Simulating children's verb inflection errors in English using an LSTM language model

Simulating children's verb inflection errors in English using an LSTM language model

Daniel Freudenthal (D.Freudenthal@liverpool.ac.uk),
Julian M. Pine (JPine@liverpool.ac.uk)
School of Psychology, University of Liverpool

Colin Bannard (Colin.Bannard@manchester.ac.uk)
Department of Linguistics and English Language, University of Manchester

Abstract

We present a computational (LSTM) model that learns to produce English (3sg and -bare) verb inflection when trained on English child-directed speech (CDS). The model is trained on input containing morphemized verbs and learns to predict the next token (word/morpheme) given a preceding sequence of tokens. The model produces the type of error (-bare for -3s) made by English-learning children while avoiding errors that children do not often make (-3s for -bare). The model also shows the same type of sensitivity to input statistics that has been reported in English-learning children. Finally, we manipulated the length of the sequences the model is trained on and show that this results in the delayed acquisition of -3sg forms that is characteristic of English-learning children with Developmental Language Disorder (DLD). Taken together these results suggest that input-driven learning is a major determinant of the patterns observed in both typical and impaired acquisition of English verb inflection.

Keywords: English Verb Inflection, Developmental Language Disorder. LSTM Language Model, Input-Driven learning.

Introduction

Children in many languages go through a stage in which they produce uninflected or nonfinite forms in contexts where an inflected form is required in the adult language. Thus, children learning English may produce forms with missing 3rd person singular (3sg) -s, like *he go there* instead of the correct *he goes there*. These types of errors have become known as Optional Infinitive (OI) errors (Wexler, 1994; 1998). OI errors are very common in English-speaking children: typically developing children only reach 90% correct provision on 3sg -s when they are around 4.5 years old (Rice et al., 1998). This is in contrast to children learning (for example) Spanish, where error rates lower than 5% have been reported in children as young as 2.5 years old (Hoekstra & Hyams, 1998). The problems that English-speaking children have in acquiring (3sg) inflection extend to clinical populations. English-speaking children with Developmental Language Disorder (DLD) continue to produce OI errors for far longer than typically developing children, even when controlling for factors such as children's Mean Length of Utterance (MLU) or vocabulary size (Leonard et al., 1999).

Several explanations have been proposed for English-speaking children's difficulties with verbal morphology, many of which suggest that the nature of the input makes the English 3sg -s hard to acquire. Freudenthal et al. (2015; 2023) argue that the prevalence of the bare stem in English (which is used as the infinitive as well as all present tense forms apart from the 3sg) means that children may produce it as a default form. Experimental evidence for this suggestion has been provided by Räsänen et al. (2014), who showed that children tend to produce bare forms in 3sg contexts of those verbs that tend to occur as bare forms in the input.

A second characteristic of English input that may make 3sg inflection hard to acquire is the fact that the English process of question formation means that bare forms frequently occur after 3sg subjects. That is, phrases like *Can he go*, or *Does he go*, actually provide conflicting evidence suggesting that phrases like *he go* are acceptable English word sequences in isolation. This feature of English has also been implicated in explaining the difficulty that children with DLD have with English verb inflection. According to the *Competing Sources of Input* hypothesis (Leonard et al., 2015), children may fail to grasp that the non-finite verb *go* in these sequences is licensed because of the preceding (separated) modal or auxiliary verb *can* or *does*, and this problem may be exacerbated in children with DLD, who have been shown to have difficulties in processing long-distance dependencies (Purdy et al., 2014), and in the learning of sequential information (verbal statistical learning: Lammertink et al., 2017; serial reaction time tasks: Lum et al., 2014). Consistent with this account, Sawyer et al. (2022) have found that a useful predictor of whether a child will make a bare-form error on a given production is the relative frequency with which the particular subject-verb combination they are

producing occurs in (contextually appropriate) bare as opposed to inflected form in the input.

Such input-driven accounts were recently tested by Freudenthal et al. (2021), who used a sequential version of the Rescorla-Wagner model, and showed that models trained on (idealized) English input were slow to acquire verb inflection compared to models trained on Spanish input. Moreover, unlike the Spanish model, the English model's performance was much impaired when the model's ability to integrate information over longer phrases was degraded.

Freudenthal et al. thus provide support for the claim that questions, in particular, and the statistics of the wider input, in general, might explain the difficulty that both typically developing children and children with DLD have with English verb inflection. However, Freudenthal et al. used an idealized version of the input, which was limited to the 30 most frequent verbs, and conflated all (pronominal and lexical) 3sg subjects into a single -3S marker. Moreover, the model's input was limited to subjects and verbs (excluding all other words) and the model's task was to predict one of 3 possible inflections (-3S, bare form or progressive).

In this paper, we investigate whether a model faced with more realistic input and a more challenging task can learn English verb inflection, and whether it will show the same pattern of error as children learning English (whilst avoiding errors not produced by children). We do this by training a Language Model (LM) equipped with an LSTM cell and embedding layers to predict the next word given an input sequence of a given length. Recent work has shown that models like this are capable of learning long-distance number agreement, whether run as classifier models or performing next word prediction (Linzen et al., 2016; Gulordava et al., 2018)

The input to our model consists of sequences of child-directed speech. Importantly, the model input retains the difference between different subjects, is not limited to a specific set of verbs (outside general vocabulary limitations), and the model was trained to predict the next word in the input (with verb inflections considered as separate words). We were interested in whether the model produces the errors shown by children (-bare for -3s errors), whilst avoiding those that children do not (or rarely) produce (-3S for -bare errors). Additionally, we were interested in whether reducing the length of the sequences the model is exposed to results in the increased error rates found in children with DLD who have been argued to show decreased sensitivity to long-distance dependencies.

Methods

The model:

The model we employed was a Language Model (LM) that was trained to predict the next word in the input based on preceding input sequences of a given length. The model uses an input layer with one node for every word in the input vocabulary. The input layer connects to an embedding layer which learns distributed (semantic) representations for the words in the input during training. The embedding layer connects to an LSTM cell, which in turn connects to an output layer containing the same number of nodes as the input layer. After a given training or test trial, the output layer represents an estimated probability distribution over the model vocabulary given the input sequence. In what follows, we trained high- and low-capacity models aimed at investigating the model's ability to simulate late stages and early stages of development respectively. High-capacity models had an embedding layer of 100 nodes and 500 hidden nodes. They were trained for 100,000 batches of 32 training items with a fixed learning rate of .001 (the default for the Adam optimizer). Low-capacity models had an embedding length of 50, and a hidden (LSTM) capacity of 100 nodes. These models were trained for 250,000 batches with an initial learning rate of .0001 which was increased by .0001 every 5,000 batches to a maximum of .001.

Model Input:

We trained the model using child-directed speech (CDS) obtained from CHILDES (see project repository at https://github.com/cbannard/inflection_errors_cogsci23 for details). The input consisted of a mix of UK and US English corpora. We performed minor filtering on the CHILDES files (we removed punctuation and mark-up). For information on the corpora used see the github repository. The total amount of input was approximately 2 million utterances, restricted to a maximum length of 15 words. We reduced the model vocabulary to words that occurred in the input a minimum of 10 times – words in the input that had a frequency of less than 10 were replaced with an <unk> marker. We used the FLAIR (Akbik et al., 2019) tagger to assign Part of Speech (POS) categories to words in the input. We morphemized items assigned a 3sg present tense (FLAIR VBZ), non-3sg present tense (VBP) or verb base form (VB) tag, by changing them into their base form and adding a -3S or -BARE morpheme (e.g., goes -> go -3S, go -> go- BARE) as the next item in the sequence. The inflection was thus encoded as a separate word, and hence a potential target in the next word prediction task.

Model training:

Words in the input that had a frequency of less than 10 were replaced with an <unk> marker. This reduced the model vocab from ~50,000 to ~11,000 words. The input was split into a training and test set. The training set contained all items that did not end in -3S or -BARE, plus 80% of items that *did end* in -3S or -BARE. The test set contained the remaining 20% of items ending in -3S or -BARE. The test set was further reduced to contain only items where the POS-tag of the pre-verbal element was a (singular or plural) (pro)noun. We also excluded items containing an <unk> or 'XXX' (unintelligible) marker from testing. The test set thus contained items ending in Subject-Verb-Inflection sequences like *I go -BARE, Anne want -3S, Does he run -BARE, She eat*

-3S. We only tested the model on sequences that ended in a subject-verb sequence that occurred in the training set. Individual models were run and averaged over 5 different instantiations of training and test-set (5-fold validation).
Models were trained on sequences of tokens (morphemized input words). Sequences did not cross utterance boundaries (i.e., were restricted to single utterances). For a given sequence length (n), the sequence ending in the *ith* word in an utterance contained the *n-1* words preceding it, with padding markers added as necessary (due to utterance boundaries) up to the sequence length. Thus, for sequence length 5, the model predicted the 5th word on the basis of words 1-4. The second word in the utterance was predicted by 3 padding markers followed by the first word in the utterance.

The training set consisted of ~9.5 million sequences, most of which ended in an item other than -3S or -BARE. The test set consisted of ~70,000 items of which ~7,500 ended in -3S and the rest in -BARE. Testing took place every 5,000 batches. Models were tested on all -3S targets and a random sample of 10,000 -BARE targets, and accuracy (proportion correct) was recorded for -3S and -BARE targets separately. In testing we only considered the relative production probability of the -3S and -BARE form, and assigned to the model's output the one that was most likely given the preceding context (that is, we ignored other, potentially more probable words). In practice, however, the model considered few other forms with high likelihood beyond the very early stages of training.

Results

Performance of high-capacity models

We start by examining the extent to which the model is capable of learning English verb inflection when given substantial resources. Fig.1 shows the performance of high-capacity models (embedding of length 100, 500 hidden (LSTM) nodes, and learning rate of .001) trained for 100,000 batches on sequences of length 5, 6, 8 and 10 (-3S targets only). As can be seen in Figure 1, the model is able to reach around 95% correct performance, with the models trained on the longest sequences performing best. Given sufficient capacity, the model is thus capable of learning the task, but it shows limited developmental effects. The model's ability to simulate developmental change (i.e., earlier developmental phases) was then examined by training models with lower capacity (embedding length 50, 100 hidden nodes and a learning rate that gradually increases from .0001 to .001).

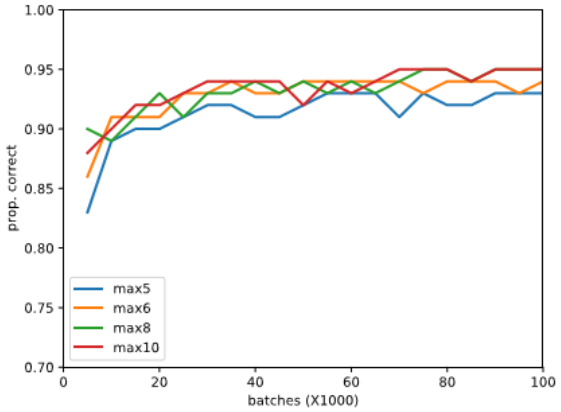

Fig. 1: -3S performance of high-capacity models trained on sequences of length 5,6,8,10.

Performance of low-capacity models

Figure 2 shows the results for low-capacity models trained on sequences of length 5 for a total of 250,000 batches. The figure shows accuracy levels for -3S and -BARE targets separately. The model achieves low initial accuracy scores for -3S targets which gradually rise to peak at around 90% correct. By contrast, accuracy for -BARE targets is near 100% throughout. Since the testing procedure considers the relative probability of -3S and -BARE forms only, all errors on -3S targets involve the production of a -BARE form (and vice-versa). The model thus starts out producing many BARE for -3S errors, and at the same time produces very few -3S for -BARE errors (as is the case in English-learning children). With increased exposure the model is increasingly capable of distinguishing the contexts that require a -3S and -BARE response, and accuracy on -3S targets increases as a result.

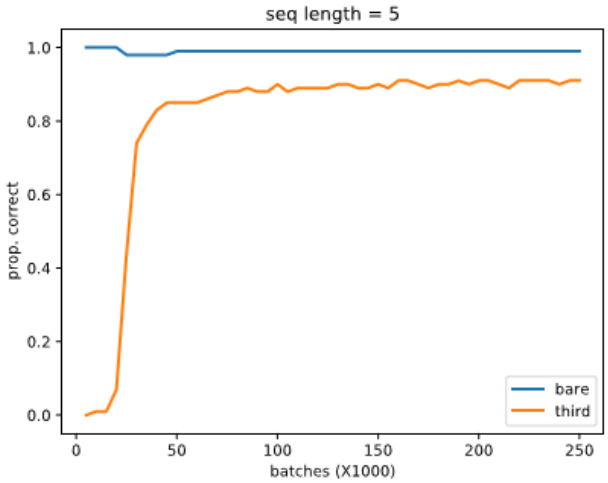

Fig. 2: -3S and -BARE Performance of low-capacity models trained on sequences of length 5.

Fig. 2 shows that the model can show high accuracy on -BARE targets combined with increasing accuracy on -3S targets, when collapsed over all test items.

Understanding errors

In order to understand what is driving the errors seen, we built a series of regression models examining whether the model's behaviour reflects the same factors as that of children. As noted in the introduction, one prominent claim is that children's errors reflect them defaulting to the most frequent form for the given verb being produced. In the spirit of this prediction, we tested whether the model's predictions of a non-finite form in contexts where a finite form is required reflect the relative frequency with which that verb is seen in bare form across all contexts in child-directed speech. We analysed one model at an intermediate stage (100,000 batches) of training and regressed the probability of seeing the bare form on the output layer on the frequency of the bare form, controlled for the total frequency of the verb. We found that the input frequency of the bare form explains variance above the total verb frequency ($F(1,7383) = 209.37$; $p < 2.2e\text{-}16$), with standardized coefficients indicating that the probability of the model predicting a bare form decreases with the total frequency of the verb (B=-.584), but increases with the frequency of the bare form of the verb across all contexts (B =.597).

We next examined the possibility that it is a high rate of exposure to instances of the particular subject-verb combination in the input that drives errors. We followed the same process as for the verb frequencies but using the total frequency of the subject verb sequence and its frequency in bare form. We found that the frequency of the bare form of the subject-verb combination, like that of the verb, explains variance above the total verb frequency ($F(1,7383) = 816.48$; $p < 2.2e\text{-}16$), with standardized coefficients indicating that the probability of the model predicting a bare form decreases with the total frequency of the verb (B=-0.143), but increases with the frequency of bare form of the verb across all contexts (B = 0.121). For completeness, we next repeated the process for the relative frequency of any bare form verb appearing with each subject, and found the same pattern. Finally, we built a model containing all 6 predictors for all three models, and found that all terms explained unique variance while all maintaining the same direction of effect. This suggests that our LSTM model is making independent use of all of these sources of information in making its predictions.

In summary, then, the model shows lower rates of (-Bare for -3S) error for phrases that are of higher frequency, and higher rates of error for sequences that are more likely to occur with bare forms. The frequency of the bare form explains the most unique variance (controlling for total frequency of the conditioning form) in the model regressed on the identity of the Subject (partial R^2 = .281), followed by the Subject-Verb sequences (partial R^2 = .10), and the identity of the Verb (partial R^2 = .027). The partial R^2 for all of the bare form frequencies combined was .24, The R^2 for the model with all six predictors was .32, with the three individual models explaining .29 (Subjects), .13 (Subject-Verb) and .03 (Verbs) % off the variance.

We also ran regressions at an earlier stage of training (35,000 batches). These models largely replicated the pattern above, with the full model explaining 32% of the variance. However, these models showed a far greater influence of the Verb statistics, with the three individual models explaining 18% (Verbs), 15% (Subjects) and 7% (Subject-Verb sequences) of the variance.

Performance on individual test items

Figure 3 looks at the model's performance on 4 individual items in more detail (averaged over 5 model runs). It plots model performance (across 250,000 batches) on a high frequency verb (*go*) with four individual subjects requiring a -3S response - two pronominal and two lexical. Plotted in Figure 3 are the absolute production probabilities for the -BARE and -3S form. Figure 3 shows different performance trajectories for the different test items. The phrases with (frequent) pronominal subjects (*he*, *that one*) are learned much faster than the phrases with (less frequent) lexical subjects (*mummy*, *the lion*). A similar pattern has been reported in English-learning children (Pine et al., 2008). However, there are also differences within the lexical and pronominal classes. The model has more difficulty learning the -3S suffix for sequences ending in *mummy go* compared to *the lion go*, possibly reflecting the increased use of the word *mummy* as a 3sg subject in questions.

It is worth noting that, after the early stages of training, the summed production probabilities of the -BARE and -3S form are close to one, reflecting the fact that (for these stimuli) the model does not assign a high probability to words other than -BARE or -3S. Finally, in the early stages, the model prefers the -BARE form over the -3S form for all four phrases, reflecting the fact that the -BARE form acts as a default form in the initial stages. The default -BARE response is replaced with a -3S response as evidence for it becomes available, but this happens at different speeds for different phrases.

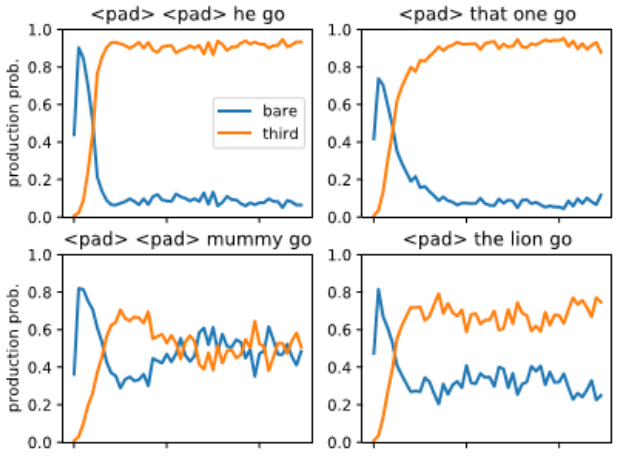

Fig. 3: Performance of low-capacity models trained on sequences of length 5 on individual test items.

Performance of models trained on short sequences:

Figures 2 and 3 showed performance of models trained on sequences of length 5. For the majority of training and test items, sequences of this length will contain all the evidence that is required for the model to determine the correct response. Short targets (e.g., *he go -3s*, or *mummy want -3S*) will be padded out to length 5, which will set them apart from modal or auxilary phrases like *can he go -BARE*, or *does mummy want -BARE)*. Moreover, since *can* and *does* are included in the longer phrases, the model can learn that the -BARE response in these contexts is associated with the modal *can* and the auxiliary *does* rather than the 3sg subject that follows them.

Figure 4 shows the model's performance when the sequences it is trained and tested on are reduced to length 4 and 3 (thus depriving it of information needed to distinguish -3S from -BARE contexts). This reduction in context length serves as a potential model for a hypothesized deficit in the processing of sequential stimuli in children with DLD. Figure 4 shows an increase in error rates on -3S targets relative to Figure 2. This increase is particularly noticeable for sequences of length 3 (*he go -3S/-BARE)*, but also apparent for sequences of length 4, where distinguishing information is removed from sequences with multi-word subjects (*the lion go -3S/BARE).*

Figure 4 also shows a small increase in errors on -BARE targets. This increase likely results from the model being tested on phrases like ~~*Does*~~ *he want -BARE*, where the absence of *does* licenses a -3S response. The combined pattern of increasing -BARE for -3S errors and low error on -BARE targets is consistent with that found in children with DLD, and thus provides support for the notion of a sequence learning deficit as a (partial) explanation of the verb inflection deficit in DLD.

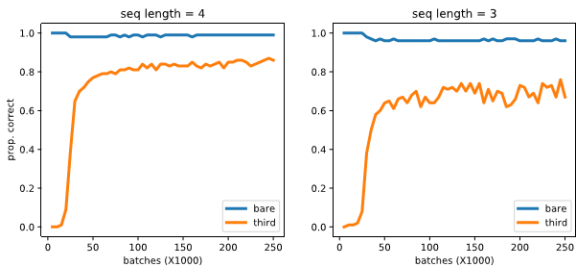

Fig. 4: Performance of low-capacity models trained on sequences of length 4 and 3.

Conclusions

The main conclusions to be drawn from the research described in this paper are a) that an LSTM Language Model that learns to predict the next word in the sequence is capable of learning English verb inflection when presented with English child-directed speech containing morphemized verb forms (when given sufficient capacity, the model is capable of reaching near 95% correct on the test set), and b) that when the model capacity and learning rate are restricted, the model starts to produce errors in a way that mirrors those made by young children - the errors that the model produces are largely restricted to -BARE for -3S errors, while the reverse (-3S for -BARE errors) are rare (as is the case in English-learning children).

These results are in line with those reported by Freudenthal et al. (2021), but the task faced by the current model is more challenging and more realistic than that faced by the Rescorla-Wagner model, which learned to predict one of three targets on the basis of idealized input (subject-verb sequences that conflated all 3sg subjects, and was restricted to 30 frequent verbs).

The current model also showed the same effects of releative input frequency as children. The regression analyses and plots for individual test items showed that the model produces fewer errors in 3sg contexts for items that are high frequency, and more errors for items that tend to occur with a bare form. Similar results have been reported in English-learning children. Sawyer et al. (2022) argue that children are more likely to produce -BARE for -3S errors on subject-verb sequences that occur with bare forms, while Pine et al. (2008) show that children are less likely to produce -BARE for -3S errors with (frequent) pronominal compared to lexical subjects.

Such input-driven effects on children's learning of verb inflection may not be surprising, but it is worth bearing in mind that -BARE for -3S errors (also referred to as OI errors) have often been explained in terms of the slow unfolding of innate linguistic principles (e.g., Wexler 1994, 1998). The kind of input-driven effects reported here are difficult to reconcile with this kind of explanation.

Our manipulation of sequence length suggests that the slow acquisition of verb inflection in English-learning children with DLD may reflect the same (but magnified) input-driven biases that operate in typically developing children. That is, children with DLD may have difficulty processing the fronted auxiliary or modal in questions, and hence may be slow, relative to TD children, to grasp the fact that phrases like 'he go' are not permissible in isolation. This is in contrast to some approaches that assume the deficit in DLD is related to procedural learning (e.g., Ullman & Pierpoint, 2005), but is in line with a recent proposal, calling for interventions for children with DLD that alter the distribution of the input to draw children's attention to the contexts in which bare forms are and are not permissible (Leonard et al., accepted). However, the actual mechanism for simulating DLD (reducing the length of sequences available to the model) is rather crude. We are currently in the process of investigating whether more sophisticated manipulations of the model's gating (and learning) mechanisms, can yield similar differences in performance in models that are exposed to the same lengths of training sequences.

Taken together, our simulations thus suggest that the slow acquisition of verb inflection in English children is largely driven by input factors (the high frequency of the English bare form, and the English process of question formation). However, they also suggest that the performance of TD and DLD children may be best viewed as lying on a continuum with the input effects apparent in TD children being amplified in children with DLD, as the result of a diminished ability to integrate information that unfolds over time.

Acknowledgements

Daniel Freudenthal, Julian Pine, and Colin Bannard are members of the International Centre for Language and Communicative Development (LuCiD) at the Universities of Liverpool and Manchester, for which the support of the Economic and Social Research Council [ES/L008955/1 and ES/S007113/1] is gratefully acknowledged.

References

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S. & Vollgraf, R. (2019). FLAIR: An easy-to-use framework for State-of-the-Art NLP. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*.

Freudenthal, D., Gobet, F. & Pine, J.M. (2023). MOSAIC+: A crosslinguistic model of verb-marking errors in typically developing children and children with Developmental Language Disorder. *Language Learning.* https://doi.org/10.1111/lang.12580

Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2015). Defaulting effects contribute to the simulation of cross-linguistic differences in Optional Infinitive errors. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Meeting of the Cognitive Science* Society (pp. 679– 684). Austin, TX: Cognitive Science Society.

Freudenthal, D., Ramscar, M., Leonard, L. B., & Pine, J.M. (2021). Simulating the Acquisition of Verb Inflection in Typically Developing Children and Children with Developmental Language Disorder in English and Spanish. *Cognitive Science*, *45,* e12945

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*

Hoekstra, T., & Hyams, N. (1998). Aspects of root infinitives. *Lingua*, *106*, 81–112.

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3473–3486. https://doi.org/10.1044/2017_JSLHR-L-16-0439

Leonard, L., Deevy, P., Bredin-Oja, S.L. & Schroeder. M.L. (accepted). Sources of Misinterpretation in the Input and their Implications for Language Intervention with English-speaking children. *American Journal of Speech-Language Pathology*.

Leonard, L. B., Fey, M. E., Deevy, P., & Bredin-Oja, S. L. (2015). Input sources of third person singular –s inconsistency in children with and without specific language impairment. *Journal of Child Language*, *42*(4), 786–820. https://doi.org/10.1017/S0305000914000397

Leonard, L., Miller, C., & Gerber, E. (1999). Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *42*, 678–689.

Linzen, T., Dupoux, E. & Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics 4, 521-535*

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex, 51*, 1–10.

Pine, J.M., Conti-Ramsden, G., Joseph, K.L., Lieven, E.V.M., & Serratrice, L. (2008). Tense over time: testing the Agreement/Tense Omission Model as an account of the pattern of tense-marking provision in early child English *Journal of Child Language*. *35*(1), *55-75*

Purdy, J. D., Leonard, L. B., Weber-Fox, C., & Kaganovich, N. (2014). Decreased sensitivity to long distance dependencies in children with a history of specific language impairment: Electro-physiological evidence. *Journal of Speech, Language, and Hearing Research*, *57*, 1040–1059. https://doi.org/10.1044/2014_JSLHR-L-13-0176

Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest-frequency form? *Journal of Child Language*, *41*(4), 756–779. https://doi.org/10.1017/S0305000913000159

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of Tense acquisition in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, *41*, 1412–1431.

Sawyer, H., Bannard, C. & Pine, J.M. (2022). Young Children's Verb-Marking Errors Are Predicted by the Relative Frequency of Contextually Appropriate Uninflected Sequences in Well-Formed Child-Directed Speech: A Pre-registered Corpus Analysis. *Psyarxiv*. https://psyarxiv.com/5k4eu/

Ullman, M., & Pierpoint, E. (2005). Specific language impairment is not specific to language:The Procedural Deficit Hypothesis. *Cortex*, *41*, 399-433.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivation in child grammar. In N.

Hornstein & D. Lightfoot (Eds.), *Verb Movement* (pp. 305-350). Cambridge: Cambridge University Press.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, *106*(1-4), 23-79.