Title
Double PP Constituent Ordering Preferences in English Early Child Language

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Liu, Zoey
Namdar, Lauren E
Wulff, Stefanie
et al.

Publication Date
2023

Peer reviewed

Double PP Constituent Ordering Preferences in English Early Child Language

Zoey Liu
University of Florida

Lauren Namdar
University of California, Davis

Stefanie Wulff
University of Florida

Kenji Sagae
University of California, Davis

Abstract

What determines children's production of syntactic alternations? This study takes up this question with the double PP construction in English as the test case (e.g., ***write*** *[$_{PP_1}$ on the paper] [$_{PP_2}$ with this new pen]*). Leveraging data of spontaneous child-parent interactions, we investigate the roles of dependency length and parent input frequency, with the latter being operationalized as lexical frequency and contextual predictability. We found that when child and parent data was combined, all three factors turned out to have significant predictive power, with dependency length having the most pronounced role. Results from the developmental trajectory of DLM as well as logistic regression analysis suggest that child production of constituent ordering preferences starts becoming more parent-like after the age of 30 months.

Keywords: syntactic alternation; dependency length; input frequency; child language

Introduction

When two constituent orderings are permissible, language users' choice for one or the other ordering is known to be multifactorially governed: different linguistic constraints jointly determine constituent ordering preferences. This raises the question when and how children acquire these diverse constraints, their relative weightings, and their potential interactions as children are acquiring alternating constituent orders.

While syntactic alternations have been researched extensively, the focus has mostly been on written data produced by adults (Heller, Bernaisch, & Gries, 2017; Gries & Adelman, 2014; Gries, 2017; Hawkins, 2014; Wasow, 1997); investigations of constituent ordering preferences in spoken data are comparatively few, with most existing studies focusing on the English dative alternation (Bresnan, Cueni, Nikitina, & Baayen, 2007; Engel, Grafmiller, Rosseel, & Szmrecsanyi, 2022; Szmrecsanyi et al., 2017).

Even fewer studies have investigated syntactic ordering preferences in child speech, most of them again limited to the dative alternation in English (De Marneffe, Grimm, Arnon, Kirby, & Bresnan, 2012; van den Bosch & Bresnan, 2015; Liu & Wulff, 2023). De Marneffe et al. (2012) presented a multifactorial analysis of a small data set with 530 utterances taken from child speech. Using the same data set, follow-up experiments by van den Bosch and Bresnan (2015) examined individual variation of children's production of the dative alternation. In recent work, Liu and Wulff (2023) studied the developmental trajectories of dependency length minimization (Ferrer-i Cancho, 2004) in child spoken data, looking at 8,046 instances of the dative construction.

This study contributes to existing literature on constituent ordering preferences in early child language in English, using the double prepositional phrase (hereafter PP) construction as the test case (e.g., ***write*** *[$_{PP_1}$ on the paper] [$_{PP_2}$ with this new pen]*). Using transcripts of naturalistic child-parent interactions, we examine the roles of dependency length (Gibson, 1998) and parent input frequency (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Barnes, Gutfreund, Satterly, & Wells, 1983). While our focus is on child speech, we considered parents' production patterns as developmental benchmarks and analyzed their speech as well where permissible.

Related Work

Dependency length

The principle of Dependency Length Minimization (hereafter DLM (Ferrer-i Cancho, 2004)) predicts that words or phrases that are syntactically dependent on each other prefer to occur closer together, which in turn minimizes the overall dependency length of the sentence and thereby facilitates online processing efficiency (Gibson et al., 2019).

There has been considerable amount of work exploring the effect of dependency length in syntactic variation (Gildea & Temperley, 2010; Gulordava & Merlo, 2015; Hawkins, 1990; Liu, 2020); and the preference for DLM has been claimed to be a syntactic universal for human languages (Futrell, Levy, & Gibson, 2020; Futrell, Mahowald, & Gibson, 2015). However, prior studies have mostly examined written data, with only a few exceptions. Liu (2019) compared syntactic orderings in spontaneous spoken English to those in (formal) writing; the results demonstrated that the preference for shorter dependencies is significantly weaker in spoken domain. Kramer (2021) analyzed spoken data in seven languages taken from YouTube channels; his findings illustrated that the tendency for DLM is weaker in speech than in written texts for head-initial languages (e.g., English), yet the opposite pattern was observed for head-final languages. Aside from looking at corpora of spoken data, others (Liu, Upreti, Kramer, & Namboodiripad, 2022) have studied the role of dependency length in predicting acceptability judgments, using *audio* stimuli; however, no pronounced effect was found for dependency length.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

These mixed results suggest that DLM may be modality-specific. Our study contributes to this line of work; examinations of naturalistic child-parent interactions not only allow us to explore DLM in spoken registers further (at least in English), but also help us understand how ordering preferences vary across different interlocutors.

Input frequency

While little is known about the relation between parental input and children's acquisition of syntactic alternations specifically, in more general terms, parental input frequency has been found to impact syntactic development in children in various ways. For instance, the amount of parental speech is positively correlated with child mean length of utterance (Barnes et al., 1983) as well as with more rapid syntactic development in first-borns, as they tend to receive more parental input than their siblings (Erika, 1997; Hoff-Ginsberg, 1998). Also, parents fine-tune their input to their child's language ability: overall they introduce increasingly complex syntactic constructions gradually, and while in earlier stages, they provide children with a less diverse set of lexical realizations of a syntactic construction and restrict input to repeated exposure to prototypical renderings that aid children in figuring out meaning (Goldberg, Casenhiser, & Sethuraman, 2004), they then gradually introduce a more diverse set of lexical realizations of a given syntactic construction; in short, high informativity and low diversity in earlier stages are replaced by low informativity and high diversity in later stages (Sethuraman, 2004).

Here we operationalize parent input frequency in two ways: lexical frequency and contextual predictability. Compared to dependency length, investigations of these two factors in syntactic orderings have received little attention. Gustafsson (1976) demonstrated that the more frequent word tends to occur first in English binomials; similarly, Wulff (2003) showed the more frequent adjective prefers to appear first in English adjective orderings. These studies, however, focused on constituent order at the lexical level, instead of attending to larger/more complex syntactic structures.

Besides lexical frequency, previous research has looked at both structural predictability (Rajkumar, van Schijndel, White, & Schuler, 2016) and lexical predictability (Levy & Jaeger, 2007) in syntactic alternations. The former attends mainly to the structural level; for example, for a dative construction which can be realized as a double object structure or a prepositional object structure, which structure is more predictable given the preceding syntactic parse tree structure (Rajkumar et al., 2016)? Lexical predictability, on the other hand, focuses on how likely it is that certain word(s) will appear given preceding lexical context (e.g., whether the relativizer will be omitted in relative clauses (Levy & Jaeger, 2007)). Given the structural properties of the double PP construction, we use lexical predictability as a proxy for contextual predictability in this study.

Experiments

Data and preprocessing

We resorted to the CHILDES database (MacWhinney, 2000) for child and parent production data. We extracted transcripts of naturalistic child-parent conversations from the English-NA and the English-UK sections of CHILDES, accessed via the `childes-db` interface (Sanchez et al., 2019). In particular, we focused on utterances produced by typically developing children and their parents. The part-of-speech (POS) tags of these utterances were automatically derived with the publicly open natural language processing library Stanza (Qi, Zhang, Zhang, Bolton, & Manning, 2020); each utterance was then automatically assigned a syntactic dependency tree using DiaParser (Attardi, Sartiano, & Yu, 2021); this dependency parser has been demonstrated to achieve good performance for child spoken language in English (Liu & Prud'hommeaux, 2022).

We extracted double PP constructions from these utterances (e.g., (1a)[1]), taking advantage of their POS tags and dependency parses. Specifically, we searched for VPs where the head verb has exactly two PP dependents occurring on the same side and the two PPs appeared adjacent to each other. The POS tag of the head verb in the VP was always VERB, which denotes lexical verbs (excluding auxiliaries). The nominal head of the PP had one of four POS tags: NOUN (lexical noun), NUM (numeral), PRON (pronoun), and PROPN (proper noun), and the dependency relation between the PP and the head verb was always oblique.

(1) a. **write** [$_{PP_1}$ on the paper] [$_{PP_2}$ with this new pen]

b. **write** [$_{PP_1}$ with this new pen] [$_{PP_2}$ on the paper]

After the initial round of data extraction, an author of this paper with advanced training in dependency linguistics manually went through every utterance for inspection. Overall, four types of instances were excluded from analysis, illustrated below (relevant elements are in italics): cases where the second PPs were repetitions of the first, e.g., (2a); cases where the first PPs were partial repetitions of the second, as

[1]Examples provided here are adapted from utterances initially extracted from CHILDES.

in (2b); when the nominal heads of the PPs were discourse fillers but they were not assigned the correct POS tags automatically, as in (2c); and lastly, utterances like (2d), where the parser erroneously parsed the word *like*, a discourse marker, as a preposition. (Note that the total number of cases similar to (2c) and (2d) is smaller than 10).

(2) a. go [with me] [*with me*].
b. put it [*in a*] [in a line].
c. eat [*with ah*] [near the table]
d. it got [*like bumps*] [on it].

Removal of such cases led to a total of 12,035 utterances subjected to analysis (Child: 1,817; Parent: 10,218). Although our focus is on child speech, we turned to parent production as references in analysis where necessary. Although prior study has shown that English has both head-initial (two PPs occurring postverbally) and head-final (two PPs appearing preverbally) double PP structures (Liu, 2022), our data set turned out to contain only the former.

Measures for factors

Following the methodologies outlined in Liu (2022), we examined the effects of three linguistic factors in PP ordering preferences: dependency length (Ferrer-i Cancho, 2004), lexical frequency (Wulff, 2003), and contextual predictability (Levy & Jaeger, 2007). In what follows, we describe our measures of each factor.

Dependency length Take examples (1a) and (1b) as illustrations. Both have the same head verb, *write*, with the same two PP dependents appearing postverbally; the order of the PPs in the two utterances is switched. Based on predictions by DLM, the order of (1a) will be preferred since the PP of shorter length, *on the paper*, is placed closer to the head verb, thus shortening the overall dependency length of the sentence.

Here we ask whether dependency length affects PP orderings. For each original utterance in our data set, we first measured its overall dependency length. We then constructed its syntactic alternative via switching the order of the two PPs (e.g., construct (1b) based on (1a)), and measured its overall dependency length. Lastly, we computed the difference of overall dependency length between the original utterance and its structural alternative; a positive value indicates that the original utterance has shorter dependencies, thus supporting the prediction of DLM.

In their examinations of the dative alternation in English, Liu and Wulff (2023) found a strong tendency for DLM across children at different developmental stages (indexed by age). To see whether the same preference is also observable for the double PP construction, we likewise used age as an index of children's developmental stage. We created age bins spanning 6-month intervals and grouped all double PP structures produced by children (and by parents accordingly) into their corresponding age bins. For all the data within each age bin, we calculated the DLM ratio, DLM_r. We counted the number of utterances that demonstrate a preference for shorter dependencies (N_{short}) and the number of instances where the opposite pattern holds (N_{long}). Then DLM_r was computed as $\frac{N_{short}}{N_{long}}$; significance testing was conducted using bootstrapping (Efron & Tibshirani, 1994) with 10,000 iterations. A DLM_r value significantly larger than 1 indicates that there is a pronounced preference for shorter dependencies.

While our main focus is on child speech, we also calculated corresponding DLM_r measures for parent production.

Lexical frequency Prior work has demonstrated that there is a preference for the constituent of higher frequency to occur first in a syntactic alternation (Gustafsson, 1976; Wulff, 2003). With double PP constructions, the general prediction will be that the more frequent PP will tend to occur first.

Our approximation of the lexical frequency for each PP in child production data is as follows (Braginsky, Yurovsky, Marchman, & Frank, 2019): (1) given a double PP structure in our data set, we noted which corpus in CHILDES the structure is from; (2) we estimated unigram counts of all lemma types in parent speech (not just parents' production of the double PP construction) from that corpus; (3) for each PP in the structure, we took the unigram count of the adposition and that of the lemma of the lexical head; for every case where their unigram count was missing, we assigned it with a frequency of 1; the unigram counts were then normalized by the number of tokens in the corpus where the structure is from; (4) the lexical frequency of each PP was computed as the product of the unigram probability (log transformed) of the preposition and that of the lemma of the lexical head.

$$Lexical\,Frequency(on\,the\,paper) = P(on) * P(paper) \qquad (1)$$

After obtaining the lexical frequency of both PPs in a double PP structure this way, we calculated the difference of their lexical frequencies; a positive value means that the PP of higher frequency appears first.

Note again that lexical frequency functions here as an approximation of input frequency. While it is easy enough to estimate lexical frequencies of the input children receive from their parents as done here, the lexical frequencies of the input directed to parents could not be estimated the same way, for the plain reason that this data is not available in CHILDES: there is no conversational data between the parents and their own parents, or between the parents and other parents/caregivers. To remedy that, we chose to approximate the lexical input frequencies of parents using data from the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992), which contains naturalistic telephone conversations between adults, and the spoken section of the International Corpus of English (ICE) (Greenbaum, 1988); preprocessing led to $\sim$1.6 million tokens.

Contextual predictability The general prediction of contextual predictability here is that the PP that is more predictable given preceding sentential context will prefer to occur first. The computation for the contextual predictability of each PP is similar to that of lexical frequency, which was measured as the product of the respective conditional probability of the preposition and the lexical head given preceding sentential context. For instance, the contextual predictability of every PP in example (1a) is computed as follows.

$$Pred(on\,the\,paper) = P(on|write) * P(paper|on\,write) \quad (2)$$

$$Pred(with\,this\,new\,pen) = P(with|write) * P(pen|with\,write) \quad (3)$$

To estimate word predictability, we turned to neural language models (LM), which generate word-by-word conditional probabilities and are able to deal with (training) data sparsity much better compared to *n*-gram LMs. In particular, we used long short-term memory (LSTM) models (Hochreiter & Schmidhuber, 1997). We first extracted *all parent production data* in English from CHILDES ($\sim$18 million tokens); then split it into training/validation sets at a 4:1 ratio (when approximating predictability for parent data, we again relied on *the combined Switchboard and ICE data*). We used the training set to build LSTMs with different combinations of embedding sizes ({50, 100, 200}), numbers of hidden layers ($\{1, 2\}$), and numbers of hidden units in each layer ({50, 100, 200}). Each model was trained with a batch size of 20. These models were evaluated using the test set with early stopping. Based on their evaluation performance, the final model architecture had an embedding size of 100 and two hidden layers, with 50 hidden units in every layer (see also Portelance, Degen, and Frank (2020)).

For each double PP structure, we computed the differences in the contextual predictability of the two PPs; a positive value suggests that the more predictable PP occurs first.

Regression modeling

To estimate the predictive effect of each factor, we applied logistic regression modeling. We randomly selected half of the original utterances produced by children from our data set; these utterances stayed as they were and we coded their ORDER as 1. For each utterance in the remaining half, we constructed its structural variant via switching the order of the two PPs, then coded the ORDER of the variant as 0. Hence in the final data that we subjected to regression analysis, half were the original utterances, meanwhile the other half were the constructed syntactic alternatives. The ORDER of these utterances was the dependent variable in the regression model.

With fixed-effects, we included the measures of the three factors, namely, dependency length, lexical frequency, and contextual predictability. For instance, with lexical frequency, for each utterance, we used the *difference value* between the lexical frequencies of the two PPs as the fixed effect. At the same time, we controlled for the effects of the definiteness and the pronominality of the lexical head in each PP, along with the age of the child and utterance length. Note that the child utterances that we analyzed ($N = 1,817$) were produced by more than 246 children (we could not count the exact number due to that some speaker information is missing in CHILDES), which is a relatively large number. This means that we did not have enough data points for most individual children to include them as random effects. (Of all the children, 127 each produced just one double PP structure.)

The regression model was trained to predict the original utterance order (ORDER = 1). The final formula for the model was determined with step-wise forward regression (reaching the maximal structure with the aforementioned factors and their interactions) by comparing the Akaike Information Criterion (AIC) score of different regression models. We performed logistic regression for parent data in the same way. (In this case, the age factor for each utterance produced by the parent corresponds to the age of their child at the time.)

In our regression models, a coefficient value significantly larger than 0 means the factor has a positive effect on the order. Specifically, a positive coefficient value for, e.g., lexical frequency, indicates that the larger the lexical frequency difference is between the two PPs, the more likely the more frequent PP appears first.

Results

Regression analysis

Results derived from the optimal logistic regression model for child data and that for parent data are presented in Table 1. Each coefficient value can be interpreted in two respects. Take dependency length as an example. It appears that child production of the double PP construction shows a pronounced preference for shorter dependencies ($\beta = 0.58$) when looking at just child data; in addition, this indicates that the larger the length difference is between the two PPs in each structure, the stronger the tendency for the PP of shorter length to appear closer to the head verb is. This corresponds to previous findings in Hawkins (1990) that looked at PP orderings in written English yet lacked proper statistical tests. In contrast, we found no significant effects for lexical frequency and contextual predictability for child speech when analyzing child data alone.

Table 1: Coefficient values (β) and 95% confidence intervals (CIs) for each factor of interest in child and parent speech; these results are based on individual optimal logistic regression models fit to child and parent data separately.

Role	**Factor**	β	95%CI
Child	dependency length	0.58	(0.44, 0.73)
	lexical frequency	0.11	(-0.01, 0.24)
	predictability	0.08	(-0.05, 0.21)
Parent	dependency length	0.46	(0.41, 0.52)
	lexical frequency	0.22	(0.18, 0.26)
	predictability	0.07	(0.02, 0.12)

On the other hand, when analyzing just parent data, the three factors all have significant effects for parent production

(Table 1). Similar interpretations can be drawn for lexical frequency and contextual predictability as we did for dependency length with child production; namely, there is a preference for the more frequent or more contextually predictable PP to appear first (and closer to the head verb), and that this preference is modulated by the difference value of the particular factor between the two PPs. Looking across the three factors, dependency length has the strongest predictive power, followed by lexical frequency, which has a more pronounced role than contextual predictability.

Since the results shown in Table 1 are based on separate regression analyses of child and parent data, these results do not speak to potential systematic differences in the effect for each factor when comparing speaker roles (child or parent). Thus, we combined child and parent data and performed logistic regression modeling the same way as described in the previous section, except this time we included speaker role (child or parent) as an additional factor. The optimal model after step-wise regression turned out to not contain significant interactions between speaker role and each of dependency length, lexical frequency, and contextual predictability, suggesting that there are no systematic differences in the roles these factors play in child and parent speech. Overall, the model showed significant effects for all three factors, indicating that *when considering child and parent production together*, dependency length, lexical frequency, and contextual predictability have significant roles in PP ordering.

Table 2: Coefficient values (β) and 95% confidence intervals (CIs) for each factor of interest; these results are based on the optimal logistic regression model fit to the combination of child and parent data.

Factor	β	95%CI
dependency length	0.47	(0.42, 0.52)
lexical frequency	0.22	(0.18, 0.26)
predictability	0.06	(0.02, 0.10)

The developmental trajectory of DLM

Now we zoom into the developmental trajectory of DLM in the double PP construction in child speech. Figure 1 contrasts the values of DLM ratio (DLM_r) in child and parent production within each age bin. (There were fewer than 100 instances produced by children within the age ranges of 12 to 18 months and 18 to 24 months, as well as fewer than 100 utterances by parents when children are above 72 months old). It seems that there is no pronounced difference regarding the values of DLM_r across most age bins; in addition, DLM_r does not become significantly larger than 1 until children are within the age range of 54 to 60 months. In other words, patterns of DLM did not arise robustly in the double PP construction before this age range, at least based on the data that we have. At the same time, the average DLM_r oscillates upwards (mostly) after the age range of 30-36 months, indicating a trend towards DLM starting that age.

When DLM emerged between 54 to 60 months in child production, the DLM_r (1.70 (1.37, 2.10)) is already comparable to that in parent production (1.49 (1.11, 2.00)). In fact, the extent of DLM in this age range is also comparable to that observed in naturalistic (telephone) conversations between adults (1.83 (1.44, 2.41)) (Liu et al., 2022), measured from the Switchboard corpus (Godfrey et al., 1992).

To add statistical rigor to our results as well as to investigate what linguistic constraints possibly lead to the varying extents of (or lack of) DLM in different age ranges, we fit linear regression models to *all the original utterances* produced by children. (Thereby this regression data did not include any constructed structural variants as we did when predicting PP orders.) For each original utterance, we computed its dependency length difference with their structural variant, as well as the difference values of lexical frequency and contextual predictability between the two PPs. The regression models predict the dependency length differences with all the other fixed effects (and their interactions) we used when fitting logistic regression for the PP ordering preferences. The optimal model after step-wise regression shows significant effects for utterance length, lexical frequency, along with the pronominality and definiteness of the two PPs.

These results have several implications. First, the lack of a pronounced effect for the age factor supports our initial descriptive analysis that the preference for shorter dependencies does not seem to depend on age. Second, the longer the utterance length is, the stronger tendency there is for shorter dependencies. Given that (mean) utterance length can also be taken as an index of children's developmental stage, this suggests as children's linguistic production develops (e.g., develops to produce longer utterances), they do tend to show a more pronounced preference for DLM, yet this is not exactly bound by age alone. Third, the different extents of DLM also result from the joint efforts of other factors. In particular, there is a tendency for the PP that is closer to the head verb to be pronominal and definite, which, accordingly, affects the overall dependency length of the utterance.

The overall patterns for DLM across different age ranges of children go against the findings reported in Liu and Wulff (2023), which demonstrated strong preferences for DLM across the developmental trajectory in children 18 months and older. One plausible explanation is the total number of analyzable instances in child speech is much smaller for the double PP construction in comparison to the number of dative structures ($N = 8,046$) examined in Liu and Wulff (2023). Keeping that in mind, these results also suggest the extent of DLM is construction-specific, that between the dative alternation and the double PP construction, the preference for DLM is possibly stronger in the former. We can speculate regarding underlying reasons. First, the dative alternation is arguably more syntactically complex in the sense that it includes core arguments such as the direct (and indirect) objects, and constructing the structural variant for a dative structure involves more than just simply switching the order of the constituents,

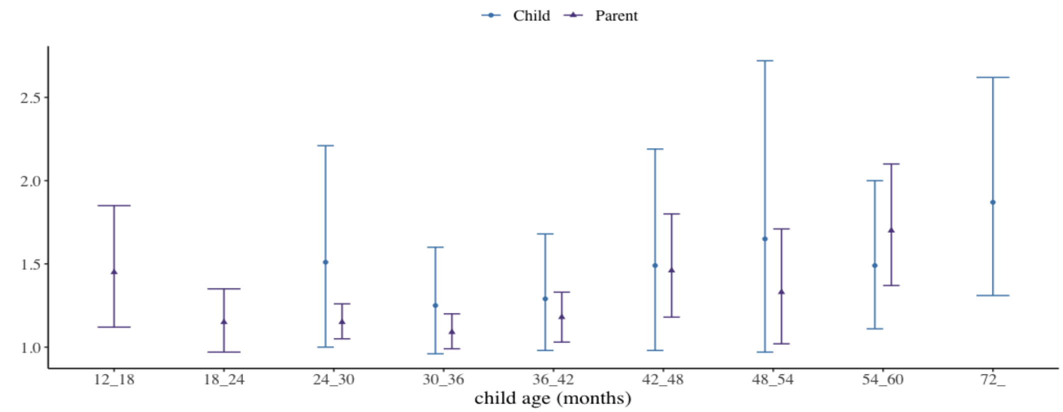

Figure 1: DLM ratios in the double PP construction in child and parent speech; a missing error bar within a specific age bin means there were fewer than 100 instances produced by that particular speaker role (child or parent) during that age range.

as is the case with the double PP structures. Second, because of the multifactorial nature of syntactic orderings, other factors that influence the dative alternation, such as verb semantic class and animacy, might modulate the effect of DLM. This is to be tested via regression analysis on the dative alternation, which we leave for future work.

Discussion

Using the double PP construction as the test case, we examined the roles of dependency length and parent input frequency in predicting constituent ordering preferences in child speech, relying on parent production as a reference point. We approximated parent input frequency with two measures, lexical frequency and contextual predictability. When child and parent data was combined, the logistic regression analysis indicates that all three factors have significant predictive power, with dependency length being the most effective. In addition, the effect for each of the three factors does not differ systematically between child and parent, suggesting that the role of dependency length, lexical frequency and contextual predictability is not constrained by specific speaker roles.

As a way to strengthen the observations above, we took the optimal logistic regression model derived from the combination of child and parent double PP utterances, then measured its prediction accuracy for the utterances produced by each individual speaker role. The model achieved an accuracy of 0.66 (out of 1) for child data, and 0.63 for parent data. These numbers, albeit reaching a reasonable level, suggest that there is still some amount of variation in the data that is not explained or captured by the all the linguistic constraints investigated here; there are possibly other factors at play that impact PP ordering preferences in child and parent speech. Earlier work (largely on English) on the double PP construction has also considered the traditional ordering rule for postverbal adverbials more broadly, which is Manner before Place before Time (MTP) (Quirk & Leech, 1985); later empirical studies (Liu, 2019; Wiechmann & Lohmann, 2013) have confirmed this rule to have a pronounced effect in PP orderings. With proper manual annotations, future studies can explore MTP as well as how it interacts with the other factors of interest examined here for child and parent production. It is possible that the inclusion of MTP will lead to an optimal logistic model with more explanatory power of the PP ordering preferences.

Due to the relatively smaller number of child utterances containing the double PP construction in our study, we were unable to compare ordering preferences between child and parent speech within individual age spans. Even the three factors of interest do not seem to impact ordering preferences differently in children vs. parents as described above – yet this does not strictly address the question: at what developmental stage do children's syntactic ordering preferences start to become more parent-like?

As a preliminary attempt (of which the results are suggestive rather than conclusive because of data sparsity), we took the optimal logistic regression model fitted to parent speech alone and applied it to child data (Gries & Adelman, 2014)). Our logic is if at a certain developmental stage, children's production starts to resemble that by parents, the model *tailored* towards parent speech should be able to predict reasonably when it comes to data produced by children during that developmental stage. To that end, we measured the model's prediction accuracy for child speech within each age bin. A higher accuracy indicates that child production during this age range is comparatively more parent-like. The results showed the accuracy for children under 24 months of age was 0.48 (out of 1), and 0.56 between 24 and 30 months; afterwards, the accuracy scores increase to a relatively stable value of 0.62. These results suggest that the age of 30 months could be a turning point along children's developmental trajectory for their production of the double PP construction. With the CHILDES database constantly growing, future work based on more data from more children could confirm this hypothesis.

References

Attardi, G., Sartiano, D., & Yu, Z. (2021). *Diaparser attentive dependency parser.*

Barnes, S., Gutfreund, M., Satterly, D., & Wells, G. (1983). Characteristics of adult speech which predict children's language development. *Journal of Child Language*, *10*(1), 65–84. doi: 10.1017/S0305000900005146

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019, 06). Consistency and Variability in Children's Word Learning Across Languages. *Open Mind*, *3*, 52-67. Retrieved from `https://doi.org/10.1162/opmi_a_00026` doi: 10.1162/opmi$_{a0}$0026

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69–94). KNAW.

De Marneffe, M.-C., Grimm, S., Arnon, I., Kirby, S., & Bresnan, J. (2012). A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes*, *27*(1), 25–61.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Engel, A., Grafmiller, J., Rosseel, L., & Szmrecsanyi, B. (2022). Assessing the complexity of lectal competence: the register-specificity of the dative alternation after give. *Cognitive Linguistics*.

Erika, H.-G. (1997). *Language development*. Pacific Grove, CA: Brooks/Cole.

Ferrer-i Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70*(5), 056135.

Futrell, R., Levy, R. P., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language*, *96*(2), 371–412.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1 - 76.

Gibson, E., Futrell, R., Piantadosi, S., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407.

Gildea, D., & Temperley, D. (2010). Do Grammars Minimize Dependency Length? *Cognitive Science*, *34*(2), 286-310.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 517–520).

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*(3), 289-316.

Greenbaum, S. (1988). A proposal for an international computerized corpus of english. *World Englishes*, *7*(3), 315–315.

Gries, S. T. (2017). Syntactic alternation research: Taking stock and some suggestions for the future. *Belgian Journal of Linguistics*, *31*(1), 8–29.

Gries, S. T., & Adelman, A. S. (2014). Subject realization in japanese conversation by native and non-native speakers: Exemplifying a new paradigm for learner corpus research. In *Yearbook of Corpus Linguistics and Pragmatics 2014* (pp. 35–54). Springer.

Gulordava, K., & Merlo, P. (2015). Structural and lexical factors in adjective placement in complex noun phrases across Romance languages. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning* (pp. 247–257).

Gustafsson, M. (1976). The frequency and 'frozenness' of some English binomials. *Neuphilologische Mitteilungen*, *77*, 623–637.

Hawkins, J. A. (1990). A Parsing Theory of Word Order Universals. *Linguistic Inquiry*, *21*(2), 223–261.

Hawkins, J. A. (2014). *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press.

Heller, B., Bernaisch, T., & Gries, S. T. (2017). Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes. *ICAME Journal*, *41*(1), 111–144.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied psycholinguistics*, *19*(4), 603–629.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology*, *45*(3), 337-374. doi: https://doi.org/10.1016/S0010-0285(02)00500-5

Kramer, A. (2021, February). Dependency lengths in speech and writing: A cross-linguistic comparison via YouDePP, a pipeline for scraping and parsing YouTube captions. In *Proceedings of the Society for Computation in Linguistics 2021* (pp. 359–365). Online: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2021.scil-1.36`

Levy, R., & Jaeger, T. F. (2007). Speakers Optimize Information Density Through Syntactic Reduction. In *Proceedings of the 20th Conference on Neural Information Processing Systems (NIPS).*

Liu, Z. (2019, August). A comparative corpus analysis of PP ordering in English and Chinese. In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)* (pp. 33–45). Paris, France: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/W19-7905` doi: 10.18653/v1/W19-7905

Liu, Z. (2020). Mixed evidence for crosslinguistic dependency length minimization. *STUF - Language*

Typology and Universals, *73*(4), 605 - 633. doi: https://doi.org/10.1515/stuf-2020-1020

Liu, Z. (2022). A multifactorial approach to crosslinguistic constituent orderings. *Linguistics Vanguard*.

Liu, Z., & Prud'hommeaux, E. (2022). Data-driven parsing evaluation for child-parent interactions. *arXiv preprint arXiv:2209.13778*.

Liu, Z., Upreti, R., Kramer, M. A., & Namboodiripad, S. (2022). Does one size fit all in crosslinguistic dependency length minimization? In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Liu, Z., & Wulff, S. (2023, March). The development of dependency length minimization in early child language: A case study of the dative alternation. In *Proceedings of the seventh international conference on dependency linguistics (depling, gurt/syntaxfest 2023)* (pp. 1–8). Washington, D.C.: Association for Computational Linguistics. Retrieved from `https://aclanthology.org/2023.depling-1.1`

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (Vol. 1). Psychology Press.

Portelance, E., Degen, J., & Frank, M. C. (2020). Predicting Age of Acquisition in Early Word Learning Using Recurrent Neural Networks. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* (Vol. 42).

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020, July). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 101–108). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-demos.14

Quirk, G. S., Randolph, & Leech, G. (1985). *A Comprehensive Grammar of the English Language*. Longman, London.

Rajkumar, R., van Schijndel, M., White, M., & Schuler, W. (2016). Investigating locality effects and surprisal in written English syntactic choice phenomena. *Cognition*, *155*, 204–232.

Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2019). childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior research methods*, *51*(4), 1928–1941.

Sethuraman, N. (2004). Influence of parental input on learning argument structure constructions. *On-line proceedings supplement of Boston University child language development 28*.

Szmrecsanyi, B., Grafmiller, J., Bresnan, J., Rosenbach, A., Tagliamonte, S., & Todd, S. (2017). Spoken syntax in a comparative perspective: The dative and genitive alternation in varieties of English. *Glossa: a journal of general linguistics*, *2*(1), 1–27.

van den Bosch, A., & Bresnan, J. (2015, September). Modeling dative alternations of individual children. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning* (pp. 103–112). Lisbon, Portugal: Association for Computational Linguistics. doi: 10.18653/v1/W15-2414

Wasow, T. (1997). End-Weight from the Speaker's Perspective. *Journal of Psycholinguistic Research*(3), 347–361.

Wiechmann, D., & Lohmann, A. (2013). Domain minimization and beyond: Modeling prepositional phrase ordering. *Language Variation and Change*, *25*(1), 65–88.

Wulff, S. (2003). A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics*, *8*(2), 245–282.