EMPIRICAL STUDY

Toddlers' Verb-Marking Errors Are Predicted by the Relative Frequency of Uninflected Sequences in Well-Formed Child-Directed Speech: A Pre-registered Corpus Analysis

Hannah Sawyer[a], Colin Bannard[b], Julian Pine[a]

[a]University of Liverpool [b]University of Manchester

Author notes / acknowledgements

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

The data that support the findings of this study are openly available in CHILDES at http://doi.org/10.21415/T54G6D (Manchester corpus) and http://doi.org/10.21415/T5C31W (Conti-Ramsden 3 corpus). Analysis code is available on Open Science Framework at http://osf.io/ef8bm. The support of the Economic and Social Research Council (North West Social Science Doctoral Training Partnership (NWSSDTP) [ES/P000665/1] and International Centre for Language and Communication Development (LuCiD) [ES/L008955/1]) is gratefully acknowledged. We would also like to thank Daniel Freudenthal for his contribution to our data extraction code. Authors declare no conflicts of interest.

Correspondence concerning this article should be addressed to Hannah Sawyer, University of Liverpool: Department of Psychology, Eleanor Rathbone Building, Bedford Street South, Liverpool, L69 7ZA, hshsawy2@liverpool.ac.uk.

Abstract

Verb-marking errors such as ‘she play football’ and ‘Daddy singing’ are a hallmark feature of English-speaking children’s speech. We investigate the proposal that these errors are input-driven errors of commission, arising from the high relative frequency of subject+unmarked verb sequences in well-formed child-directed speech. We test this proposal via a pre-registered corpus analysis, and ask at what level the effects occur: is it the relative frequency of specific subject+unmarked verb sequences in the input that is important, or is it simply that verbs become entrenched, such that their frequency of appearance with any third-person singular subject accounts for errors? We find that the best predictor of children’s verb-marking errors is the relative frequency of unmarked forms of *specific* subject+verb sequences. Our results support the proposal that children's apparent omissions of certain grammatical morphemes are in fact input-driven errors of commission and provide insight into the mechanisms by which this occurs.

Key words: Language Development, Verb Marking Errors, Corpus Analysis, Child-Directed Speech.

It is well established that children are more fluent in producing words and phrases that occur frequently in caregiver speech than words and phrases that do not. There is also evidence that frequent forms have an advantage in grammatical development, with children being less likely to make errors of omission, such as omitting plural marking from nouns, for frequent words (Matthews & Theakston, 2006). But could the high relative frequency of particular words or phrases also cause children to make errors? In this paper, we investigate this issue with a focus on verb-marking errors such as ‘she play football’ and ‘Daddy singing’, which are a hallmark feature of young English-speaking children’s speech. We use a corpus analysis to test the idea that these errors reflect the extraction of sequences like ‘she play’ and ‘Daddy singing’ from longer sequences in which they are grammatical (e.g., ‘Does she play football?’ and ‘I can hear Daddy singing’ respectively). We also investigate the level at which any such sequence effects occur. For example, we ask whether they reflect the influence of specific subject+verb sequences (e.g., ‘she + play’) or of more abstract patterns (e.g., ‘third person singular (3sg) subject + play’). By establishing specific links between the subject+verb sequences in children’s errors and subject+verb sequences in the input, our results provide strong support for the view that these sequences are learned directly from the language that children hear.

Background

Verb-marking errors are a hallmark feature of children’s early speech. In English, two of the most common errors involve: 1) using the bare form of the verb when either the 3sg or past tense form is required (e.g., producing ‘she play football’ instead of ‘she plays football’ or ‘she played football’), and 2) using a bare progressive participle when an auxiliary and progressive combination is required (e.g., producing ‘Daddy singing’ instead of ‘Daddy is singing’; Brown, 1973). However, why children make such errors is still a matter of some debate.

Since verb-marking errors do not occur in the input, Nativist researchers have taken them as evidence of an underlying difference between the child and the adult grammar (e.g., Hoekstra & Hyams, 1998; Hyams, 1996; Legate & Yang, 2007; Rizzi, 1993; Wexler, 1994; 1998). These researchers argue that children have innate knowledge of inflection onto which they map the forms of the language they are learning, but that differences between the child and the adult grammar result in the incorrect use of certain verb forms in certain contexts. For example, according to Wexler’s (1998) Optional Infinitive account, verb-marking errors reflect a stage of development in which young children have set all the inflectional and phrase structure parameters of their language, but their grammar allows them the option of using untensed verbs in main clauses.

By contrast, Constructivist researchers have taken verb-marking errors as evidence of a lack of knowledge on the part of the child (e.g., Aguado-Orea & Pine, 2015; Bates & MacWhinney, 1989; Goldberg, 2019; Lieven, 2010; Tomasello, 2003). These researchers argue that children acquire knowledge of inflection gradually by analysing the input to which they are exposed, and that both their correct and their incorrect use of verb forms reflect the frequency statistics and semantic-distributional properties of the language they are learning. More specifically, they point out that sequences such as ‘she play’ and ‘Daddy singing’ do occur in the input within longer structures such as ‘Does she play football?’ and ‘I can hear Daddy singing’, and argue that verb-marking errors can be learned directly from the input as a result of the faulty processing of these longer structures. For example, in an elicitation study, Theakston et al. (2003) reported that children were more likely to produce bare stem errors with novel verbs (e.g., 'This one mib') when these verbs were introduced exclusively as bare forms in questions (e.g., 'Will it mib?'). Similar findings were also reported by Finneran and Leonard (2010) who repeated Theakston et al.'s (2003) study using an expanded list of sixteen novel verbs.

Verb-marking errors can also be observed in the speech of children with Developmental Language Disorder (DLD; previously referred to as Specific Language Impairment). DLD refers to a significant deficit in language ability that cannot be attributed to hearing loss, neurological damage, or gross developmental disorder (Leonard, 2014). While most typically developing (TD) children have mastered the use of tense and agreement morphemes by the time they reach five years of age, children with DLD produce verb-marking errors for longer than TD children and at higher rates than both age-matched and language-matched controls (Leonard et al., 1997; Rice et al., 1995; Rice et al., 1998).

Recently, Leonard and his colleagues have developed a unified input-driven account of the pattern of verb-marking errors in TD children and children with DLD: the Competing Sources of Input (CSI) account (Leonard & Deevy, 2011; Leonard et al., 2015). According to this account, children extract 3sg subject + unmarked form sequences from longer sentences and questions in the input because they are insensitive to material earlier in the sentence which makes these sequences grammatical in context. The period during which this faulty processing occurs is relatively short-lived for TD children but is significantly protracted in children with DLD (Leonard et al., 2015).

Leonard et al. (2015) argue that verb-marking errors are conditioned by the way subject + unmarked verb sequences are used in the input. However, the level at which these sequences influence verb-marking errors is not fully specified in the account. One of our goals in this work is to address this gap empirically. One possibility is that children extract specific subject +

unmarked verb sequences from the input and use them in contexts in which the equivalent subject + marked verb sequence is required. For example, the child may extract 'she play' from 'Does she play football?' and then produce verb-marking errors such as 'she play tennis'. Alternatively, it may be that the child extracts a more abstract pattern from such utterances, which is then used to produce verb-marking errors with a range of different subjects (e.g., 'Mummy play tennis', 'Daddy play basketball').

The CSI account predicts effects at the level of the subject+verb sequence. However, a third possibility is that verb-marking errors reflect the relative frequency with which verbs occur in bare versus inflected form regardless of context. In fact, there is already some evidence for this possibility in the literature. For example, in an elicited production study, Räsänen et al. (2014) found that the tendency to make bare stem errors in 3sg contexts with particular verbs was predicted by the relative frequency with which those verbs occurred in bare versus 3sg form in English child-directed speech (CDS). This finding has since been replicated by Kueser et al. (2018) in a group of children with DLD and a group of language-matched controls.

One reason we might expect sequence effects is that 3sg subject + unmarked verb sequences (e.g., 'He go') and 3sg subject + marked verb sequences (e.g., 'He goes') have the same, or very similar, semantics, and are therefore likely to compete for selection in the child's system. This is, of course, not the case for all bare and 3sg forms of the verb, since many bare forms (e.g., the 'go' in 'I go') will not have 3sg semantics. However, since input effects could occur at the level of the verb or the sequence, all three of the possible effects discussed above will be considered in the present study.

The Current Study

In the current study, we used a corpus analysis to test the predictions of the CSI account (Leonard et al., 2015). To do this, we looked for input effects on the tendency to produce unmarked forms (both bare stems and bare progressive participles) in 3sg contexts, in a group of TD children and a group of children with DLD. Both populations were included to investigate whether input effects varied between them.

Since it is unclear whether specific 3sg subject + unmarked verb sequences or unmarked verbs with any 3sg subject in the input contribute to verb-marking errors, we included the following input measures in our analyses: subject+verb sequence bias (i.e., the rate at which *specific* subject+verb sequences occur in bare versus inflected form, e.g., 'she play' vs. 'she plays') and verb bias in *3sg contexts* (i.e., the rate at which verbs occur in the input in 3sg contexts in bare versus inflected form, e.g., 'any 3sg subject + play' vs. 'any 3sg subject + plays'). Moreover, since a relation has already been established between the tendency to make errors on particular verbs and the relative frequency of bare versus inflected forms of the verb in the input, we also included a measure of verb bias in *any context* (i.e., the rate at which verbs occur in the input in the bare versus 3sg form, e.g., 'play' vs. 'plays'). This enabled us to investigate whether there were effects of sequence bias over and above any more general effect of verb bias. This variable is not included in the main analysis of progressive errors reported below as it was not pre-registered (due to there being no precedent for it in the previous literature). However, we later realised that it would be a useful variable to look at and therefore included it in an exploratory analysis. Overall, we address the following research questions:

1. Is the relative frequency of contextually appropriate unmarked forms in the input related to children's tendency to make verb-marking errors in their speech?

2. Which of the three measures of relative frequency of unmarked forms (subject+verb sequence bias, verb bias in *3sg contexts* and verb bias in *any context*) is the most valuable predictor of verb-marking errors?

3. Do the three measures of relative frequency of unmarked forms explain unique variance in children's verb-marking errors?

4. Does the effect of the best measure of relative frequency of unmarked forms vary dependent on whether the child is TD or diagnosed with DLD?

Methods

The corpus analysis consisted of four general phases: 1) Extraction of child-produced 3sg and progressive subject+verb sequences from a set of target corpora; 2) Identification of verb-marking errors; 3) Collection of bias statistics from CDS in English (UK and USA); and 4) Mixed-effects logistic regression modelling to determine whether our bias statistics are predictive of verb-marking errors and, if so, which statistic has the greatest predictive value. All stages of the corpus analysis were pre-registered on the Open Science Framework (OSF) [available at: osf.io/ef8bm].

Data

Twelve TD children (females *N*=6, males *N*=6) from the Manchester corpus (Theakston et al., 2001) with an average age of 1;11 (years;months) (range=1;8–2;0) at the beginning of data collection and four children with DLD (females *N*=1, males *N*=3) from the Conti-Ramsden 3 corpus (Joseph et al., 2002) with an average age of 3;3 (range=2;6–4;0) at the beginning of data collection were included in this study. The two groups (TD and DLD) were similar in Mean Length of Utterance (in words; MLUw) at the beginning of data collection (Manchester MLUw *Mean*=1.47 *SD*=0.32, Conti-Ramsden 3 MLUw *Mean*=1.67 *SD*=0.78) and there was no significant difference in the mean MLUw between the two groups ($t(14)=.75$, $p=.467$).

Both corpora contain transcripts of naturalistic interactions between the target children and their caregivers (predominately their mothers) in their normal home environment, spanning at least a one-year period. In total, the Manchester corpus contained 224,829 child utterances (*Mean utterances per child*=18,735.75, *SD*=3259.25, range=13,321–24,914), with a total of 512,555 words produced (*Mean words per* child=42,712.92, *SD*=9096.01, range=28,246–62,805). The Conti-Ramsden 3 corpus contained a total of 43,666 child utterances (*Mean utterances per child*=10,916.50, *SD*=3302.48, range=8598–15,787), with a total of 111,892 words produced (*Mean words per child*=27,973, *SD*=11,640.70, range=16,330–43,510). For further demographic information, see Supplementary Table 1 in Appendix S1.

Child Corpora

We prepared the Manchester and Conti-Ramsden 3 corpora for analyses using a script written in

Python 3 [available at: osf.io/ef8bm]. The following adjustments were made to all transcripts to ensure they were consistent throughout: punctuation (commas), embedded returns (where utterances spanned multiple lines) and parentheses when surrounding an alphanumeric character were removed; variant forms on the main tier were normalised by replacing approximate child productions with the transcriber-inferred intended form (e.g., 'toming [: coming]' was replaced with 'coming'); and other variant forms without transcriber-inferred interpretations were replaced (e.g., 'hasta' was replaced with 'has to'). We did not replace the contraction 'gonna' with 'going to' because young children may not be aware that the correct target is a progressive form. In addition to the corpora preparation details outlined in the OSF pre-registration document [available at: osf.io/ef8bm], error (e.g., [*]) and omission (e.g., 0is) markings were also removed and spaces were added before and after angle brackets (which were used to indicate any retracings or overlaps in the utterances).

Next, we automatically extracted child-produced 3sg and progressive subject+verb sequences from the target corpora. The Python script analysed the text lines for the child tier only and used the %mor tier to identify possible (pro)noun and verb combinations. It then searched for items marked as (pro)nouns (or proper nouns) and items marked as (main) verb or progressive (prog and presp). Finally, the %mor codes were stripped to arrive at the text representation of the lemma. This resulted in a list of base subjects and verbs. 'y' and 'ie' were added to items marked as diminutives and then lists of progressives and 3sg forms for each verb were generated. We then searched the text lines for items in the subject list that were followed by items in the verb list. For present tense items, subjects were directly followed by the verb (e.g., she play/plays). For progressive items, there were three possible combinations: subject followed directly by verb (e.g., Daddy singing), subject followed by the auxiliary 'is' and then verb (e.g., Daddy is singing), and subject plus contracted auxiliary 'is' and then verb (e.g., Daddy's singing).

For the purpose of identifying verb-marking errors, we extracted the full utterances of which each candidate 3sg and progressive subject+verb sequence formed a part from the prepared target corpora using the UNIX command-line tool *grep*. We (the three authors working together) then manually coded these candidate utterances for error using the following criteria: utterances containing a verb-marking error (i.e., where the unmarked form was used incorrectly in place of the marked form) were coded as 1 (e.g., baby go in there) and utterances where the correct marked verb form was used were coded as 0 (e.g., baby goes in there). Any utterances that were excluded were coded as NA to be removed during analysis (e.g., if the utterance was a question, except for tag questions). For a list of exclusion criteria and examples, see Supplementary Table 2 in Appendix S1.

To ensure coding was consistent for all utterances across both corpora, we applied a number of rules (e.g., retracings were only counted once, and items where the subject could be singular or plural [e.g., deer] were coded as singular unless obviously plural [e.g., all the deer go there]). For a list of coding rules, see Supplementary Table 3 in Appendix S1. During this process, we manually combined any variants of the same subject (e.g., 'Mum', 'Mummy', 'Mama') under one single term (e.g., 'Mum') and combined the names of all target children from both corpora under the single term 'Childname'. This yielded a total of 1673 different candidate 3sg subject+verb sequences (with a total of 8767 utterances) and 879 different candidate progressive subject+verb sequences (with a total of 3268 utterances) to which CDS statistics could be applied.

Adult Corpora

To capture bias statistics (the relative frequency of unmarked forms) which accurately reflect the nature of CDS in the English language, we took frequencies from a large collection of diverse corpora selected from the English UK and USA subsections of the CHILDES database (MacWhinney, 2000). In selecting the corpora, we applied a number of inclusion criteria (e.g., the corpora were not from atypically-developing children or clinical controls, and the corpora were not of child-adult interactions with adults other than caregivers). For a list of all inclusion criteria, see Supplementary Table 4 in Appendix S1.

Details of the 35 corpora which were included in the analysis, along with any amendments made to the original corpora, and the names of the target children that could be identified, are provided in Supplementary Table 5 in Appendix S1. The target child names that could not be identified (*N*=60) did not affect any CDS statistics as they did not appear in the transcripts.

To ensure that our CDS statistics were reliable, we based them on a very large dataset from a wider range of corpora than the target children. Moreover, since one of our inclusion criteria was that the input should be to TD children, the adults from the Manchester corpus were included but the adults from the Conti-Ramsden 3 corpus were not. While the question of whether individual variation in the input to particular children affected their language development is an interesting question, the samples of CDS for each target child were simply not big enough to pursue this question in the present study. Instead, we made the common assumption (e.g., Braginsky et al., 2019; Kueser et al., 2018; McCauley et al., 2021) that there is enough stability in the input across children that combining corpora in this way is a valid way of calculating these estimates.

To extract the CDS bias statistics, we used a Python script [available at: osf.io/ef8bm] which prepared the aggregated corpora for analysis using the same procedure as described in the child data section above. However, instead of analysing the child tier, the procedure analysed all caregiver tiers. Additionally, this script combined variants of the same subject under one single term (e.g., 'Dad', 'Father' 'Dada' 'Daddy' were all combined under the single term 'Dad') and combined all target children's names under one single term 'Childname'. For a list of subjects and their variants, see Supplementary Table 6 in Appendix S1. During this process, we applied a number of rules (e.g., candidate multi-word subjects were only treated as such when they were specifically marked, such as Pooh+Bear). For a list of rules and examples, see Supplementary Table 7 in Appendix S1.

CDS Statistics

Subject+verb sequence bias statistics were generated directly by the Python script. To calculate the verb bias in *3sg contexts* statistics (i.e., where the subject was always 3sg), we used the script to identify every possible 3sg subject+verb sequence for each individual verb (regardless of whether the subject+verb sequence also appeared in the child utterances or not). The frequency counts for each verb were then added together, and the bias – unmarked frequency divided by total frequency for the verb - was calculated (e.g., the verb bias in *3sg contexts* statistic for the verb 'help' was calculated using the frequencies from multiple 3sg subject+verb combinations such as 'Mum help', 'Dad help', 'he help", 'that help', etc.). To calculate the verb bias in *any* *context* statistics (i.e., regardless of whether the subject was 3sg or not), we used the UNIX command-line tool *grep* to identify the total number of times each verb appeared in the full set of CDS corpora, in either the bare stem or 3sg form. The bias was then calculated using these counts. To ensure we obtained as reliable an estimate of bias as we could without having to omit too much of our sample, we required that each subject+verb sequence and verb, in both *3sg contexts* and in *any context,* occurred at least 10 times (unmarked and marked form total) in our CDS sample to remain in the analysis.

The unmarked utterances from which each 3sg or progressive subject+verb sequence, or verb were derived were spot-checked for whether they were correct. We did not spot-check marked utterances on the assumption that these would be correct in CDS. Exceptions to this rule were the verbs ‘has’ and ‘does’, which were spot-checked for whether they were being used as an auxiliary or a main verb. We extracted the utterances using the UNIX command-line tool *grep*. However, it was not possible to manually code the full sample of CDS in this way for all items. Instead, we spot-checked a maximum of 10 utterances for each 3sg or progressive subject+verb sequence, or verb. This meant that if there were less than 10 instances of a subject+verb sequence or verb, then all were checked, while if there were more than 10 instances, a random sample of 10 utterances was selected.

These utterances were coded by the authors using the following criteria: utterances where the unmarked form of the verb was correctly used, or the marked forms ‘has’ and ‘does’ were used as a main verb, were coded as 1 and utterances where they were not (e.g., where the candidate verb in the utterance was a noun) were coded as NA and excluded. For a list of exclusion criteria and examples, see Supplementary Table 8 in Appendix S1. The verb bias in *any context* statistics were only coded for whether the candidate verb was used as a main verb or not.

After the utterances were spot-checked, we adjusted the unmarked and marked frequencies for each subject+verb sequence and verb based on the proportion of utterances deemed to be correct (i.e., those coded as 1). For example, if 10 unmarked utterances were spot-checked and 7 were coded as NA (excluded), we assumed that 70% of the total amount of unmarked utterances for that subject+verb sequence or verb would also be excluded. As a result, 0.7 times the total frequency count for that subject+verb sequence or verb was subtracted from the original count generated. Once this process was complete, if there were less than 10 instances of the subject+verb sequence or verb in total (unmarked and marked form combined), then these items were removed from the analysis.

Final 3sg and Progressive Data

This process yielded a total of 359 different 3sg subject+verb sequences (with a total of 5770 utterances) and 238 different progressive subject+verb sequences (with a total of 2222 utterances) remaining to be analysed. For a list of 3sg and progressive subject+verb sequences and their bias statistics, see Supplementary Table 9 in Appendix S1. Both raw and adjusted counts (from spot-checking) were used in the analyses outlined below to ensure that any pattern of results was not contingent on our coding decisions.

Inter-Rater Reliability

Reliability coding was performed for 20% of the tokens of the final datasets (3sg *N*=1554, Progressive *N*=445) by an independent coder. Agreement between the three authors' coding (working together) and the independent coder was 'excellent' for both the 3sg (Cohen's Kappa=0.92) and progressive (Cohen's Kappa=0.97) data.

Models

To evaluate the relationship between the bias statistics (subject+verb sequence bias, verb bias in *3sg contexts* and verb bias in *any context*) and verb-marking errors, we conducted mixed-effects logistic regression modelling in R (version 4.0.3) using the lme4 package (version 1.1-27.1; Bates et al., 2015). The main analysis (research questions 1 to 4) and sensitivity analyses were pre-registered on OSF [available at: osf.io/ef8bm]. In all models, the dependent variable was the presence of a verb-marking error (1) or correct use of a marked verb form (0). Any remaining NAs were removed from the analysis (3sg *N*=1387, progressive *N*=338). The final datasets and analysis code are available on OSF [available at: osf.io/ef8bm].

All three bias statistics were scaled and centred and then used as predictors (fixed effects) for predicting this binary dependent variable. In all models, there were by-child random effects on all terms, to reflect the fact that the target children may differ in the extent to which their verb-marking errors could be predicted by each of the different bias statistics. There was also a random effect of verb on the intercept. We preregistered including this term only in an additional analysis. However, to account for variance in verb-marking errors that may be caused by any other properties of individual verbs, we decided to also take this more conservative approach in our main analysis. We ran analyses over the 3sg and the progressive data separately. Only the TD data from the Manchester corpus were used to address research questions 1-3, because combining the datasets may have obscured any potential effects present in each dataset individually. However, a composite dataset (which combined the TD and DLD data) was used in addressing research question 4 to explore whether the bias statistics affected these two groups in a similar or different way. For completeness, models addressing research questions 1-3 for the DLD group alone can be found in Appendix S2. All reported significance tests were based on model comparisons in order to avoid the consequences of multicollinearity. For all research questions, except research question 1, an alpha of .05 was used.

To address research question 1, we constructed single predictor models for each of the bias statistics to help understand the relationship between these predictor variables and verb-marking errors. We ran Bonferroni-corrected likelihood-ratio tests to compare each of the single predictor models to otherwise identical models without the fixed effect term.

To address research question 2, we calculated AIC values for each of the single predictor models constructed in research question 1, and these were used to determine which bias statistic provided the best fit to the data. We considered the model with the lowest AIC value to be the 'best model' and ΔAIC values were used to infer the level of support for each of the remaining models. ΔAIC was the difference between the AIC of each of the remaining models and the AIC of the 'best model'. The conventional rules of thumb for interpreting the ΔAIC values are: values <2 indicate that the candidate model is almost as good as the best model, values 4-7 indicate considerably less support for the candidate model and values >10 indicate that there is no support for this model providing the best fit to the data (Burnham & Anderson, 2002; Fabozzi et al., 2014). Following this logic, it could be determined whether our 'best model' did indeed provide the best fit to the data.

To address research question 3, we performed likelihood ratio tests to determine whether each bias statistic was uniquely predictive of verb-marking errors. We used a drop-one procedure to establish which bias statistic(s) explained variance over and above that explained by any of the other bias statistics. The full model for the 3sg analysis included all three bias statistics (subject+verb sequence bias, verb bias in *3sg contexts* and verb bias in *any context*) as fixed effects and the full model for the progressive analysis included subject+verb sequence bias and verb bias in *3sg contexts* as fixed effects. We compared the full models to three subsequent models for the 3sg analysis and two subsequent models for the progressive analysis, each of which left out the fixed effect term for a different predictor variable (bias statistic).

To address research question 4, we added a group*bias interaction to the 'best' single predictor model (as determined by the AIC values in research question 2) to explore whether the effect of the best predictor (bias statistic) of verb-marking errors varied as a function of whether the child was TD or diagnosed with DLD. We ran this model over a composite dataset (which included both the TD and DLD data) and we used model comparisons to see whether adding the interaction term improved the fit to the data.

We ran sensitivity analyses to see whether any of the decisions made during the corpus analysis impacted the pattern of results. The full set of analyses were run with the following adjustments to the data: 1) sequences containing potential past tense verbs (e.g., hit) were removed from the data 2) sequences containing potential past tense verbs and potential past participles (e.g., come) were removed from the data 3) raw bias statistics were used instead of adjusted bias statistics (from spot-checking) and 4) sequences containing the subjects ‘Mum’ or ‘Childname’ were removed from the data.

Additional exploratory analyses were pre-registered on OSF [available at: osf.io/ef8bm] and run using absolute frequency statistics rather than bias statistics for each of the given variables (subject-verb sequences, verbs in *3sg contexts*, verbs in *any context*) as there has been some discussion about the contributions of both absolute and relative frequency to these types of errors (Tatsumi et al., 2018). To do this, we re-run the full set of analyses over the TD data from the Manchester corpus first, using total frequency variables (unmarked and marked frequencies added together) and then, using unmarked frequency variables (whilst controlling for total frequency). The analyses were repeated using log transformed and not log transformed values.

Finally, we ran non-pre-registered exploratory analyses over the TD data from the Manchester corpus, firstly to determine whether each bias statistic was uniquely predictive of verb-marking errors even when potential effects of age were considered by using a drop-one procedure. We also re-ran the full set of analyses (research questions 1-4) with the addition of a progressive verb bias in *any context* variable to see whether this affected the pattern of results. Progressive verb bias in *any context* was calculated to be the rate at which progressive verbs occur in the input in the bare form versus with the auxiliary ‘is’ (e.g., ‘eating’ vs. ‘is eating’).

Results

Descriptive Statistics

Table 1 reports descriptive statistics for each of the bias measures. As can be seen from the table, the mean bias varies between 0.36 and 0.88 and the median bias varies between 0.35 and 0.93. There is also substantial variation for each of the bias measures. Table 2 reports the error rate for both TD and DLD children. As can be seen from the table, there was a higher error rate for both groups in the 3sg data compared to the progressive data, but children with DLD only made more errors than TD children on the progressive utterances. The balance between the two outcomes (error or correct) was sufficient for conducting logistic regressions in both the 3sg and progressive data (King & Zeng, 2001).

Table 1

Descriptive Statistics for the Bias Measures

Data	Statistic	Mean bias	95% CIs	Median bias	Range
3sg	Subject+verb sequence bias	0.43	0.40;0.46	0.40	1 (0–1)
	Verb bias in 3sg contexts	0.43	0.39;0.47	0.46	0.91 (0–0.91)
	Verb bias in any context	0.88	0.86;0.90	0.93	0.66 (0.34–1)
Progressive	Subject+verb sequence bias	0.38	0.35;0.41	0.35	1 (0–1)
	Verb bias in 3sg contexts	0.36	0.33;0.39	0.36	0.75 (0.10–0.85)

Table 2

Descriptive Statistics for the 3sg and Progressive Data.

Data	Group	Number of errors	Total number of utterances	Error rate
3sg	TD	3018	3634	83%
	DLD	524	749	70%
	Total	3542	4383	81%
Progressive	TD	510	1532	33%
	DLD	186	352	53%
	Total	696	1884	40%

Research Question 1: Is the relative frequency of contextually appropriate unmarked forms in the input related to (TD) children's tendency to make verb-marking errors in their speech?

Table 3 shows single predictor models for each of the bias statistics. As can be seen from the table, in the 3sg analysis, all three bias statistics were significant predictors of the rate of verb-marking errors (see Table 2 for error rate). For all bias statistics, there was a 'small' to 'medium' effect size. In the progressive analysis, subject+verb sequence bias was also a significant predictor of the rate of verb-marking errors (see Table 2 for error rate) which had a 'small' effect size. Verb bias in *3sg contexts* was not, although the trend was still in the expected direction. The positive slope for all fixed effects indicates that more verb-marking errors were made as bias increased towards 1 (i.e., fully biased towards the unmarked form). Overall, these results provide robust evidence for input effects on verb-marking errors in the 3sg analysis, as well as some evidence for input effects in the progressive analysis.

Table 3

Results of the Single Predictor Models for the 3sg and Progressive Analysis.

Bias statistic	B	Std Error	χ^2	df	P-value	95% CIs
3sg						
Subject+verb sequence bias	0.88	0.17	14.53	1	<.001	0.52;1.24
Verb bias in 3sg contexts	0.59	0.18	10.84	1	<.001	0.23;0.97
Verb bias in any context	0.72	0.16	13.15	1	<.001	0.38;1.08
Progressive						
Subject+verb sequence bias	0.31	0.10	5.71	1	.017	0.11;0.50
Verb bias in 3sg contexts	0.22	0.13	2.68	1	.101	-0.05;0.47

Significance of terms assessed via a Bonferroni-corrected likelihood ratio test relative to null model. Confidence intervals were calculated in R using confint and the bootstrapping method with a total of 500 simulations. For some bias statistics, this created convergence errors. However, such errors never occurred in more than 10% of the simulations.

Research Question 2: Which of the three measures of relative frequency of unmarked forms (subject+verb sequence bias, verb bias in 3sg contexts and verb bias in any context) is the most valuable predictor of (TD) children's verb-marking errors?

In both the 3sg and progressive analyses, subject+verb sequence bias provided the 'best' fit to the data (as indicated by the smallest AIC value). The ΔAIC values for the other candidate models were greater than 10, indicating that there was no support for the other bias statistics providing as good a fit to the data (see Table 4). These results provide evidence for sequence effects on the patterning of both 3sg and progressive errors, which supports the CSI account. They also clarify the level at which these sequence effects occur.

Table 4

AIC and ΔAIC values for individual predictor variables

Predictor Variable	DAIC
3sg	
Subject+verb sequence bias	AIC*
Verb bias in 3sg contexts	81.96
Verb bias in any context	30.31
Progressive	
Subject+verb sequence bias	AIC*
Verb bias in 3sg contexts	14.30

AIC is the model with the smallest AIC value (best fit to the data).*

Research Question 3: Do the three measures of relative frequency of unmarked forms explain unique variance in (TD) children's verb-marking errors?

Table 5 shows the results of the drop-one analysis. As can be seen from the table, in the 3sg analysis, the drop-one method revealed that subject+verb sequence bias and verb bias in *any context* accounted for significant unique variance in verb-marking errors. However, verb bias in *3sg contexts* did not.

This suggests that verb bias in *3sg contexts* accounted for a subset of the variance explained by the other bias measures. For completeness, we investigated whether this overlap occurred with subject+verb sequence bias and/or verb bias in *any context*. To do this, the drop-one analysis was repeated first without subject+verb sequence bias in the comparison model, and then without verb bias in *any context*. Verb bias in *3sg contexts* accounted for significant unique variance when subject+verb sequence bias was removed from the comparison model ($\chi^2_{(1)}$=5.14, p=.023) but not when verb bias in *any context* was removed ($\chi^2_{(1)}$=1.09, p=.297). This suggests that verb bias in *3sg contexts* explained a subset of the variance which was also explained by subject+verb sequence bias.

In the progressive analysis, neither subject+verb sequence bias nor verb bias in *3sg contexts* accounted for significant unique variance (see Table 5). This suggests that the variance explained by these measures overlapped.

Overall, these results provide evidence for sequence effects over and above the more general effect of verb bias reported in the previous literature. They thus provide strong support for the CSI account. However, they also show that it is sometimes difficult to differentiate between the two sequence bias measures.

Table 5

Results for the 3sg and Progressive Data (Research Question 3).

Bias statistic	B	Std Error	χ^2	df	P-value	95% Cis
3sg						
Subject+verb sequence bias	0.80	0.18	11.98	1	<.001	0.46;1.21
Verb bias in 3sg contexts	-0.02	0.13	0.01	1	.914	-0.28;0.29
Verb bias in any context	0.56	0.15	11.24	1	<.001	0.25;0.87
Progressive						
Subject+verb sequence bias	0.24	0.16	1.87	1	.171	-0.04;0.56
Verb bias in 3sg contexts	0.08	0.18	0.22	1	.637	-0.28;0.41

Significance of terms assessed via a likelihood ratio test using a drop-one procedure. The full model (including all bias statistics as fixed effects) was compared to three subsequent models (3sg analysis) and two subsequent models (progressive analysis), each of which left out the fixed effect term for a different bias statistic. Confidence intervals were calculated in R using confint and the bootstrapping method with a total of 500 simulations. While this created convergence errors, such errors never occurred in more than 10% of the simulations.

Research Question 4: Does the effect of the best predictor variable of verb-marking errors vary dependent on whether the child is TD or diagnosed with DLD?

Table 6 shows the results of the interaction models for both the 3sg and progressive data. As can be seen from the table, in the 3sg analysis, adding an interaction term between subject+verb sequence bias and group (TD or DLD) improved the fit to the data, as confirmed using model comparisons between the interaction model and an otherwise identical model without the interaction term as a fixed effect. The model including the interaction term also provided the 'best' fit to the data (as indicated by the smallest AIC value). The ΔAIC for the model without the interaction term was 2.21. This indicates that there was either no effect, or a radically reduced effect, of subject+verb sequence bias on verb-marking errors in the DLD group.

In the progressive analysis, model comparisons revealed that adding the interaction term did not lead to a significant improvement (see Table 6), and that the model without the interaction term provided the 'best' fit to the data (as indicated by the smallest AIC value), which reflected the fact that the effect of subject+verb sequence bias on verb-marking errors was similar for both the TD and DLD groups. However, the ΔAIC for the model with the interaction term was only 0.75 (i.e., the difference between the two models was minimal). While these results provide stronger evidence for input effects in the TD children than the children with DLD, there is some evidence for an input effect in both groups.

Table 6

Results of the Interaction Models for the 3sg and Progressive Analysis.

Bias statistic	B	Std Error	χ^2	df	P-value	95% CIs
3sg						
Subject+verb sequence bias	0.01	0.30				-0.53;0.52
Group(TD)	0.09	0.69				-1.33;1.44
Subject+verb sequence bias*Group(TD)	0.83	0.34	4.22	1	.040	0.22;1.45
Progressive						
Subject+verb sequence bias	0.10	0.19				-0.00;0.50
Group(TD)	-0.60	0.71				-0.00;0.87
Subject+verb sequence bias*Group(TD)	0.21	0.21	1.25	1	.263	-0.00;0.61

Significance of terms assessed via a likelihood ratio test between a model with, and a model without, the interaction term as a fixed effect. Confidence intervals were calculated in R using confint and the bootstrapping method with a total of 500 simulations. While this created convergence errors, such errors never occurred in more than 10% of the simulations.

Sensitivity Analyses

We ran sensitivity analyses which 1) removed potential past tense verbs (e.g., hit), 2) removed potential past tense verbs and potential past participles (e.g., come), 3) used raw bias statistics instead of adjusted bias statistics (from spot-checking) and 4) removed sequences containing the subjects ‘Mum’ or ‘Childname’. In most cases, the same pattern of results emerged. One possible exception was when removing sequences containing the subjects ‘Mum’ or ‘Childname’. Here, for the 3sg data, the best single predictor model was verb bias in *any context* rather than subject+verb sequence bias. However, there was only ‘less’ support for subject+verb sequence bias (DAIC=5.48) and this measure still explained significant unique variance over and above that of verb bias in *any context*. This is consistent with the overall pattern of results reported in the main analyses and shows that the reported results were not driven by the specifics of particular coding decisions (see Appendix S3 for further details).

Exploratory Analyses

In the above analyses we used composite measures of bias (for any given subject+verb sequence or verb X, the frequency of the unmarked form of X divided by the total frequency of X). We also ran pre-registered exploratory analysis to explore whether we would see the same pattern of results when we put the two component frequencies into models separately. There was no difference in the observed pattern when analyses were repeated using total frequency (unmarked + marked forms) nor unmarked frequency (whilst controlling for total frequency) variables instead of bias statistics (see Appendix S4 for further details).

We also ran a non-pre-registered analysis to explore the effect of age on verb-marking errors. We found that while age was a significant predictor of the rate of verb-marking errors in both the 3sg and progressive datasets, all bias statistics accounted for significant unique variance in verb-marking errors even when taking age into account, with the exception of the verb bias in *3sg contexts* statistics in the progressive analysis. This result is consistent with the findings from the main pre-registered analysis and shows that our reported effects were not confounded with age (see Appendix S5 for further details).

Finally, we repeated the main analyses with the addition of a progressive verb bias in *any context* variable to see whether this affected the results. Overall, we found that there was no significant change to our overall pattern of results. In addition, while verb bias in *any context* was not a significant predictor of progressive verb-marking errors, there was a trend when using both raw (p=.048) and adjusted (from spot-checking) counts (p=.067). The positive slope in both cases indicates that more verb-marking errors were made as bias increased towards 1 (i.e., fully biased towards the unmarked form). This extends previous findings on the more general verb bias measures which, to our knowledge, has only been explored in 3sg data (see Appendix S6 for further details).

Discussion

In this paper, we investigated whether verb-marking errors such as ‘she play football’ and ‘Daddy singing’ arise from the high relative frequency of contextually appropriate unmarked words and word sequences in the input. We conducted a pre-registered corpus analysis testing the idea that children's verb-marking errors reflect the extraction of unmarked sequences from longer structures in the input (e.g., ‘Does she play football?’ and ‘I can hear Daddy singing’ respectively). We also investigated the level at which any such sequence effects occurred by using a subject+verb sequence bias measure (i.e., the rate at which specific subject+verb sequences occur in bare versus inflected form, e.g., ‘she play’ vs. ‘she plays’) and a verb bias in *3sg contexts* measure (‘any 3sg subject + play’ vs. ‘any 3sg subject + plays’). Finally, we investigated whether any observed effects of sequence frequency remained once a more general effect of verb bias was included in our models - a verb bias in *any context* measure (i.e., the rate at which verbs occur in the input in the bare versus 3sg form e.g., ‘play’ vs. ‘plays’).

Overall, we found support for the CSI account of both 3sg and progressive errors. In both analyses, all input measures performed significantly above chance in predicting errors (except for verb bias in *3sg contexts* in the progressive analysis), with more errors being made as bias increased. These findings provide strong support for the claim that verb-marking errors can be learned directly from the input as a result of the faulty processing of longer structures such as questions and double-verb constructions (e.g., Finneran & Leonard, 2010; Theakston et al., 2003). The best predictor variable was subject+verb sequence bias which provided a better fit to the data than the already established verb bias in *any context* measure (e.g., Kueser et al., 2018; Räsänen et al., 2014).

To give a concrete example, in the 3sg analysis, children made errors on sequences such as ‘Mum help’ (100% error rate) which had a high bias score (1.00) as they always occurred in the input in the unmarked form in questions such as ‘Shall mum help?’. By contrast, children made few errors (17% error rate) on sequences such as ‘He say’ which had a low bias score (0.30) as they occurred more often in the input in the marked form in declaratives such as ‘Let’s see what he says’. Similarly, in the progressive analysis, children made errors on sequences such as ‘Car coming’ (63% error rate) which had a high bias score (0.82) as it occurred more often in the input in the unmarked form in questions such as ‘Is there a car coming?’ and made few errors on sequences (11% error rate) such as ‘It raining’ which had a low bias score (0.20) as they occurred more often in the input in the marked form in declaratives such as ‘It’s raining now’. These findings both provide specific support for the CSI account and clarify the level at which the proposed effect of input occurs.

Overall, our findings suggest that there are input effects on children’s verb-marking errors. To relate back to the theoretical debate outlined in the introduction, this is consistent with Constructivist accounts of language learning. Constructivists argue that sequences such as ‘she play’ and ‘Daddy singing’ occur in the input within longer structures (e.g., ‘Does she play football?’ and ‘I can hear Daddy singing’ respectively) and that children’s verb-marking errors reflect the inappropriate use of these sequences. The best predictor variable in our study was the relative frequency of unmarked specific subject+verb sequences which provides evidence for these accounts. By contrast, our results are not consistent with Nativist accounts of language learning, such as Wexler's OI hypothesis, which argue that verb-marking errors are not shaped by the input but instead reflect innate differences between the child and the adult grammar.

Of course, the extent to which these findings generalise to other languages remains an open question. Sequence effects on verb-marking errors are likely to be particularly strong in English because of the relatively high frequency of 3sg + unmarked verb sequences in questions (e.g., Does Dolly want a biscuit?) as a result of subject-auxiliary inversion. Such effects are less likely in languages in which questions are predominantly marked using intonation (e.g., Spanish: '¿Dolly quiere una galleta?') or subject-main verb inversion (e.g., German: 'Will Dolly einen Keks?'). However, it is possible that sequence effects may result in agreement errors with conjoined subjects in Spanish (e.g., 'Mama y Papa quiere' on the basis of 3sg sequences such as 'Papa quiere') and in optional infinitive errors in German (e.g., 'Dolly tanzen' = 'Dolly to dance' on the basis of modal questions such as 'Kan Dolly tanzen?'). Freudenthal and colleagues have shown that a computational model (MOSAIC) that 'learns' optional infinitive errors from modal declaratives and questions can simulate cross-linguistic variation in the pattern of verb-marking errors in English, Dutch, German and Spanish (Freudenthal et al., 2007; 2009). The possibility that there are sequence effects on verb-marking errors in other languages is therefore an interesting question for future research.

In addition to establishing specific sequence effects, we also found evidence to support the more general verb bias effects that have already been reported (Kueser et al., 2018; Räsänen et al., 2014). While the verb bias in *any context* measure did not provide the best fit to the data, it still performed significantly above chance in predicting verb-marking errors and accounted for significant unique variance over and above that of the sequence effects in the 3sg analysis. We also extended these findings by looking for a verb bias in *any context* effect in the progressive analysis (i.e., the rate at which progressive verbs occur in the input in the bare form versus with the auxiliary is e.g., ‘eating’ vs. ‘is eating’) which, to our knowledge, has not been considered in the previous literature. Our results were less clear here, but still showed that more verb-marking errors were made as bias increased.

We ran exploratory analyses which investigated the effects of absolute frequency and age. The results showed that there was no difference to the basic pattern of results when analyses were run using absolute frequency variables rather than bias statistics, and that our bias statistics accounted for significant unique variance even when taking into consideration the effect of age.

We also conducted a number of pre-registered sensitivity analyses including a) removing potential past tense verbs (e.g., hit), b) removing potential past tense verbs and past participles (e.g., come), c) using raw instead of adjusted bias statistics (from spot-checking) and d) removing sequences containing the subjects ‘Mum’ or ‘Childname’. In virtually all cases, the same pattern of results emerged, which suggests that our findings were not driven by the specifics of particular coding decisions. One small variation was found when sequences containing the subjects ‘Mum’ or ‘Childname’ were removed from the analysis. In this case, there was a slight change to the pattern of results in the 3sg analysis. Here, subject+verb sequence bias was not the best performing model; instead, verb bias in *any context* was. It could, then, be argued that children make more verb-marking errors on such sequences because they treat ‘Mum’ as if it were the second-person singular subject ‘you’, and their name as if it were the first-person singular subject ‘I’, both of which take the bare form of the verb in English.

However, it is important to note that subject+verb sequence bias still explained significant unique variance over and above the verb bias in *any context* measure in this analysis. We can therefore be confident that the subject+verb sequence bias effect we found in the main analysis was not driven by the inclusion of these sequences; they only boosted the effect. This finding is consistent with our overall pattern of results and provides further support for the view that verb-marking errors are learned from unmarked sequences in the input.

Overall, we found stronger evidence for input effects in TD children compared to children with DLD. Although there was no interaction between subject+verb sequence bias and group in the progressive analysis, there was a significant interaction in the 3sg analysis, which reflected the fact that there was no, or a radically reduced, effect of subject+verb sequence bias in the DLD group. This finding contrasts with previous research, which has found that input effects were apparent, and in fact stronger, for children with DLD, who may be more likely to learn frequent sequences from the input without appreciation of the wider linguistic context (e.g., Leonard & Deevy, 2011; Leonard et al., 2015). However, it is compatible with Constructivist accounts of language learning which argue that all children re-use sequences taken directly from the input (as in the use of 'she play' and 'Daddy singing' in inappropriate contexts) and not just those experiencing difficulties in their grammatical development.

It is also the case that the DLD corpus we used (the Conti-Ramsden 3 corpus), although currently the richest available on the CHILDES database, is still considerably sparser than the TD corpus (the Manchester corpus). In total, the children with DLD only provided 1178 utterances in the 3sg data (46% of which used the verb 'go') and 352 utterances in the progressive data (41% of which used the verb 'going'). In comparison, the TD children produced 3634 utterances in the 3sg data and 1532 utterances in the progressive data. The most common verbs were also ‘go’ and ‘going’ which were used in 26% and 41% of the utterances respectively. The limited number of utterances, as well as the limited range of different subject+verb combinations in the DLD data, is likely to have reduced the chances of finding input effects, and, as a result, we should interpret the difference in our results for TD and DLD children with caution. It also underlines the need for richer naturalistic corpora for children with DLD to inform our understanding of the relation between the errors that they make and the distributional properties of the language that they are learning.

The main advantage of conducting a corpus analysis, rather than using an experimental method, was that it allowed us to look for a relationship between real verb-marking errors in naturally occurring children’s speech and real CDS. The risk of unmeasured confounding is of course greater with observational data. However, we took a number of measures to minimise this risk. The use of detailed preregistration minimises any impact of researcher bias. We included a number of sensitivity checks in which the data were coded in a variety of different ways, to check that our findings were not the results of bias in our coding decisions. Furthermore, the focus of the study on predicting errors of commission sidesteps at least one confounding pathway that typically exists when input frequency is claimed to affect production. Where a relationship is observed between the frequency with which a caregiver produces a particular form and the facility with which a child produces that form, there is a clear risk that an unmeasured third variable is actually causing the privileged status of the form for both speakers. Since we are looking at verb-marking errors, which by their very nature do not typically occur in CDS (so that the child is doing something that the caregiver is not), we can discount this ubiquitous confounding pathway.

Conclusion

The aim of this study was to test the idea that verb-marking errors such as ‘she play football’ and ‘Daddy singing’, which are a hallmark of young children’s speech, can be understood as errors of commission driven by frequency in the input that they receive. Our results provide strong support for this idea, specifically evidence that the probability of seeing an error for a particular form is related to the relative frequency (in the input) of marked and unmarked forms of the verb and the relative frequency of marked and unmarked subject+verb sequences. Our results thus provide support for the CSI account of children’s verb-marking errors, as well as clarifying the level at which CSI effects occur. They also illustrate the power of focused and pre-registered corpus analyses in illuminating the relation between the errors children make and the distributional properties of the input they receive.

References

Aguado-Orea, J. & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS One*, *10*(3). https://doi.org/10.1371/journal.pone.0119613.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The cross-linguistic study of sentence processing* (pp. 3–73). New York: Cambridge University Press. https://doi.org/10.2307/1423201.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind: Discoveries in Cognitive Science*, *3*, 52-57. https://doi.org/10.1162/opmi_a_00026.

Brown, R. (1973). *A First Language: The Early Stages.* Cambridge, MA: Harvard University Press.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag. https://doi.org/10.1007/b97636.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. New Jersey: John Wiley & Sons. https://doi.org/10.1002/9781118856406.

Finneran, D. A., & Leonard, L. B. (2010). Role of linguistic input in third person singular -s use in the speech of young children. *Journal of Speech, Language and Hearing research*, *53*(4), 1065-1074. https://doi.org/10.1044/1092-4388(2009/09-0056).

Freudenthal, D., Pine, J. M. & Gobet, F. (2009). Simulating the referential properties of Dutch, German and English root infinitives. *Language Learning and Development, 5*, 1–29. https://doi.org/10.1080/15475440802502437.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science*, *31*(2), 311- 341. https://doi.org/10.1080/15326900701221454.

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton, NJ: Princeton University Press. https://doi.org/10.2307/j.ctvc772nn.

Hoekstra, T., & Hyams, N. (1998). Aspects of root infinitives. *Lingua*, *106*(1-4), 81-112. https://doi.org/10.1016/S0024-3841(98)00030-8.

Hyams, N. (1996). The underspecification of functional categories in early grammar. In H. Clahsen (ed). *Generative perspectives on language acquisition* (pp. 91-127). Amsterdam: John Benjamins. https://doi.org/10.1075/lald.14.07hya.

Joseph, K. L., Serratrice, L., & Conti-Ramsden, G. (2002). Development of copula and auxiliary BE in children with specific language impairment and younger unaffected controls. *First Language, 22*(2), 137-172. https://doi.org/10.1177/014272370202206502.

King, G. & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, *9*(2), 137-163.

Kueser, J. B., Laurence, L. B., & Deevy, P. (2017). Third person singular –s in typical development and specific language impairment: input and neighbourhood density. *Clinical Linguistics and Phonetics, 32*(3), 232-248. https://doi.org/10.1080/02699206.2017.1342695.

Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, *14*(3), 315-344. https://doi.org/10.1080/10489220701471081.

Leonard, L. B. (2014). *Children with Specific Language Impairment (2nd Edition).* Cambridge, NA: MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001.

Leonard, L. B., & Deevy, P. (2011). Input distribution influences degree of auxiliary use by children with specific language impairment. *Cognitive Linguistics, 22*(2), 247-273. https://doi.org/10.1515/COGL.2011.010.

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of english-speaking children with specific language impairment. *Journal of Speech, Language and Hearing Research, 40*(4), 741-752. https://doi.org/10.1044/jslhr.4004.741.

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.

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua, 120*(11), 2546–2556. https://doi.org/10.1016/j.lingua.2010.06.005.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk (3rd Edition)*. Mahway, NJ: Lawrence Erlbaum. https://doi.org/10.1177/026565909200800211.

McCauley, S. M., Bannard, C., Theakston, A., Davis, M., Cameron-Faulkner., & Ambridge, B. (2021). Multiword units lead to errors of commission in children's spontaneous production: "What corpus data can tell us?", *Developmental Science*, *24*(6). https://doi.org/10.1111/desc.13125.

Matthews, D. E., & Theakston, A. L. (2006). Errors of omission in english-speaking children's production of plurals and the past tense: the effects of frequency, phonology, and competition. *Cognitive Science*, *30*(6), 1027-1052. https://doi.org/10.1207/s15516709cog0000_66.

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., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research, 38*(4), 750-863. https://doi.org/10.1044/jshr.3804.850.

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*(6), 850-863. https://doi.org/10.1044/jslhr.4106.1412.

Rizzi, L. (1993). *Some notes on linguistic theory and language development: The case of root infinitives. Language Acquisition, 3(4),* 371-393. https://doi.org/10.1207/s15327817la0304_2.

Tatsumi, T., Ambridge, B., & Pine, J. M. (2017). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: an elicited production study of Japanese. *Cognitive Science*, *42*(2), 555-577. https://doi/10.1111/cogs.12554.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb argument structure: an alternative account. *Journal of Child Language, 28*(1), 127-152. https://doi.org/10.1017/S0305000900004608.

Theakston, A. L. Lieven, E V., & Tomasello, M. (2003). The role of the input in the acquisition of the third person singular verbs in English. *Journal of Speech, Language and Hearing Research, 46*, 863-83. https://doi.org/10.1044/1092-4388(2003/067).

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press. https://doi.org/10.2307/j.ctv26070v8.1.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations. In D. Lightfoot, N. Hornstein (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. https://doi.org/10.1016/S0024-3841(98)00029-1.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Corpora Preparation and Coding Details

Appendix S2. DLD Analysis

Appendix S3. Sensitivity Analyses

Appendix S4. Absolute Frequency Analysis

Appendix S5. Age Analysis

Appendix S6. Progressive Verb in Any Context Bias Analysis