ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

The Complex Role of Utterance Length on Grammaticality: Multivariate Multilevel Analysis of English and Spanish Utterances of First-Grade English Learners

Anny Castilla-Earls,[a] David J. Francis,[b] and Aquiles Iglesias[c]

[a] Department of Communication Disorders and Sciences, University of Houston, TX [b] Texas Institute of Measurement, Evaluation, and Statistics, University of Houston [c] University of Delaware, Newark

ARTICLE INFO

Article History:
Received August 7, 2020
Revision received November 16, 2020
Accepted August 26, 2021

Editor-in-Chief: Stephen M. Camarata
Editor: Mary Alt

https://doi.org/10.1044/2021_JSLHR-20-00464

ABSTRACT

Purpose: This study examined the relationship between utterance length, syntactic complexity, and the probability of making an error at the utterance level.
Method: The participants in this study included 830 Spanish-speaking first graders who were learning English at school. Story retells in both Spanish and English were collected from all children. Generalized mixed linear models were used to examine within-child and between-children effects of utterance length and subordination on the probability of making an error at the utterance level.
Results: The relationship between utterance length and grammaticality was found to differ by error type (omission vs. commission), language (Spanish vs. English), and level of analysis (within-child vs. between-children). For errors of commission, the probability of making an error increased as a child produced utterances that were longer relative to their average utterance length (within-child effect). Contrastively, for errors of omission, the probability of making an error decreased when a child produced utterances that were longer relative to their average utterance length (within-child effect). In English, a child who produced utterances that were, on average, longer than the average utterance length for all children produced more errors of commission and fewer errors of omission (between-children effect). This between-children effect was similar in Spanish for errors of commission but nonsignificant for errors of omission. For both error types, the within-child effects of utterance length were moderated by the use of subordination.
Conclusion: The relationship between utterance length and grammaticality is complex and varies by error type, language, and whether the frame of reference is the child's own language (within-child effect) or the language of other children (between-children effect).
Supplemental Material: https://doi.org/10.23641/asha.17035916

The purpose of this study is to enhance our understanding of the relationship between the length and complexity of an utterance and the probability of making an error at the utterance level in Spanish and English. Developmental measures of utterance length, such as mean length of utterances (MLUs), and broad indices of grammaticality, such as mean percentage of grammatical utterances (PGU) without errors, are often used in the literature to describe typical and atypical language development in Spanish (e.g., Restrepo, 1998; Simon-Cereijido & Gutiérrez-Clellen, 2007) and English (e.g., Guo et al., 2019; Scott & Windsor, 2000). However, the relationship between measures of utterance length and grammatical errors is not well understood, particularly in bilingual children. Developmentally, utterances are initially short with many omissions of both words and

Correspondence to Anny Castilla-Earls: annycastilla@uh.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

grammatical structures as vocabulary and grammar are still developing. As children further develop their language skills, they increase the length of their utterances and make fewer errors of omission and more errors of commission (i.e., substitutions and additions). Over time, children produce utterances that continue to increase in length, are more syntactically complex, and contain fewer errors overall (Brown, 1973). Yet, Simon-Cereijido and Gutiérrez-Clellen (2007) found that some Spanish-speaking children who produce very short utterances made no errors. This finding could be a result of either the opportunities for errors (number of words in an utterance) or a function of utterance complexity. Importantly, not all errors are the same. Errors of omission and errors of commission signal different stages of child language development and may have different associations with utterance length (Brown, 1973). In addition, the relationship between utterance length, errors, and complexity might be mediated by the learning status of a language. In Spanish-speaking English learners (ELs; i.e., children who learn Spanish at home and English at school), different relationships might be observed for Spanish, the first language (L1), and English, a language in early stages of development.

To fully understand the relationship between utterance length, complexity, and errors, it is important to examine how this association works at the utterance level. Most of the current studies that offer some linkage between MLU, sentence complexity, and errors are conducted at the child level (e.g., Baron et al., 2018; Castilla-Earls & Eriks-Brophy, 2012; Rice et al., 2006; Simon-Cereijido & Gutiérrez-Clellen, 2007), which means that relationships are driven by individual differences between children. Using between-children effects to describe utterance-level effects exemplifies what researchers have termed an *ecological fallacy*: an incorrect inference that a relationship at one level of analysis applies to another level of analysis that was not examined (Francis et al., 2018; Robinson, 1950). In this study, we examine utterance length, complexity, and errors in a large sample of Spanish-speaking ELs. ELs present with a unique scenario to investigate the relationship between utterance length, complexity, and errors since we can examine this relationship in both the L1 (Spanish) and the emerging language (English).

Trade-Off Effects in Sentence Formulation

Bock (1982) proposed a widely used theoretical framework of sentence formulation. Within this sentence formulation framework, basic components such as lexical identification, morphological specification, structural scaffolding, and structural assembly are combined to produce a message in a sentence. Of particular interest to this study are the interactions between these components that work under a limited capacity working memory system. These interactions predict trade-off effects, such that increases in complexity in one component might result in decreases in accuracy in another component (Bock, 1982). This theoretical framework applies to a mature language system; therefore, developmental errors are not well represented within this theory. However, this framework can give us information about how processing demands might affect the production of errors in child language utterances.

A handful of studies have examined the potential trade-offs proposed by Bock in child language. In English-speaking children, Bloom (1990) investigated the relationship between subject omission and verb phrase length in the three children studied by Brown (1973). Bloom found that utterances with subject omission errors had longer verb phrases than utterances with subjects. Masterson and Kamhi (1992) studied the potential trade-off effects between grammatical accuracy, phonology, and syntax in a small group of children with and without language disorders. They found that most phonologically complex utterances were syntactically simple, while utterances that had simpler phonological complexity included both simple and complex utterances, which suggests a trade-off between syntactic and phonological complexity. Grela (2003) and Grela and Leonard (2000) examined the influence of language complexity on the omission of verbs in two groups of English-speaking children with typical language skills: 3-year-olds and 5-year-olds. Their results suggested that 3-year-olds made more verb omission errors and subject errors when they produced utterances with greater complexity in argument structure in comparison to 5-year-olds. Lastly, Owen (2010) examined the accuracy of past tense in compound and complex utterances in a group of English-speaking children using a sentence completion task. Her results showed that all children made similar errors in compound clauses but more past tense errors in the second clause of complex clauses, suggesting an effect of complexity on verb accuracy since complex clauses were more likely to be produced with errors.

There is also evidence in support of the sentence formulation model proposed by Bock (1982) from research conducted in Spanish-speaking children with language disorders. Simon-Cereijido and Gutiérrez-Clellen (2007) examined the utility of spontaneous language measures to identify Spanish-speaking children with language disorders. They found that children who are often misidentified as not having a language disorder tend to have short and simple utterances with few errors. Simon-Cereijido and Gutierrez-Clellen argued that the lack of errors was due to the children relying on simple syntax when complex language was required, supporting the limited capacity aspect of sentence formulation (Bock, 1982).

The previous studies suggest that there is potentially a trade-off between the length and complexity of an utterance and the probability of making errors. The general finding is that long utterances and/or complex utterances are more likely to present with errors. However, there are many unresolved questions. First, does the type of error have an effect on this relationship? It is possible that omission errors are likely to occur in shorter sentences and commission errors in longer sentences (Brown, 1973). Second, how does the relationship between utterance length and complexity predict errors? Although longer utterances tend to be more complex, the relationship between length and complexity may not be linear.[1] Last, is the relationship between length, complexity, and errors different by level of language development? In the case of ELs, different patterns of errors might be seen in their two languages, as the level of language development of an individual child may differ across the two languages with considerable individual differences in language skills across children in each language. It is important to disentangle the relationship between errors, utterance length, and complexity to enhance our understanding of potential sentence production trade-offs in child language.

Between-Children and Within-Child Effects

Studies on child language development have mostly focused on describing differences between individual children (i.e., interindividual differences), and between groups of children. For example, we want to know how development affects children's performance on a task, and to answer this question, we could compare the task performance of children who differ from one another in age. Such a design relies on between-children differences to guide inferences about how performance will change as children develop. However, differences between children do not tell us the complete story about the development of children's language abilities and can be misleading when we attempt to infer how factors will affect the development of individual children based on relationships that have been established between children. Intra-individual differences are the hallmark of development and do not always follow from interindividual differences (Nesselroade, 1991).

Within-child differences are important for understanding the language functioning of individual children. Within-child effects (i.e., intra-individual differences) represent the variability and covariability among measures for an individual child. Common approaches to studying within-child effects involve the multilevel analysis of longitudinal data. These analyses allow the investigation of intra-individual change over time while also investigating interindividual differences in intra-individual change (Francis et al., 2018). In such instances, the emphasis is simultaneously on how each individual's behavior changes over time and on factors that explain why development differs from one individual to the next. Questions such as how Spanish and English vocabulary codevelop in bilingual children are questions that provide different answers when examined within child versus between children (Branum-Martin et al., 2009).

Importantly, within-child and between-children effects can be studied simultaneously using multilevel modeling (MLM). MLM allows the separation of total covariation between measures into covariation that resides within child and covariation that resides between children with potentially different predictors of each type of relationship. In this study, we examine within-child and between-children influences on the probability of making an error using sentence length and subordination at the utterance level. Within-child effects allow us to examine the potential trade-offs of utterance length and complexity on the probability of an utterance to include an error for a child. It is then possible to estimate if the probability of an utterance to include an error increases or decreases when a child produces utterances that are longer or shorter than their average-length utterance. Between-children effects allow us to estimate whether children who produce utterances that are, on average, longer or shorter than the average utterance for all children tend to produce more or fewer errors than children with average-length utterances.

This Study

In this study, we examine the relationship between errors and the length and complexity of utterances using multivariate MLM with utterance as the unit of analysis nested within language and person. The study is innovative in that we describe (a) differences across utterances that are produced by the same child and differences between utterances produced by different children; (b) errors of omission and errors of commission separately because relationships between utterance length, complexity, and error probability may vary as a function of error type; and (c) differences between utterances in Spanish and utterances in English in a group of young ELs to investigate the potential role of language development in the relationship between utterance length, complexity, and errors.

The research questions guiding this investigation were as follows: (1a) What is the effect of utterance length on the probability of making an error of omission in Spanish and English? (1b) How do these relationships compare within child and between children? (1c) What is the effect of subordination on the probability of making an omission error? (2a) What is the effect of utterance length on the probability of making an error of commission in Spanish and English? (2b) How do these relationships compare within and between children? (2c) What is the effect of subordination on the probability of making a commission error?

[1]Utterances can be long and simple (i.e., The boy said hello to the frog, the big dog, and the turtle) but also short and complex (i.e., The boy saw what he did).

Following the sentence formulation framework proposed by Bock (1982), we predict a trade-off between utterance length, complexity, and errors. As utterances increase in length and complexity, more errors are predicted to occur. However, from a developmental standpoint, we expect differences between the type of errors and between Spanish and English samples. We hypothesize that increases in utterance length and complexity will result in (a) a decrease of errors of omission and an increase of errors of commission at the child level; (b) larger trade-off effects for English than Spanish considering that English is the second language for these children, which might impose higher processing demands; and (c) trade-offs for within-child effects due to increased processing demands, but not for between-children because children who produce longer utterances, on average, are believed to have more developed language skills and, therefore, fewer errors would be expected.

Method

Participants

Children in this study are a subset of the children in Castilla-Earls et al. (2019). The participants were 830 children who were attending first grade (age 6;7 [years; months]; *SD* = 5 months) in one of two states in the southcentral and western United States. At the time of recruitment, which occurred in kindergarten, these children had not been identified as having a disability by their school districts. However, it is reasonable to expect that at least some of these children would later be identified as having a language disability. Because we did not systematically obtain subsequent information on ability status from the school district, it is possible that ability status may have subsequently changed. Thus, it is expected that this sample might have included some children with language disorders. There were 422 girls and 408 boys. These children were considered to be ELs by their schools (e.g., the English skills of these children were considered to be insufficient to perform adequately in an English classroom[2]) and were enrolled in either bilingual English–Spanish instruction (*n* = 542) or English-only instruction with language learning support (*n* = 285). Bilingual instruction programs included transitional bilingual, dual language, and maintenance programs, with the majority of the children enrolled in transitional bilingual programs. These bilingual instruction programs are very similar in early grades with more differentiation occurring after third grade. The children were considered to be from a low socioeconomic background as indexed by average mother education of 9 years and 79% of children with reported family income below $30,000 per year (Branum-Martin et al., 2014). Children were selected for this study because they produced a story retell in both English and Spanish during the fall of first grade. Students who were repeating first grade were excluded from this study. Only children who produced stories of at least 10 utterances in each language were included in the study to ensure that children had enough productive skills in both Spanish and English to measure the relationship between utterance length, errors, and complexity. The study participants produced a total of 1,660 story retells (see Table 1 for descriptive statistics for the story retells). The utterances produced in these narratives served as the unit of analysis in this study. The secondary analysis presented in this study received exempt institutional review board approval at the University of Houston in 2016.

Measures

All measures in this study were calculated at the utterance level and analyzed separately for Spanish and English using the Systematic Analysis of Language Transcripts (SALT) software (Miller & Iglesias, 2019).

Utterance Length

MLU is perhaps one of the most well-known measures of child language development. In English, MLU is traditionally calculated as the total number of morphemes divided by the total number of utterances (Brown, 1973). In Spanish, MLU is conventionally calculated in words (Gutiérrez-Clellen et al., 2000). As children grow older, changes in MLU are observed in Spanish and English (e.g., Castilla-Earls & Eriks-Brophy, 2012; Gutiérrez-Clellen et al., 2000; Rice et al., 2006; Rojas & Iglesias, 2013). Utterance length in words (ULw) was used in this study to account for utterance length.

Sentence Complexity

Subordination index (SIu) is a measure of clausal density that represents sentence complexity. Utterances with more than one clause are considered to be more complex than utterances with a single clause. There is evidence that complex syntax, as measured by SIu, continues to develop through the school years for English- and Spanish-

[2]Children in this study were attending schools in Texas and California. These states have specific EL identification procedures that have changed over time. At the time of data collection, both states administered a home language survey at the time of first school enrollment. This home language survey would have identified the student as a language minority student if a language other than English was spoken at home. If the student was identified as a language minority at school entry, they were administered a language proficiency assessment. The specific assessment used varied by state, but both states used an assessment and employed a specific decision-making process for determining that a student qualified for support as an EL.

Table 1. Descriptive information for story retells.

Variable	Spanish				English			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Number of utterances	38.5	13.6	10	112	36.1	12.7	10	90
Length of story in minutes	4.3	1.6	1.52	15.2	4.3	1.6	1.4	13.1
Mean length of utterance in words	5.5	0.8	2.45	8.7	6.4	1.0	1.9	10.6
Subordination index	1.1	0.1	1	1.6	1.0	0.1	0	1.6
Number of different words	80.7	23.8	23	184	73.3	26.3	13	164
Percentage of utterances with errors	.10	.10	0	.74	.44	.15	0	.93

Note. Min = minimum; Max = maximum.

speaking children (Castilla-Earls & Ericks-Brophy, 2012; Gutierrez-Clellen & Hofstetter, 1994; Scott, 1988). SIu at the utterance level (i.e., number of clauses in the utterance) was used in this study to account for sentence complexity.

Errors

The percentage of grammatical utterances in a language sample is a broad measure used in child language to index grammatical development in both Spanish and English. In both English-speaking and Spanish-speaking children, the percentage of grammatical utterances increases with age (e.g., Castilla-Earls & Eriks-Brophy, 2012; Guo et al., 2019). The number of errors of omission and the number of errors of commission at the utterance level were used in this study to examine the probability of a child producing an utterance with error.

Procedure

The general procedures for collecting the story retells were previously described in Castilla-Earls et al. (2019) and Miller et al. (2006). Children were asked to retell one of the frog stories from the Mercer Mayer (1967) wordless picture book series after having previously heard a script of the story. All audio recordings of the stories were transcribed by hand and then analyzed using SALT. The language transcription of all samples was completed by trained research assistants as part of the initial parent study using a strict protocol that is described in Miller et al. (2006). These stories were segmented into modified C-units and coded using SALT standard procedures for language sample coding for errors and SIu. Errors of omission at the word and bound morpheme level were coded with an asterisk. Incorrect past forms of irregular verbs were marked as errors of omission. Errors of commission in this study included overregularization errors [EO:_] and errors at the word level [EW:_]. See Supplemental Material S1 for examples of errors of omission and commission in Spanish and English.

Interrater reliability was calculated for 10% of the samples in this study. All samples were coded for errors of omission and commission at the utterance level by a trained research assistant, and the results were compared to the original coding in the parent study. For Spanish, interrater reliability was 95% for errors of omission and 91% for errors of commission. For English, interrater reliability was 93% for errors of omission and 91% for errors of commission. All disagreements were resolved using a third trained research assistant. We considered the error coding procedures to be reliable.

SALT software does not currently have capabilities to run analyses at the utterance level in large batches. Therefore, we created multiple databases at the utterance level using the transcript cut option. To do so, we set the transcript cut to start at Utterance 1 and end at Utterance 1. Then, we created a rectangular data file and extracted ULw, number of different words in the utterance (NDWu) SIu, and error information for all utterances. Utterances with code-switching (1,578 utterances in English and 1,016 in Spanish, approximately 5% and 3% of all utterances, respectively) were excluded for all analyses. On average, children produced 37 utterances in Spanish (*SD* = 11) and 34 utterances in English (*SD* = 11); therefore, we analyzed all data available until Utterance 60 to ensure the inclusion of the most representative data (7% of stories in Spanish and 4% stories in English had more than 60 utterances). This procedure resulted in 60 CSV files (rectangular databases), each representing an utterance number in the stories for all children in the study. Each utterance in each language was included only once, and only utterances with data were included (i.e., utterance numbers 39–60 for a child who produced 38 utterances were not included). All 60 CSV files with information at the utterance level were then concatenated to create the database for this study.

This procedure resulted in a database of utterances with information about the child (i.e., age and code), the language used to retell the story, ULw, NDWu, SIu, the number of errors of commission, and the number of errors of omission. One hundred and sixty-six utterances in Spanish and 105 utterances in English were excluded from the database because they represented group outliers in the distribution of utterance length in each language visually (17 or more words in English and 16 or more words in Spanish). Error coding was dummy coded for all utterances as error = 1 (1 or more errors of the same type in the utterance) or error = 0 (no errors in the utterance) to make it an analyzable categorical variable for logistic regression. The dummy code was unique for the type of error (omission vs. commission). The database of utterances included 30,550 utterances in Spanish and 27,497 utterances in English. There were 2,304 utterances in Spanish with commission errors and 675 with omission errors. In English, there were 7,572 with commission errors and 8,843 utterances with omission errors. A number of utterances (*n* = 779; 111 in Spanish and 1,942 in English) included both errors of commission and errors of omission, and these utterances were eliminated from the study.[3]

Analysis

Mixed-effect logistic regression was used in this study to analyze the impact of utterance length and subordination on the probability of making an error at the utterance level. All models were estimated with the meqrlogit command in Stata Version 15 (StataCorp, 2017). The outcome measure in a given analysis was a binary variable signaling the presence of an error (either of omission or commission, depending on the analysis) in an utterance (0 = no error, 1 = error). The utterance-level data were structured in a hyperunivariate form in which each utterance represented a single record in the data file with two binary outcome variables (error of omission; error of commission). This utterance-level record was linked to the child, the language used to retell the story, other descriptive data about the child, and the story retell in that language. Thus, the utterance-level records carried information that varied at the utterance level, the story/language level, and the child level.

Errors of omissions and errors of commission were analyzed separately using the following steps. The first step in the analysis was the estimation of a model that partitioned the variance of the probability of making an error in both Spanish and English. In this base model, the variance was allowed to vary as a function of the language of the story (Spanish or English). Also, a covariance between the variance in Spanish and English was included to capture the relationship between the two languages. As a result, this model included five parameters (two fixed and three random): a fixed Spanish intercept, a fixed English intercept, the variance in intercepts for Spanish, the variance in intercepts for English, and the covariance between Spanish and English intercepts. The fixed intercept in a language estimates the average log-odds of making an error of that type in that language, while the variance of the intercept in a given language estimates how the log-odds of making an error in that language differs across children.

The equation of Model 1 is $\rho_{ui_l} = \gamma_{00_l} + \cup_{0i_l} + \in_{ui_l}$, where *p* is the log-odds of the probability of making an error in utterance *u* in language *l* for child *i*. We use the subscript *l* as an index to capture the language of the narrative (Spanish or English). In this equation, the random variables define a language-specific random intercept $\cup_{0i_l}$ for person *i* in language *l,* which is allowed to vary randomly across children, and a random error residual, $\in_{ui_l}$, that captures the extent to which the log-odds for utterance *u* for child *i* in language *l* deviates from the expected log-odds for an utterance in language l, γ_{00_l}, plus the person-specific random effect, $\cup_{0i_l}$.

The goal of all subsequent models was to explain the variability across children and utterances in the probability of making an error at the utterance level. We first ran a model with chronological age, although we did not expect this variable to be significant because we had restricted the age in the sample by selecting children in the fall of first grade. Age was indeed nonsignificant for either errors of omission or errors of commission and did not improve model fit as indexed by Bayesian information criterion (BIC). Therefore, age was not included in any of the subsequent models.

To explain the potential within-child effects and between-children effects of utterance length on the probability of making an error in an utterance, we included utterance length as a predictor in Model 2. We employed person-level centering of utterance length to isolate the effects of utterance length within child and separate them from the between-children effects of utterance length. Thus, we included (a) ULw centered at the child mean by the language of the utterance and (b) child mean ULw in that language centered at the sample mean by the language of the utterance and language of instruction for the child. We explicitly chose to center child mean utterance length within language and language of instruction (i.e., bilingual instruction or English-only instruction) to remove differences in mean utterance length due to the language of instruction and its interaction with the language of the narrative. By using a language-and-instruction-specific mean for centering, we eliminate mean differences across languages due to the language of instruction. Given the utterance-level centering of utterance length at the child mean and centering of the child-level mean utterance length at the grand mean utterance length for that language and language of instruction, the intercepts in Model

[3]We conducted stability analyses (Rosenbaum & Silber, 2009) to assess the impact of this decision and found that the exclusion of utterances with both types of errors did not materially affect inferences from our statistical analyses (i.e., analysis of errors of omission for all utterances yielded comparable inferences to analysis of errors of omission when utterances with both error types were excluded). Joint analysis of the two error types proved difficult for estimating the covariance between error types at the utterance level, which joint analysis would require to address nonindependence between-errors probabilities at the utterance level.

2 represent the average probability of an error occurring in an utterance of average length for the child when spoken by a child whose mean utterance length is average for the sample in that language for children in that language of instruction. This model separates the effects of utterance length on the log-odds of making an error into effects within child and effects between children. The coefficient associated with the within-child-centered utterance length variable estimates the difference in log-odds of making an error when the child makes an utterance in that language that is longer or shorter than their typical utterance in that language. The coefficient associated with the between-children predictor estimates the difference in the log-odds of an error for children who, on average, make longer or shorter utterances in that language than other children in the same language of instruction. Thus, the former coefficient reflects intra-individual differences associated with the child making utterances that are longer or shorter than average for them in a given language, whereas the latter coefficient reflects interindividual differences associated with children having a mean utterance length that is longer or shorter than the sample average utterance length in a given language and language of instruction. The equation for Model 2 is $\rho_{ui_t} = \gamma_{00_t} + \gamma_{10} * ULwc + \gamma_{20} * ULwg + \cup_{0i_t} + \in_{ui_t}$.

We attempted to include a random effect for within-child utterance length in the models but were unsuccessful in estimating this model with additional random effects. Examination of simpler deconstructed models for the type of errors by language (i.e., commission errors in Spanish) revealed very small estimates for random effects of utterance length. We attributed the difficulty in estimating a random effect for utterance length at the child level to the fact that children might not have enough utterances at different values of utterance length for these slopes to be estimated with sufficient precision to mitigate shrinkage back to the mean within-child relationship. It is important to keep in mind that this slope estimates how the log-odds of making an error varies as a function of utterance length for a given child. We proceeded without a random effect for utterance length at the child level.

In Model 3, the effect of SIu and the interaction between utterance length and SIu were included. SIu was centered at 1 in this analysis. Therefore, the intercepts in Model 3 represent the average probability of making an error for an utterance with an SIu of 1 that is of average length for the child, for a child whose mean utterance length is average for the sample in that language for children in the same language of instruction as the child. Utterances with a subordination score equal to or higher than 2 were collapsed into a single group to represent subordination. The coefficient associated with subordination thus represents the difference in log-odds between utterances with no subordination and utterances with subordination. The equation for Model 3 is $\rho_{ui_t} = \gamma_{00_t} + \gamma_{10} * ULwc + \gamma_{20} * ULwg + \gamma_{30} * SIu + \gamma_{40} * ULwc + \cup_{0u_t} + \in_{ui_t}$.

To make the results of the models more interpretable, the results are presented in the text by converting log-odds into probabilities. Probability is a number between 0 and 1 that describes how likely an utterance is to have an error. Numbers closer to 0 denote very low probability, and numbers closer to 1 signify certainty. The formula to convert log-odds to probabilities is p = exp(log odds)/[1 + (exp(log-odds))]. Probabilities can also be interpreted as the PGU that have an error. For example, a probability of 1 indicates that 100% of the utterances have an error while a probability of 0 indicates that none of the utterances have errors.

Results

Errors of Omission

The results of Model 1 showed that, on average, the probability of making an error of omission is .013 in Spanish and .264 in English in this base model with no predictors (descriptive results at the utterance level are presented in Table 2; fixed and random effects are presented in Table 3). Therefore, only about 1.3% of all utterances in Spanish included errors of omission, while about 26% of the utterances in English included an omission error.

Table 2. Descriptive statistic for utterances.

	Spanish n = 30,550				English n = 27,497			
Variable	*M*	*SD*	Min	Max	*M*	*SD*	Min	Max
Utterance length	5.6	2.70	1	16	6.5	2.60	1	17
Subordination index	1.2	0.38	1	4	1.1	0.33	0	3
Number of different words	5.3	2.41	1	16	6.1	2.14	1	17
Errors of omission	.02	.14	0	1	.32	.47	0	1
Errors of commission	.07	.26	0	1	.27	.45	0	1

Note. Min = minimum; Max = maximum.

Table 3. Fixed and random effects for errors of omission.

Variable	Model 1			Model 2			Model 3		
	Estim.	*SE*	Sig	Estim.	*SE*	Sig	Estim.	*SE*	Sig
Fixed effects									
Spanish	−4.334	.070	***	−4.383	.072	***	−4.410	.075	***
English	−1.022	.036	***	−1.088	.034	***	−1.192	.035	***
ULwc Spanish				0.108	.015	***	0.113	.020	***
ULwc English				−0.083	.007	***	−0.099	.007	***
ULwg Spanish				−0.089	.071		0.097	.072	
ULwg English				−0.465	.037	***	−0.479	.038	***
SIu Spanish							0.371	.161	***
SIu English							0.331	.081	***
SIu × ULwc Spanish							−0.075	.042	
SIu × ULwc English							0.077	.023	***
Variance components									
Spanish	0.747	.113		0.753	.114		0.755	.111	
English	0.886	.059		0.712	.049		0.705	.048	
Cov Spanish–English	−0.041	.059		−0.029	.053		−0.043	.053	
BIC		33,289			32,836			31,873	

Note. Estim. = estimate; *SE* = standard error; Sig = significance; ULwc = utterance length in words child centered; ULwg = utterance length in words grand mean centered; SIu = subordination index at the utterance level; COv = covariance; Bayesian information criterion = BIC.

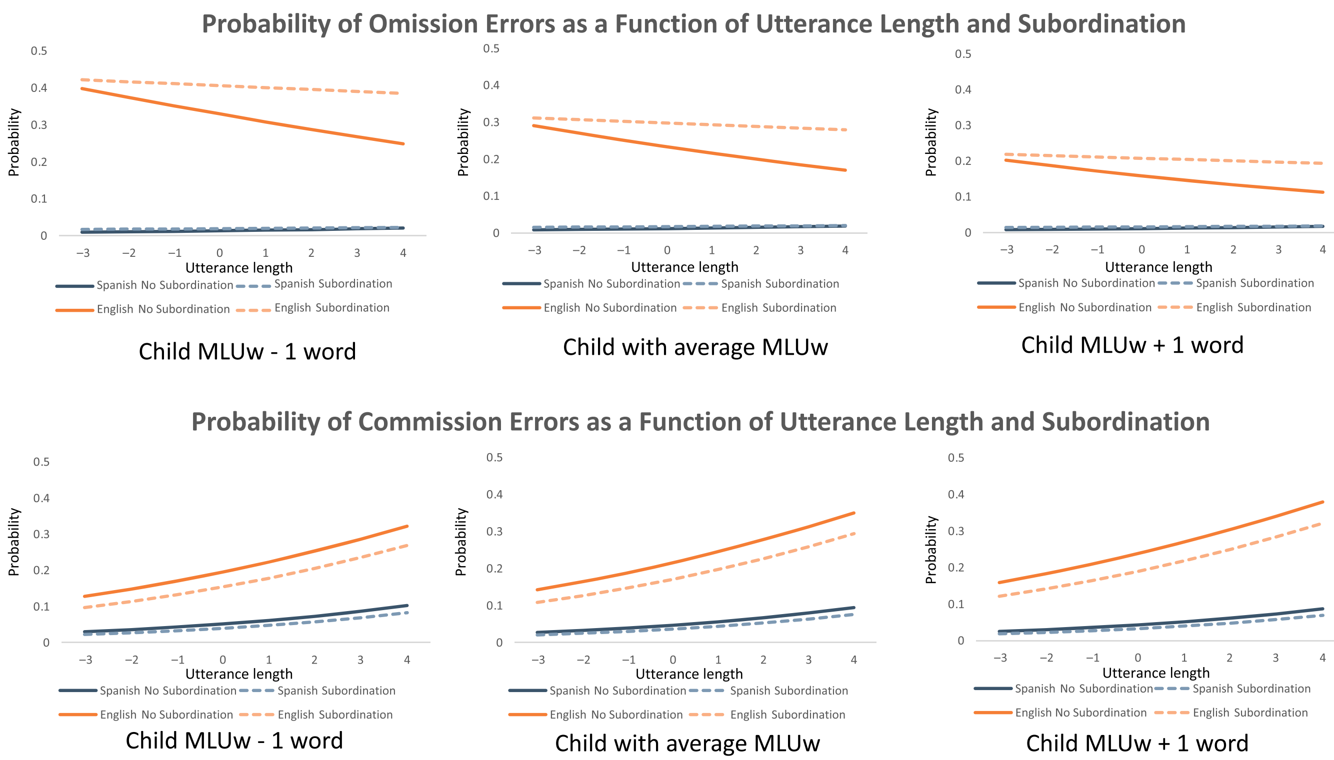
***Significance below .05.

Model 2 examined within-child and between-children effects of utterance length and showed significant effects at both levels in both Spanish and English. The within-child and between-children coefficients show similar effects of increasing the length of an utterance by one word in comparison to the child's average and the sample average for both Spanish and English. In Spanish, the probability of making an omission error decreases from .012 to .006 when a child produces an utterance that is one word longer than their average utterance. However, children whose average utterance length is one word longer than the average utterance length for all children have the same probability of making an error of omission since the difference was not statistically significant for Spanish ($\gamma = -.089$, $SE = .071$). In English, the probability of producing an error of omission also declines when utterance length increases, either within child or between children. That is, within child, an increase of one word in utterance length reduces the probability of an error from .251 to .236. Similarly, children whose mean utterance length is one word longer than the sample average are less likely to make errors of omission, with the probability decreasing from .251 to .174. Therefore, within-child effects of increasing utterance length by one word are similar in Spanish and English. It is important to note that omission errors in Spanish at this age are rare and the differences in error probabilities as a function of utterance length were very small.

In Model 3, the intercept is interpreted as the probability of making an error in an utterance with an SIu of 1 that is of average length for a child whose average utterance length is at the sample average for that language and language of instruction. Therefore, the probability that a child whose mean utterance length is at the sample average makes an error in an utterance that is of average length for the child is .233 in English and .012 in Spanish for utterances without subordination. The main effect of subordination was significant for utterances in English and Spanish as a predictor of omission errors. In English, subordination increased the probability of making an error from .233 to .297. At the same time, there was evidence that subordination interacted with utterance length at the child level, further increasing the probability of error to .452 when the child produced an utterance that was one word longer than their average and that was subordinated. In Spanish, producing an utterance with subordination was associated with an increased probability of making an omission error (.012 for unsubordinated utterances vs. .017 for subordinated utterances), but the interaction between within-child utterance length and subordination in Spanish was nonstatistically significant ($\gamma = -.075$, $SE = .042$).

The between- and within-child effects of Model 3 for omission are shown in Figure 1 on the top panel. The panel is organized into a left-, middle-, and right-hand plot. In all plots of Figure 1, the *x*-axis represents the difference in utterance length in number of words with zero representing the child's average; the *y*-axis shows the probability of an error. The middle plot shows the differential effect of changes in utterance length within child for subordinated and unsubordinated utterances for a child of average mean length of utterance in words (MLUw) for their language of instruction. The left-hand panel graphs the relationship for children whose average utterance length is one word lower than the average for all children in the

Figure 1. Probability of omission errors and commission errors as a function of utterance length. The middle panel shows the differential effect of changes in utterance length and subordination for a child in comparison to their average for a child of average MLUw. In this figure, the *x*-axis represents the difference in utterance length in number of words with zero representing the child's average. The left panel shows the effect of producing an utterance by a child whose average utterance length is one word lower than the average for all children in the same language of instruction. The right panel shows the effect of producing an error by a child whose average utterance length is one word higher than the average for all children in the same language of instruction. MLUw = Mean length of utterance in words.



same language of instruction, whereas the right-hand panel shows the relationship for children whose average utterance length is one word higher than the average for all children in the same language of instruction. The choice of showing the effect of a one-word difference in mean utterance length was driven by the standard deviation in MLUw, which was close to one word in both Spanish and English. Therefore, the left- and right-hand graphs in the top panel can be viewed as being characteristic of children whose average utterance length is approximately 1 *SD* below and 1 *SD* above average for their language of instruction, with the central panel showing the relations for children whose average utterance length is average for their language of instruction.

Model 3 had the lowest value of BIC in comparison to Model 1 and Model 2, indicating an improvement in fit to the data. The unconditional covariance between English and Spanish intercepts from Model 1 was −.041, whereas the conditional covariance from Model 3 was −.043, indicating that the error rates in English and Spanish are largely independent, albeit slightly negatively correlated, but not statistically significantly so. This lack of correlation is not attributable to limited variability in the random intercepts, which was roughly .755 for Spanish in the unconditional model and .705 for English. It appears that errors of omission in one language are not especially predictive of errors in the other, although this covariance might be larger if the effects of language of instruction were uncontrolled.[4]

Errors of Commission

Model 1 shows that, on average, the probability of making an error of commission in Spanish is .049 and the probability of making an error of commission in English is .212 for a base model with no predictors (fixed and random effects are presented in Table 4). In other words, about 5% of all utterances in Spanish and about 21% of all utterances in English included at least one error of commission. In Model 2, we included child mean utterance length centered

[4]Our modeling approach controlled the language of instruction through the approach to centering. Had we not centered within the language of instruction, differences due to the language of instruction would contribute to the random intercepts. However, this effect could also be removed by entering the language of instruction as a fixed effect in the model, in which case we would expect the covariance between intercepts to be similar to estimate in the models presented. We elected for the centering approach to simplify presentation of the models and discussion of model parameters.

Table 4. Fixed and random effects for error of commission.

	Model 1			Model 2			Model 3		
Variable	Estim.	*SE*	Sig	Estim.	*SE*	Sig	Estim.	*SE*	Sig
Fixed effects									
Spanish	−2.957	.050	***	−3.052	.052	***	−3.009	.053	***
English	−1.306	.022	***	−1.332	.022	***	−1.291	.022	***
ULwc Spanish				0.163	.006	***	0.187	.013	***
ULwc English				0.161	.006	***	0.168	.011	***
ULwg Spanish				−0.119	.060	***	−0.092	.060	
ULwg English				0.146	.025	***	0.126	.026	***
SIu Spanish							−0.281	.107	***
SIu English							−0.288	.084	***
SIu × ULwc Spanish							−0.024	.032	
SIu × ULwc English							−0.008	.021	
Variance components									
Spanish	1.202	.096		1.236	.097		1.235	.099	
English	0.163	.018		0.174	.019		0.166	.019	
Cov Spanish–English	−0.029	.028		−0.053	.029		−0.050	.030	
BIC		41,309			40,101			39,873	

Note. Estim. = estimate; *SE* = standard error; Sig = significance; ULwc = utterance length in words child centered; ULwg = utterance length in words grand mean centered; SIu = subordination index at the utterance level; COv = covariance; Bayesian information criterion = BIC.
***Significance below .05.

at the grand mean to estimate between-children effects and within-child utterance length centered at the child-level mean to estimate within-child effects. The results of Model 2 suggest that utterance length is a significant predictor of the probability of making a commission error in both Spanish and English, both within child and between children. In Spanish, the probability of producing an error increases from .045 to .053 when a child produces a Spanish utterance that is one word longer than their average Spanish utterance. At the same time, children whose average utterance length is one word above average for the sample are less likely to make an error. Specifically, the probability of making an error decreases from .045 to .040, if the child has an average utterance length that is one word longer than average for their language of instruction. In English, the probability of making an error is .209 for a child whose average length of utterance is at the mean for all children in their language of instruction when the child makes an utterance of average length for that child. This probability increases to .236 when the same child produces an utterance that is one word longer than their average. Similarly, the probability increases to .234 for children whose MLU is one word longer than average for their language of instruction. That is, children with longer average utterance length have a greater probability of making errors of commission in English, and the probability of making an error of commission goes up as children produce utterances that are longer than their average.

Model 3 included SIu in addition to the previous predictors. We found a significant decrease in the probability of making an error of commission when the utterance included subordination in English ($\gamma = -.288$, $SE = .084$) and Spanish ($\gamma = -.281$, $SE = .107$). For a child whose MLU was average for their language of instruction, the use of subordination was associated with decreased probability of making an error of commission (.215 vs. .171). Similarly, in Spanish, the probability decreased from .047 to .035 when subordination was used. The interaction between subordination and utterance length within child was nonsignificant for both Spanish utterances ($\gamma = -.016$, $SE = .024$) and English utterances ($\gamma = .008$, $SE = .021$). It is important to note that the between-children effect of utterance length for Spanish was nonsignificant in Model 3 ($\gamma = -.092$, $SE = .060$). The bottom panel of Figure 1 shows the within-child and between-children effects of utterance length and subordination. The panels are arranged as before, with the left panel showing expected outcomes for children whose mean utterance length is one word below the sample average and the right panel showing expected outcomes for children whose mean utterance length is one word above the sample average.

Similar to the results for omission, Model 3 showed a significant improvement in data fit as indexed by BIC. The covariance between Spanish and English for errors of commission was −.029 in the unconditional model (Model 1) and −.050 in Model 3. Thus, errors of commission have a small negative correlation, on average, but show no residual correlation when subordination and utterance length are controlled within language.

Discussion

The purpose of this study was to examine the relation between the probability of making an error and the length and complexity of an utterance in Spanish, the participants' L1, and English, the second language for these ELs. We examined these relations at the utterance level for two types of errors: errors of omission and errors of commission. The children in this study were, on average, 6 years of age and produced more errors in English than in Spanish, as expected, given that the children were in the process of acquiring English. The overall rate of errors in Spanish was low with only about 9% of all utterances containing an error, while the rate of errors was higher in English with about 44% of all utterances containing an error. The methodological and statistical approach used in this study allowed us to examine within-child and between-children effects of utterance length (i.e., changes in the probability of making an error as a given child varied the length of their utterances, as well as differences across children in the probability of making an error based on children's average utterance length). In general, our results suggest that there is an effect of both utterance length and complexity on the probability of making an error, but the magnitude and direction of the effect differed within child and between children and by type of error and language. We first discuss the effect of utterance length on errors of omission and commission by language. Second, we address the differences in the directionality of within-child and between-children effects by language. Third, we consider the general role of complexity. Last, we discuss the implication of the results of this study for Bock's Sentence Formulation Model.

Errors of Omission and Utterance Length

Omission errors in English were present in about 32% of the utterances in the sample. In English, the probability of making an omission error was lower for children whose average length of utterance was longer than the average child in the same language of instruction. Regarding within-children effects, the probability of making an error of omission also decreased when the child produced an utterance that was longer than their average utterance. Thus, in English, the language that is less developed in these children, the probability of omission errors decreases with increases in utterance length, both when the child produces utterances whose average length of utterance is above average for the language of instruction as well for utterances that are longer than their average.

The results of Spanish suggest a similar pattern to English. Omission errors were infrequent and present in only about 2% of the utterances. Considering that 6-year-old children have acquired most of the language skills except for some complex syntax, the low rate of omission errors seen in Spanish for the children in this study is not surprising. In Spanish, the probability of making an error of omission declined as mean utterance length increased. That is, when a child produced an utterance in Spanish that was longer than their average utterance, the probability of making an error of omission decreased. Thus, a given child was less likely to produce an error of omission if they produced an utterance that was longer than their average utterance. However, children whose average utterance length in Spanish was above average showed the same probability of making errors of omission as children with average or below-average utterance length.

These findings are consistent with a developmental perspective on language acquisition: As children learn the language, they are acquiring more words and language structures and, therefore, they have fewer omission errors and longer utterances (Brown, 1973). Those children who produce English utterances that are longer, on average, are considered to have better English skills in general (i.e., longer MLUw compared to peers), which might explain why they make fewer errors of omission than children with average or below-average utterance length. This between-children effect of utterance length was nonsignificant for Spanish when subordination was included in the model predicting the probability of errors of omission. Errors of omission were very infrequent in Spanish (only 2% of utterances). Most children made no errors (n = 439) or only one error (n = 239) of omission in Spanish. Consequently, power for detecting effects of utterance length on the probability of errors in Spanish between children was low in these first-grade children whose L1 was Spanish. If we had sampled Spanish from an earlier developmental period, it is reasonable to expect that the rate of error production might have been higher, allowing greater power for detecting differences between children.

Errors of Commission and Utterance Length

In English, about 27% of all utterances included errors of commission. Between-children and within-child effects of utterance length were in the same direction. The probability of making a commission error was positively related to mean utterance length. Thus, children whose MLU was above average showed a higher probability of making errors of commission than children whose MLU was average or below average. When a child produced an utterance that was longer than their average utterance, there was also an increase in the probability of making a commission error. Thus, in English, a child was more likely to make an error of commission in their longer utterances, and children who had longer utterances, on average, also made more errors of commission. These results are also consistent with a developmental approach to language acquisition: As these ELs are acquiring more words and language structures in English, their utterances grow in length, but the newly acquired words and language forms are not always used correctly (Brown, 1973). A child might be more likely to make an error of commission during longer utterances because the longer utterance taxes the language system more as it is trying newly acquired language structures, which is consistent with Bock's (1982) Sentence Formulation Model. Interestingly, children whose MLU was above average made more errors of commission in English. This relationship may exist between children because, although their longer mean utterance indicates greater language skills in English than their classmates in the same language of instruction, these children are still learning English with increased processing demands, which leads to increased error probabilities.

In Spanish, commission errors were present in about 7% of all utterances. The model for commission errors in Spanish suggested that when a child produced an utterance that was longer than their average utterance, the probability of making an error of commission increased slightly. The directionality is the same in English, but the magnitude of the effect is significantly smaller. Interestingly, children with longer mean utterances showed no differences in error rates. Thus, in Spanish, there was not a statistically significant between-children effect for errors of commission, but only once subordination was included in the model. When subordination was not controlled, children with longer mean utterance length were less likely to produce errors, which is consistent with the idea that the effects of utterance length between children on error production reflect the effects of proficiency. More proficient children produce longer utterances, on average, and are less likely to produce errors. When subordination is included in the model, between-children effects of utterance length are not significant, which may again stem from the overall low error rate in Spanish and the small number of students who produced them, which would result in lower power for detecting unique effects of utterance length between persons.

We interpret these findings for English and Spanish as longer utterances taxing the language system within the child in a way that yielded more commission errors. The higher processing demands of producing longer utterances than the child's average might result in a trade-off effect, as suggested by Bock (1982), between utterance length and accuracy. We believe that children who produced longer utterances than their average utterance experience trade-off effects of sentence formulation, whereas the between-children relation to mean utterance length reflects the effects of language development on error production.

Within-Child and Between-Children Effects of Utterance Length in Spanish and English

Our results point to one key difference between the results for Spanish and English. In Spanish, the effect was the same for both types of errors but differed for the within child and between children. Specifically, in Spanish, within-child effects suggested an increase in commission errors and a decrease of omission errors with longer utterances, but between-children effects were nonsignificant for both errors of omission and errors of commission. In English, the direction of the effect was consistent for within-child and between-children effects and varied by type of error. Both within-child and between-children effects suggested a decrease in omission errors and an increase in commission errors with longer utterances.

Because of the differences between Spanish and English, we estimate that the relationship between errors and utterance length is dependent on the level of language development. When a child has mostly acquired the language, such as is the case in this study for Spanish, the likelihood of errors is low in general. However, when a child attempts to produce an utterance that is longer than their average utterance, the child is more likely to produce an error of commission because the processing demands for the child increases. When a child is learning the language, English in this study, the likelihood of errors is higher in general and differences in the within-child and between-children effects are similar for different error types because even children with longer utterances than average are still learning the language, which increases the processing demands during sentence formulation.

It is important to note that although the trade-off effect for Spanish was statistically significant, it was relatively small. Our results are in agreement with other studies that found trade-off effects in monolingual children (e.g., Bloom, 1990; Masterson & Kamhi, 1992; Owen, 2010). However, the size of the trade-off for the L1 of ELs may be different for tasks that use narrative language, such as it was the case in this study, than for more structured tasks (e.g., sentence completion task in Owen, 2010). In general, the trade-off between utterance length and commission errors is greater for a language in development in comparison to the L1, as indicated by the magnitude of the effect and the similarity between within-child and between-children effects.

Subordination

The results of this study suggest that subordination exerted a significant moderating influence on the probability of making an omission error as a function of utterance length. Producing a subordinated clause significantly increased the probability of making errors of omission in Spanish and English in comparison to utterances without subordination, controlling for utterance length. The effect was the opposite for errors of commission. In both Spanish and English, producing a subordinated clause while keeping the utterance at the average length for the child decreased the probability of making an error of commission. The interaction between SIu and utterance length was only significant for errors of omission in English.

These findings suggest that subordination taxed language production differently for errors of omission and errors of commission. We observed more omission errors in Spanish and English when subordination was involved, which is consistent with the results of Bloom (1990), Grela (2003), Grela and Leonard (2000), and Owen (2010) for monolingual English-speaking children in that higher rates of errors were observed in clauses with higher complexity. In contrast, we observed a decline in errors of commission as a function of utterance length in utterances that were subordinated, suggesting that a combination of increased utterance length and subordination boosted the reduction in errors in both Spanish and English.

Implications for Sentence Formulation Models

The results of this study offer some support for the sentence formulation framework proposed by Bock (1982). We observed trade-off effects between errors of commission and utterance length and complexity at the child level, and these trade-off effects were larger in English than in Spanish. We interpret these findings to be supportive of a trade-off effect because sentence formulation in English poses greater processing demands in these ELs than sentence formulation in Spanish, which is the L1 for these children. However, the results of this study add an important developmental piece to models of sentence formulation. The trade-off effects are smaller in magnitude for the L1 in comparison to a language in development. We hypothesize that if we were to observe these children in Spanish when they were younger, we would have observed larger trade-off effects in Spanish, which is consistent with the results of Grela (2003) and Grela and Leonard (2000) who observed that younger children had more errors in complex utterances in comparison to older children.

Clinical and Research Implications

The children in this study included a large sample of children with a wide range of language skills. Although we did not study children with language disorders specifically, we examined between-children effects of utterance length for those children who performed at the average level, and above and below average. For children whose average length of utterance was one word lower than the average for all children in the same language of instruction (see Figure 1, leftmost panel), we observed that the probability of making an error of commission increased in English with increasing utterance length, while the probability of making an error of omission decreased. Therefore, children whose mean utterance length was below average were more likely to have fewer errors of commission, although more errors of omission in English. These results are in agreement, in principle, with Simon-Cereijido and Gutiérrez-Clellen (2007), who suggested to closely examine utterance length when a child suspected of a language disorder is not making errors. Interestingly, we did not observe a between-children effect of increasing average utterance length for Spanish. It is possible that between-children effects of utterance length do not exist in Spanish because there is not a developmental difference in the relationship between errors and utterance length once the language is mostly acquired. It is important to further examine the within-child and between-children effects of errors, utterance length, and complexity, and error production in a group of children with language disorders, who presumably would have lower MLU, on average. Any increase in complexity may affect the probability of making errors differently in children with developmental language disorders, as suggested previously by Leonard et al. (2000) and Owen (2010), due to the limited processing abilities seen in these children.

From a research perspective, there are two main implications from this study. First, the knowledge gained in this study adds insight into the length and complexity of stimulus items. Test developers might use the information regarding the effects of utterance length and complexity on the probability of errors at the child level to design items that capture higher processing–dependent skills. Second, the results of this study showed some differences between within-child and between-children effects. Most of the research on child language has focused on group differences. Although important, between-groups comparisons do not always capture the subtleties involved in learning at the individual level. More research exploring both within-child and between-children effects should be conducted to better understand individual differences in language learning.

Limitations and Future Directions

Although this study included a large number of children, the number of utterances produced by each child in each language was not large across a wide array of utterance lengths within child, which affected our ability to include a random effect for utterance length in the statistical models. A second limitation is that, although we included children whose MLU was below average for children in a given language of instruction but not identified with a disability in kindergarten, it is not clear whether any of the participants were at risk of being later identified with a language disability. Future studies should include children with and without language disorders to identify potential effects of language abilities on the relationship between errors, utterance length, and sentence complexity. Third, we did not have specific information regarding the English language exposure for these children other than the language of instruction. English exposure may play a moderating effect for the effects seen in this study (i.e., children with more English exposure might have fewer errors). Last, it is important to note that increases in utterance length and the use of subordinated clauses are not the only factors influencing errors. Other factors such as memory limitations, efficiency and speed in lexical retrieval, and language ability also impact the production of errors in child language (McKee et al., 2017).

Conclusions

The purpose of this study was to examine the relationship between the probability of making errors and utterance length and complexity. The results of this study present three main contributions to the field of language acquisition. First, it is pivotal to examine errors of omission and errors of commission separately to be able to observe the differential effects of utterance length on the two types of errors. Second, for ELs, the relationship between errors, utterance length, and complexity depends on the language: The trade-off between utterance length and errors was greater for the language in development (English) in comparison to the L1 (Spanish). Third, subordination seems to tax the language production system, resulting in a higher likelihood of errors of omission. The relationship between utterance length and grammaticality is complex and varies by language, error type, and whether our frame of reference is the child's own language or the language of other children. The results of this study support a general framework of sentence formulation that predicts trade-off effects between utterance length, complexity, and accuracy.

Acknowledgments

Research reported in this publication was supported by National Institute on Deafness and Other Communication Disorders Award Number K23DC015835 granted to Anny Castilla-Earls, National Institute for Child Health and Human Development Award Number HD39521 granted to David Francis, and U.S. Department of Education Award Number R305U010001 granted to David Francis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Baron, A., Bedore, L. M., Peña, E. D., Lovgren-Uribe, S. D., López, A. A., & Villagran, E. (2018). Production of Spanish grammatical forms in U.S. bilingual children. *American Journal of Speech-Language Pathology, 27*(3), 975–987. https://doi.org/10.1044/2018_ajslp-17-0074

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 21*(4), 491–504. https://www.jstor.org/stable/4178692

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*(1), 1–47. https://doi.org/10.1037/0033-295X.89.1.1

Branum-Martin, L., Mehta, P. D., Carlson, C. D., Francis, D. J., & Goldenberg, C. (2014). The nature of Spanish versus English language use at home. *Journal of Educational Psychology, 106*(1), 181–199. https://doi.org/10.1037/a0033931

Branum-Martin, L., Mehta, P. D., Francis, D. J., Foorman, B. R., Cirino, P. T., Miller, J. F., & Iglesias, A. (2009). Pictures and words: Spanish and English vocabulary in classrooms. *Journal of Educational Psychology, 101*(4), 897–911. https://doi.org/10.1037/a0015817

Brown (1973). *A first language: The early stages*. Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Castilla-Earls, A. P., & Eriks-Brophy, A. (2012). Spontaneous language measures in monolingual preschool Spanish-speaking children. *Revista de Logopedia, Foniatría y Audiología, 32*(3), 97–108. https://doi.org/10.1016/j.rlfa.2012.04.001

Castilla-Earls, A. P., Francis, D., Iglesias, A., & Davidson, K. (2019). The impact of the Spanish-to-English proficiency shift on the grammaticality of English learners. *Journal of Speech, Language, and Hearing Research, 62*(6), 1739–1754. https://doi.org/10.1044/2018_JSLHR-L-18-0324

Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the simple view of reading to account for variation within readers and across texts: The complete view of reading (CVR i). *Remedial and Special Education, 39*(5), 274–288. https://doi.org/10.1177/0741932518772904

Grela, B. G. (2003). The omission of subject arguments in children with specific language impairment. *Clinical Linguistics & Phonetics, 17*(2), 153–169. https://doi.org/10.1080/0269920031000061812

Grela, B. G., & Leonard, L. B. (2000). The influence of argument-structure complexity on the use of auxiliary verbs by children with SLI. *Journal of Speech, Language, and Hearing Research, 43*(5), 1115–1125. https://doi.org/10.1044/jslhr.4305.1115

Guo, L. Y., Eisenberg, S., Schneider, P., & Spencer, L. (2019). Percent grammatical utterances between 4 and 9 years of age for the Edmonton Narrative Norms Instrument: Reference data and psychometric properties. *American Journal of Speech-Language Pathology, 28*(4), 1448–1462. https://doi.org/10.1044/2019_AJSLP-18-0228

Gutiérrez-Clellen, V. F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech and Hearing Research, 37*(3), 645–654. https://doi.org/10.1044/jshr.3703.645

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children. *Language, Speech, and Hearing Services in Schools, 31*(1), 88–98. https://doi.org/10.1044/0161-1461.3101.88

Leonard, L. B., Miller, C. A., Grela, B., Holland, A. L., Gerber, E., & Petucci, M. (2000). Production operations contribute to the grammatical morpheme limitations of children with specific language impairment. *Journal of Memory and Language, 43*(2), 362–378. https://doi.org/10.1006/jmla.1999.2689

Masterson, J. J., & Kamhi, A. G. (1992). Linguistic trade-offs in school-age children with and without language disorders. *Journal of Speech and Hearing Research, 35*(5), 1064–1075. https://doi.org/10.1044/jshr.3505.1064

Mayer, M. (1967). *A boy, a dog and a frog*. Penguin Books.

McKee, C., McDaniel, D., & Garrett, M. F. (2017). Children's performance abilities. In *The handbook of psycholinguistics* (pp. 491–515). Wiley & Sons, Inc. https://doi.org/10.1002/9781118829516.ch22

Miller, J. F., Heilman, J., Nockerts, A., Iglesias, A., Fabiano, L., & Francis, D. J. (2006). Oral language and reading in bilingual children. *Learning Disabilities Research & Practice, 21*(1), 30–43. https://doi.org/10.1111/j.1540-5826.2006.00205.x

Miller, J. F., & Iglesias, A. (2019). *Systematic Analysis of Language Transcripts (SALT), Research Version 2019* [Computer software]. SALT Software.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 92–105). American Psychological Association. https://doi.org/10.1037/10099-006

Owen, A. J. (2010). Factors affecting accuracy of past tense production in children with specific language impairment and their typically developing peers: The influence of verb transitivity, clause location, and sentence type. *Journal of Speech, Language, and Hearing Research, 53*(4), 993–1014. https://doi.org/10.1044/1092-4388(2009/09-0039)

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1398–1411. https://doi.org/10.1044/jslhr.4106.1398

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research, 49*(4), 793–808. https://doi.org/10.1044/1092-4388(2006/056)

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*(3), 351. https://doi.org/10.2307/2087176

Rojas, R., & Iglesias, A. (2013). The language growth of Spanish-speaking English language learners. In *Child development* (Vol. 84, Issue 2, pp. 630–646). Wiley Society for Research in Child Development. https://doi.org/10.1111/j.1467-8624.2012.01871.x

Rosenbaum, P. R., & Silber, J. H. (2009). Amplification of sensitivity analysis in matched observational studies. *Journal of the American Statistical Association, 104*(488), 1398–1405. https://doi.org/10.1198/jasa.2009.tm08470

Scott, C. M. (1988). Producing complex sentences. *Topics in Language Disorders, 8*(2), 44–62. https://doi.org/10.1097/00011363-198803000-00006

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324–339. https://doi.org/10.1044/jslhr.4302.324

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*(2), 317–339. https://doi.org/10.1017.S0142716407070166

StataCorp. (2017). *Stata Statistical Software: Release 15*. StataCorp, LLC.