old (since 1-1 role bias is no longer present but no more than two arguments are being generalized).

In future, it would be interesting to incorporate lexicalization into the model presented in this paper, as this feature seems likely to bridge the gap between this model and BabySRL in transitive settings. Lexicalization should also help further distinguish modifiers from arguments and improve the overall accuracy of the model.

It would also be interesting to investigate how well this model generalizes to languages besides English. Since the model is able to use the verb position as a semi-permeable boundary between canonical subjects and objects, it may not work as
well in verb-final languages, and thus makes the prediction that filler-gap comprehension may be acquired later in development in such languages due to a greater reliance on hierarchical syntax.

Ordering is one of the defining characteristics of a language that must be acquired by learners (e.g. SVO vs SOV), and this work shows that filler-gap comprehension can be acquired as a by-product of learning orderings rather than having to resort to higher-order syntax. Note that this model cannot capture the constraints on filler-gap usage which require a hierarchical grammar (e.g. subjacency), but such knowledge is really only needed for successful production of filler-gap constructions, which occurs much later (around 5 years; de Villiers and Roeper, 1995). Further, the kind of ordering system proposed in this paper may form an initial basis for learning such grammars (Jackendoff and Wittenberg, in press).

8 Acknowledgements

Thanks to Peter Culicover, William Schuler, Laura Wagner, and the attendees of the OSU 2013 Fall Linguistics Colloquium Fest for feedback on this work. This work was partially funded by an OSU Dept. of Linguistics Targeted Investment for Excellence (TIE) grant for collaborative interdisciplinary projects conducted during the academic year 2012-13.

References


