The Role of Developmental Levels in Examining
the Effect of Subject Types on the Production of
Auxiliary *is* in Young English-Speaking Children

Ling-Yu Guo,a Amanda J. Owen Van Horne,b and J. Bruce Tomblinb

**Purpose:** Prior work (Guo, Owen, & Tomblin, 2010) has shown that at the group level, auxiliary *is* production by young English-speaking children was symmetrical across lexical noun and pronominal subjects. Individual data did not uniformly reflect these patterns. On the basis of the framework of the gradual morphosyntactic learning (GML) hypothesis, the authors tested whether the addition of a theoretically motivated developmental measure, tense productivity (TP), could assist in explaining these individual differences.

**Method:** Using archival data from 20 children between age 2;8 and 3;4 (years;months), the authors tested the ability of 3 developmental measures (TP; finite verb morphology composite, FVMC; mean length of utterance, MLU) to predict use of auxiliary *is* with different subject types.

**Results:** TP, but not MLU or FVMC, significantly improved model fit. Children with low TP scores produced auxiliary *is* more accurately with pronominal subjects than with lexical subjects. The facilitative effect of pronominal subjects on the production of auxiliary *is*, however, was not found in children with high TP scores.

**Conclusion:** The finding that the effect of subject types on the production accuracy of auxiliary *is* changed with children's TP is consistent with the GML hypothesis.

**Key Words:** tense, gradual morphosyntactic learning hypothesis, auxiliary *is*, subject types

Young children tend to variably use tense and agreement morphemes throughout the preschool years. For instance, even within the same lab visit, children will sometimes produce and sometimes omit auxiliary *is* in experimental tasks (e.g., Grela & Leonard, 2000) and spontaneous speech (Hadley & Rice, 1996; Leonard, 1995). One question facing researchers is how to explain this variable performance. Some attempts have been made to link the variability to properties of the sentences themselves, such as prosody (Gerken & McIntosh, 1993), sentence complexity (Grela & Leonard, 2000; Owen, 2010), and frequency of sentence elements (Pine, Conti-Ramsden, Joseph, Lieven, & Serratrice, 2008; Wilson, 2003).

To fully understand the developmental pattern of tense and agreement, it is important to explore how the influence of sentence properties (e.g., sentence complexity, frequency of sentence elements) changes with children’s knowledge of tense and agreement. Most of the current accounts tend to ignore the potential interaction between sentence properties and children’s developmental levels partly because these accounts do not have a theoretically motivated measure of the development of abstract knowledge of tense and agreement. For instance, the usage-based approach (Tomasello, 2003) predicts that young children’s production of auxiliary *be* forms should be more accurate with pronominal subjects than with lexical noun phrase (NP) subjects. This is because young children have not acquired abstract knowledge of auxiliary *be*. Pronoun + auxiliary *be* (e.g., *He’s running*) combinations are likely to be learned as a whole unit, supporting children’s production. Though the usage-based approach suggests that the difference between subject types in the production accuracy of auxiliary *be* will diminish as children acquire the adultlike
The GML Hypothesis

The GML hypothesis (Hadley, Rispoli, Fitzgerald, & Bahnsen, 2011; Rispoli & Hadley, 2011; Rispoli et al., 2009) is grounded in universal grammar. It argues that the development of the tense-and-agreement system involves setting up the [+tense] parameter, learning the specific surface forms of individual tense and agreement morphemes, and automatizing the incorporation of these morphemes into sentence production.

Adopting the assumption of the variational learning model (Legate & Yang, 2007; Yang, 2002), the GML hypothesis assumes that the [+tense] parameters are innately available to children as a component of the grammar instantiated in their mental representations without any default setting (Legate & Yang, 2007). Although the [+tense] parameter requires tense marking in sentences, the [-tense] parameter does not. Children gradually set the [-tense] parameter in a probabilistic manner based on the ambient input. For children whose ambient language marks tense (e.g., Spanish, English), they will gradually eliminate the [-tense] parameter in favor of the [+tense] parameter in a probabilistic manner as they identify which grammar they should adopt for their own utterances. When children hear an utterance with overt tense marking, the learning algorithm reinforces [+tense] grammar by incrementally boosting the probability that a child will select a grammar that contains [+tense] and “punishes” [-tense] other grammar by reducing the probability that that grammar will be chosen, leading to a gradual convergence on the grammar associated with the target language.

The GML hypothesis further argues that setting up the [+tense] parameter is necessary but not sufficient for producing tense and agreement morphemes. In order to do so, children have to learn the surface forms of these morphemes as well as their associated morphosyntactic features (e.g., [+past], [+aspect]). In addition, the morphosyntactic features must come to trigger the tense and agreement morphemes automatically in the real-time procedures of grammatical encoding (Rispoli & Hadley, 2011).

The production of tense morphemes in sentences may occur both via direct activation of memorized forms and via construction of inflected forms through grammatical encoding that involve semantic and morphosyntactic processing (Rispoli & Hadley, 2011; Rispoli et al., 2009). Direct activation is a route of language production through which “direct associative connections between referential content and phonetic content” (Bock, 1982, p. 24) are used, omitting the use of true morphosyntactic representations (Rispoli et al., 2009). Direct activations occur because they are highly frequent in the input. For instance, combinations of pronominal subjects and contracted auxiliary is are more likely to be produced via direct activation because these pronoun + contracted auxiliary is combinations are highly frequent and may be learned as limited scope formulae. Reliance on direct activation to produce tense morphemes is especially likely to be the case early in development because young children are still learning the tense-and-agreement system.

With development, children must shift from primary reliance on direct activations to grammatical encoding (Bock & Levelt, 1994) as the number of messages they attempt to express expands. The production of tense and agreement morphemes via grammatical encoding should increase with the development of the tense system. Auxiliary is with lexical NP subjects is more likely to be produced via grammatical encoding. This is because the lexical NP + auxiliary is combinations are relatively low in frequency and thus are less likely to be memorized as a unit (Wilson, 2003). Thus, the production of lexical NP + auxiliary is combinations in sentences may reflect the developmental trajectory of the tense system.

Predicting Children’s Use of Auxiliary Is With Their Developmental Levels of Tense

Within the GML hypothesis, the degree to which children produce tense morphemes via grammatical
encoding should be observable through the diverse use of these morphemes. To that end, Hadley and colleagues have developed a new measure of development of tense—and-agreement use—the TP score (Hadley & Holt, 2006; Hadley & Short, 2005). The TP score is computed by identifying cases of the productive use of morphemes unified under the [+tense] parameter: copula be, auxiliary be, third person singular -s, past tense -ed, and auxiliary do. The child receives a point for each sufficiently different use of these morphemes, with up to 5 points for each morpheme and maximum score of 25. To be counted as a sufficiently different use, verb inflections must appear on different lexical verbs, and copula be, auxiliary be, and auxiliary do must occur with different sentence subjects. All forms contracted with pronominal subjects (e.g., he’s, she’s, it’s) are excluded from analysis because they might have been produced through direct activation of memorized chunks instead of grammatical encoding and therefore would not reflect children’s true morphosyntactic knowledge (Bock, 1982; Bock & Levelt, 1994; Rispoli et al., 2009). Higher TP scores indicate that children are more advanced in producing tense morphemes via grammatical encoding, which in turn implies that children have attained a higher developmental level with regard to the tense system.

On the basis of the dual-route-processing (i.e., direct activation and grammatical encoding) assumption, the GML hypothesis makes two specific predictions on the production of auxiliary is. First, the production accuracy of auxiliary is should be higher with pronominal subjects than with lexical NP subjects in children with low developmental levels of tense (i.e., children with low TP scores) because these children are more likely to rely on direct activations of memorized chunks to produce tense morphemes. Second, the production accuracy of auxiliary is should become more symmetrical across subject types in children with high developmental levels of tense (i.e., children with high TP scores). This is because, as children’s developmental levels of tense increases, there is a general shift from production via direct activation to sentence production using grammatical encoding, and all sentence types are treated similarly when grammatical encoding is used. Nonetheless, direct activation as a strategy for production may persist for highly frequent forms beyond the point at which grammatical encoding becomes more generally available. It should also be noted that even if children shift from direct activation to grammatical encoding for sentence production, it does not mean that they will be able to produce auxiliary is at an adultlike level immediately because children still have to automate the process of incorporating auxiliary is during grammatical encoding (Hadley et al., 2011). Thus, the symmetry between subject types could occur before children’s production accuracy of auxiliary is reaches an adultlike level.

Potential Concerns in Using TP Scores to Predict the Use of Auxiliary Is

There are two potential concerns in using TP scores to predict the effect of subject types on the production accuracy of auxiliary is: (a) interdependence between the predictor and outcome variables and (b) the uniqueness of TP scores as compared with other developmental measures. First, the TP score includes the use of auxiliary be with a variety of subject types as a component of the measure. Thus, the logic may be circular: We could be predicting the use of auxiliary is with a variety of subject types in an elicited production task via a measure that includes the use of auxiliary be with a variety of subject types. With this in mind, we use TP without auxiliary be included to ensure that it is the measure as a whole and not just the presence of auxiliary be that predicts children’s performance.

A second concern is that the asymmetry of subject types in the production accuracy of auxiliary is may disappear simply because of children’s overall language development—that is, the decreasing asymmetry between subject types in the production accuracy of auxiliary is may not be unique to the increasing TP, and hence the GML hypothesis. If that is the case, then any global measure of language development will be equally predictive. To that end, we also considered whether two well-established measures of language development—mean length of utterance (MLU) and the fine verb morphology composite (FVMC)—would yield similar results.

MLU, although controversial in terms of what exactly it measures (DeThorne, Johnson, & Loeb, 2005; Eisenberg, Fersko, & Lundgren, 2001), is a well-recognized measure of general language development, particularly at younger ages (Miller, 1981; Scarborough, Wyckoff, & Davidson 1986), and it reliably differentiates typical and impaired populations even at higher MLU levels (Redmond, 2004; Rice et al., 2010). More specific than MLU, the FVMC (Goffman & Leonard, 2000; Leonard, Miller, & Gerber, 1999) is a measure of the use of tense markers in a spontaneous language sample. Instead of computing productivity of tense markers, FVMC computes the percentage of accuracy of tense markers in obligatory contexts. Specifically, it measures percentage of use in obligatory contexts of copula be, auxiliary be, third person singular -s, and past tense -ed; it does not include auxiliary do. FVMC is a reliable means of differential diagnosis of children with specific language impairment (e.g., Bedore & Leonard, 1998). Like TP and MLU, FVMC increases with age before plateauing in the late preschool years (Goffman & Leonard, 2000). FVMC differs from TP in that it counts all productions and omissions of target tense markers in obligatory contexts and makes no attempt to differentiate memorized and constructed forms. For instance, He’s running is a correct
use of auxiliary be under FVMC, but it is not included in TP. FVMC provides a measure of language development that is more specific to grammatical development than MLU and is more focused on overall accuracy of tense than TP.

**The Present Study**

To test the predictions of the GML hypothesis, we examined in the present study the extent to which the relationship between subject types and auxiliary is production changes with children’s developmental levels as measured by (a) TP scores computed without auxiliary BE (i.e., TP2), (b) mean length of utterance in morphemes (MLUm), and (c) FVMC. Specifically, is the production accuracy of auxiliary is higher with pronominal subjects than with lexical NP subjects in children with low developmental levels for each measure? Does the facilitative effect of pronominal subjects diminish in children with high developmental levels? An interaction effect between subject types and TP scores would provide support for the GML hypothesis.

**Method**

The present study is a follow-up to that of Guo, Owen, and Tomblin (2010), in which the effect of subject types on the production accuracy of auxiliary is was investigated in 20 three-year-olds. Data from Guo et al. are summarized in Table 1. At the group level, production accuracy of auxiliary is did not vary with subject types. The data from Guo et al. were therefore reanalyzed to test the predictions of the GML hypothesis. A description of how the original data were collected and analyzed follows.

**Participants**

Twenty monolingual English-speaking typically developing children with a mean age of 3;0 (range = 2;8–3;4 [years;months]) participated in this study. These 20 children (11 girls) performed below ceiling level (i.e., 90% correct; Brown, 1973) and above floor level (at least one correct use in the task or in spontaneous speech; Pine et al., 2008; Schütze, 2001) of auxiliary is use.

### Table 1. Language sample measures and experimental performance by participant.

<table>
<thead>
<tr>
<th>Childa</th>
<th>Age (years;months)</th>
<th>TP1</th>
<th>TP2</th>
<th>FVMC</th>
<th>MLUm</th>
<th>Accuracy of auxiliary is in language samples</th>
<th>Accuracy of auxiliary is in the task</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV</td>
<td>2;10</td>
<td>15</td>
<td>13</td>
<td>0.81</td>
<td>4.06</td>
<td>0.43</td>
<td>0.73 1.00</td>
</tr>
<tr>
<td>AV2</td>
<td>3;0</td>
<td>14</td>
<td>13</td>
<td>0.83</td>
<td>4.00</td>
<td>0.00</td>
<td>0.50 1.00</td>
</tr>
<tr>
<td>EE</td>
<td>2;11</td>
<td>13</td>
<td>12</td>
<td>0.90</td>
<td>4.54</td>
<td>1.00</td>
<td>0.93 0.95</td>
</tr>
<tr>
<td>LG</td>
<td>2;11</td>
<td>12</td>
<td>12</td>
<td>0.97</td>
<td>5.73</td>
<td></td>
<td>0.73 0.63</td>
</tr>
<tr>
<td>AB</td>
<td>2;11</td>
<td>12</td>
<td>12</td>
<td>0.76</td>
<td>4.26</td>
<td></td>
<td>0.82 0.67</td>
</tr>
<tr>
<td>SC</td>
<td>2;10</td>
<td>13</td>
<td>12</td>
<td>0.93</td>
<td>3.36</td>
<td></td>
<td>0.78 0.83</td>
</tr>
<tr>
<td>AH</td>
<td>2;11</td>
<td>13</td>
<td>10</td>
<td>0.71</td>
<td>4.33</td>
<td>0.80</td>
<td>0.57 0.59</td>
</tr>
<tr>
<td>CK</td>
<td>3;4</td>
<td>10</td>
<td>10</td>
<td>0.61</td>
<td>4.75</td>
<td>0.61</td>
<td>0.88 0.88</td>
</tr>
<tr>
<td>CH</td>
<td>2;8</td>
<td>9</td>
<td>9</td>
<td>0.84</td>
<td>3.93</td>
<td></td>
<td>0.67 0.24</td>
</tr>
<tr>
<td>AV</td>
<td>3;4</td>
<td>8</td>
<td>8</td>
<td>0.93</td>
<td>4.25</td>
<td>1.00</td>
<td>0.80 0.80</td>
</tr>
<tr>
<td>IR</td>
<td>2;11</td>
<td>10</td>
<td>8</td>
<td>0.76</td>
<td>5.27</td>
<td></td>
<td>0.33 0.65</td>
</tr>
<tr>
<td>AS</td>
<td>2;11</td>
<td>7</td>
<td>7</td>
<td>0.33</td>
<td>4.01</td>
<td>0.00</td>
<td>0.45 0.09</td>
</tr>
<tr>
<td>CD</td>
<td>3;3</td>
<td>7</td>
<td>7</td>
<td>0.82</td>
<td>4.12</td>
<td></td>
<td>0.67 0.27</td>
</tr>
<tr>
<td>ET</td>
<td>3;3</td>
<td>7</td>
<td>7</td>
<td>0.52</td>
<td>4.41</td>
<td>0.33</td>
<td>0.33 0.84</td>
</tr>
<tr>
<td>ZE</td>
<td>3;0</td>
<td>7</td>
<td>7</td>
<td>0.61</td>
<td>3.49</td>
<td></td>
<td>0.50 0.36</td>
</tr>
<tr>
<td>AM</td>
<td>3;2</td>
<td>7</td>
<td>6</td>
<td>0.89</td>
<td>3.81</td>
<td>1.00</td>
<td>0.80 0.80</td>
</tr>
<tr>
<td>HP</td>
<td>2;10</td>
<td>4</td>
<td>4</td>
<td>0.87</td>
<td>3.02</td>
<td></td>
<td>0.11 0.05</td>
</tr>
<tr>
<td>TK</td>
<td>3;1</td>
<td>3</td>
<td>3</td>
<td>0.70</td>
<td>4.27</td>
<td>1.00</td>
<td>0.60 0.81</td>
</tr>
<tr>
<td>KS</td>
<td>2;11</td>
<td>3</td>
<td>2</td>
<td>0.72</td>
<td>4.13</td>
<td>0.67</td>
<td>0.78 0.19</td>
</tr>
<tr>
<td>SF</td>
<td>2;11</td>
<td>2</td>
<td>1</td>
<td>0.88</td>
<td>3.75</td>
<td></td>
<td>0.38 0.50</td>
</tr>
<tr>
<td>M</td>
<td>3;0</td>
<td>8.90</td>
<td>8.15</td>
<td>0.77</td>
<td>4.18</td>
<td></td>
<td>0.62 0.58</td>
</tr>
<tr>
<td>SD</td>
<td>0;2</td>
<td>4.00</td>
<td>3.67</td>
<td>0.16</td>
<td>0.61</td>
<td></td>
<td>0.22 0.31</td>
</tr>
</tbody>
</table>

Note. TP1 = tense productivity scores with auxiliary be; TP2 = tense productivity scores without auxiliary be; FVMC = finite verb morphology composite; MLUm = mean length of utterance in morpheme; NP = noun phrase. Dashes indicate “not computed.”

*Letters in this column represent children’s initials.*
and produced at least three scorable items for each condition.

**Spontaneous Language Samples**

As a component of assessment to determine typical development, Guo et al. (2010) collected a 15- to 20-min conversational language sample for each child. The child talked to the examiner while he or she was playing with toys in a child-friendly test room. The language samples were transcribed by research assistants based on the Systematic Analysis of Language Transcripts conventions developed by Miller and Chapman (2000). The first author retranscribed four samples (20% of the data) to check reliability. The average agreement was 0.86 (SD = 0.03) in utterance segmentation and 0.89 (SD = 0.03) in morpheme-to-morpheme agreement. The mean number of utterances across participants was 132.90 (SD = 24.22), suggesting the samples were sufficiently large for further analysis.

From the language samples, children’s TP scores were computed with and without auxiliary be, MLUm, and FVMC to document their developmental levels. The reliability in identifying target morphemes was high for the TP scores (Cohen’s $\kappa = 0.90$, $SD = 0.09$) and for FVMC ($\kappa = 0.93$, $SD = 0.10$). These language sample measures of each child are presented in Table 1. It should be noted that TP scores with auxiliary be (TP1) are listed for descriptive purposes because this measure has been used in other studies. Only TP scores without auxiliary be (TP2) were used for the statistical analysis in order to avoid the potential for circularity in evaluating how well the TP score predicted experimental performance.

The average across children was 8.75 for TP scores ($SD = 4.13$, range = 2–15), 4.18 for MLU (SD = 0.61, range = 3.02–5.74), and 0.77 for FVMC (SD = 0.16, range = 0.33–0.97). TP2 was significantly correlated with TP score with TP1 ($r = .98$, $p < .0001$), but TP2 was not significantly correlated with the child’s age ($r = -.14$, $p > .05$), MLUm ($r = .31$, $p > .05$), or FVMC ($r = .19$, $p > .05$).

**Experimental Stimuli**

Fuller descriptions of the experimental stimuli, testing protocol, and scoring procedures can be found in Guo et al. (2010), but a brief description is provided here for the ease of reading. The target items consisted of 30 simple declarative transitive sentences that required the use of auxiliary is and varied with regard to subject types (i.e., pronominal high-frequency lexical NP and low-frequency lexical NP subjects). Each condition had 10 target sentences.

The subject words were three pronominal subjects (i.e., he, she, it), six high-frequency lexical subjects (i.e., eat, dog, duck, goat, Mom, pig), and five low-frequency lexical subjects (i.e., ant, deer, frog, queen, sheep). All these lexical items were repeated as subjects across target sentences, except duck and Mom. Verb frequency and sentence length were controlled across conditions. It should be noted that although Guo et al. (2010) separated the high- and low-frequency lexical NP conditions, these conditions were combined in the present analysis into one condition—the lexical NP condition. This is because there was no difference between the high- and low-frequency lexical NP conditions in Guo et al. (2010), and the GML hypothesis does not make specific predictions about lexical NPs with different frequencies.

Each child was tested individually by a trained examiner. The child was presented with a pair of drawings (i.e., nontarget and target pictures) designed to elicit the target structure. Elicitation prompts varied slightly across conditions to maximize the likelihood that children would produce pronouns or lexical NPs as subjects. During each trial, the examiner showed two pictures, briefly described them with verb phrases, and then said a full sentence describing the nontarget picture in which a plural subject was used (e.g., “The horses are wiping the floor”). Then the child was asked to describe a target picture that contained a single agent engaging in a transitive event. For instance, to elicit a pronominal item, such as “He’s eating a cookie,” the prompt was “Watching TV (point to the nontarget picture). Eating a cookie (point to the target picture). They’re watching TV (point to the nontarget picture). What’s happening to him? (point to the target picture).” To elicit a full lexical NP item, such as “The goat’s eating the car,” the prompt was “Wiping the floor. Driving the car. The horses are wiping the floor. What’s happening in this picture?” At the end of the prompt, the examiner pointed to the target picture and waited for the child’s response. The examiner used up to three additional prompts for each item, which are described in Guo et al. (2010).

The children’s responses were transcribed by the examiner and coded as correct, incorrect, or unscorable. Correct responses included productions containing auxiliary is and the target subject (e.g., He’s eating a cookie; 244 responses) or an alternative subject (e.g., He’s eating a cookie was produced as The bear’s eating a cookie). Responses with an alternative subject were reclassified according to subject frequency to an appropriate response category (73 responses). Of the responses that included an alternative subject, seven included non-nominative case pronouns (e.g., Him is eating a cookie). Incorrect responses included omission of auxiliary is using the target subject (e.g., He eating a cookie; 148 responses) or an alternative subject (e.g., The bear eating...
a cookie; 51 responses). Fifteen of the 51 alternative subjects were nonnominative case pronouns (e.g., Him eating a cookie). Agreement errors (e.g., *He is eating a cookie) were unattested in the data. Unscorable responses consisted of unrelated responses (e.g., He’s happy; 16 responses), subject omission (e.g., Eating a cookie; 61 responses), and no responses (six responses). Unscorable responses were discarded and not considered further. The reliability in coding the correctness of auxiliary is was 0.91.

**Statistical Analysis and Summary of Previous Results**

The dependent measure in Guo et al. (2010) was accuracy of auxiliary is of each scorable response for each child. Because of the dichotomous (i.e., correct or incorrect) nature of the dependent measure, a binominal logistic regression model was adopted to evaluate the effect of subject types on the production of auxiliary is. In Model A, the dependent variable (Yi) is the accuracy of auxiliary is, and the predictor variable was subject types (where SUBJ Pron refers to the pronominal condition). The lexical NP subject condition served as the reference condition, which is not explicitly listed in the model.

Model A: Logit (Yi) = β0 + β1SUBJPron.  

The production accuracy of auxiliary is did not vary with subject types (pronominal/lexical OR = 1.11, p = .65). Critically for this study, including the most obvious developmental measure—age in months—also did not significantly improve model fit, either as a main effect, χ²(1, N = 561) = 1.81, p = .18, or as an interaction term, χ²(2, N = 561) = 2.75, p = .25. Because the present study tested whether the effect of pronominal and lexical NP subjects on the production of auxiliary is would change with children’s developmental levels, Model A was used as the reduced model for testing other developmental measures and is shown in Table 2.

### Results

**TP Scores and the Effect of Subject Types**

To assess whether there was an effect of the child’s TP score in combination with the subject types on the production of auxiliary is, we performed two mixed-model binominal logistic regressions using TP2 scores as a main effect (Model B) and as an interaction term (Model C). In all of the models, each child and item were treated as random factors, and subject types and the child’s TP2 score were treated as fixed factors, as illustrated in Equation 2:

Model B: Logit (Yi) = β0 + β1SUBJPron + β2TP2  
Model C: Logit (Yi) = β0 + β1SUBJPron + β2TP2 + β3SUBJPron × TP2.

A chi-square goodness-of-fit test showed that Model C was the best fit for the data, improving model fit as compared with Model B, χ²(1, N = 561) = 5.41, p = .02, and Model A (Model A vs. Model B), χ²(1, N = 561) = 10.95, p < .001. That is, subject types had different effects on auxiliary is production, depending on the child’s TP scores. The results of Model C are reported in Table 3. To ease interpretation, a plot of the model in probability space is shown in Figure 1.

As can be seen in Figure 1, the production accuracy of auxiliary is increased with children’s TP scores, but the pattern was different depending on the subject type. The asterisk means “ungrammatical.”

---

**Table 2. Regression Model A showing the likelihood of producing auxiliary is correctly by subject type.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>0.04</td>
<td>0.22</td>
</tr>
<tr>
<td>Child</td>
<td>1.28</td>
<td>1.33</td>
</tr>
<tr>
<td>Fixed factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.53</td>
<td>0.29</td>
</tr>
<tr>
<td>Subject type (reference condition = lexical NP subjects)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronominal subjects</td>
<td>0.10</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Note. OR = odds ratio.

---

**Table 3. Regression Model C showing the likelihood of producing auxiliary is correctly by subject type and tense productivity score when auxiliary is was excluded.**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>0.07</td>
<td>0.28</td>
</tr>
<tr>
<td>Child</td>
<td>0.67</td>
<td>0.82</td>
</tr>
<tr>
<td>Fixed factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>−1.71</td>
<td>0.56</td>
</tr>
<tr>
<td>Subject type (reference condition = lexical NP subjects)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronominal subjects</td>
<td>1.28</td>
<td>0.54</td>
</tr>
<tr>
<td>Tense production (TnsProd)</td>
<td>0.28</td>
<td>0.06</td>
</tr>
<tr>
<td>Subject Type × TnsProd (reference condition = lexical NP subjects)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pronominal Subjects × TnsProd</td>
<td>−0.14</td>
<td>0.06</td>
</tr>
</tbody>
</table>

---
type used. Children with lower TP scores were more likely to produce auxiliary is accurately with pronominal subjects than with lexical NP subjects. However, the facilitative effect of pronominal subjects diminished in children with higher TP.

### Alternative Developmental Measures

We also considered whether two well-established measures of language development—MLUm and FVMC—would yield similar results, potentially negating the focus on the GML hypothesis and placing the emphasis on developmental changes in general. We tested four models, two for MLUm and two for FVMC, considering each developmental measure as a main effect and as an interaction term. Neither the model that included MLUm as a main effect, \( \chi^2(1, N = 561) = 3.25, p = .07 \), nor the model that included MLUm as an interaction term, \( \chi^2(1, N = 561) = 0.13, p = .72 \), provided a better model fit as compared with Model A. As with MLUm, the addition of FVMC did not significantly improve model fit as compared with Model A: main effect, \( \chi^2(1, N = 561) = 1.16, p = .28 \); interaction term, \( \chi^2(1, N = 561) = 0.23, p = .63 \). Thus, neither of the more general measures of language development improved model fit as compared with a simpler model.

### Discussion

In the present article, we set out to test the validity of the GML hypothesis by examining the use of auxiliary is with pronominal and lexical NP subjects in children with different developmental levels, as measured by TP2, MLUm, and FVMC. The TP score proved to be the only significant predictor of the level of accuracy in the experimental task. The production accuracy of auxiliary is was higher with pronominal subjects than with lexical NP subjects in children with low TP scores. The facilitative effect of pronominal subjects, however, diminished in children with high TP scores even if these children still produced auxiliary is below the ceiling level.

### Testing the GML Hypothesis

Recall that the GML hypothesis predicts that the production accuracy of auxiliary is should be higher with pronominal subjects than with lexical NP subjects in children with low developmental levels of tense and that the differences between pronouns and noun phrases should decrease as children’s developmental levels of tense increase. These changes are due to an initial reliance on direct activation of memorized forms followed by shift toward sentence production via grammatical encoding. Both of these predictions were supported by the findings in the present study. This suggests that most 3-year-olds have developed an abstract system of tense via setting up the [+tense] parameter and started to incorporate this system to produce tense morphemes through grammatical encoding (Rispoli & Hadley, 2011). The use of grammatical encoding to produce sentences in 3-year-olds is consistent with previous work in which revisions started to occur during sentence production by 27 months of age (Rispoli, Hadley, & Holt, 2009). Revisions presumably reveal that sentences are composed of grammatically encoded equivalent parts. The combination of an influence of TP on our model with the hesitation data suggests that children may start to produce sentences through grammatical encoding before age 3.

The asymmetrical effects of subject types in children with low TP may, however, be accounted for partly by performance limitations (Bloom, 1990). Within this performance account, children are assumed to have abstract knowledge of tense system, but their production of tense morphemes depends on the interaction between the processing capacity of the children and the processing demands of utterances that requires these morphemes. Lexical subjects tend to have more complex phrase structure, exacting more processing demands than pronominal subjects (e.g., [DP [D The] [NP dog]] and [DP [D He]], where DP refers to determiner phrase and NP refers to noun phrase). It is possible that children with low TP scores also have limited processing capacities and are sensitive to small changes in processing demands. Thus, tense morphemes may be omitted more often in utterances with higher processing demands (Leonard et al., 2000). In contrast, children with high TP scores may have relatively larger processing capacity and be less sensitive to changes in processing demands. The difference in processing demands between pronominal and lexical NP subjects is thus not observable in children with high TP scores. Further work in which processing
capacity in very young children is examined would be required to confirm this speculation.

In addition, the success of the GML hypothesis lies partly in the inclusion of a dual-route-processing model in the theoretical framework. However, Becker (2004) indicated that predicate types also affect the production of tense and agreement morphemes. That is, young children were more likely to produce copula be with temporally unbounded predicates (e.g., He’s a boy) than with temporally bounded predicates (e.g., He’s at home) even after the subject types of these sentences were controlled (Guo, Owen, & Tomblin, 2011). Because sentences with different predicate types but identical subject types should be produced via the same route in the GML hypothesis, there is no reason why the production accuracy of copula be should vary in these sentences. Consideration of how tense marking varies with subject and predicate types in the framework of the GML hypothesis is worthy of further study.

TP Scores, but not MLU or FVMC, Predict the Effect of Subject Types on the Production of Auxiliary Is

With regard to supporting the GML hypothesis, it is notable that the TP score was successful in predicting the effect of subject types on the production accuracy of auxiliary is. TP is a relatively new measure introduced by Hadley and colleagues (Hadley & Holt, 2006; Hadley & Short, 2005) in the last 5 years. This study supports the TP score as a measure worth further study despite using a different approach to calculation.

In contrast, neither MLUm nor the FVMC predicted children’s use of auxiliary is with different subject types. One reason that MLUm was not a significant predictor might be that the MLU levels in the current participants were close to or above 4.0, a point at which it no longer reflects morphosyntactic development accurately (Brown, 1973). The other reason could be that MLU does not distinguish between formulaic and directly encoded speech. Thus, increases in MLU do not necessarily reflect children’s development of morphosyntax in general or development of tense and agreement specifically.

Though the FVMC was specifically designed to document children’s development of tense and agreement, it did not predict the use of auxiliary is, which could be because the FVMC overestimates the child’s tense development (Rispoli et al., 2009). Although type-based measures (e.g., the TP score) place the emphasis on productive/nonmemorized forms, token-based measures (e.g., the FVMC) include directly activated or memorized forms, emphasizing general accuracy and communication skills. For instance, one of the children in this study, HP, produced only one correct item each for the pronominal subject and low-frequency lexical NP subject conditions in the experimental task but had an FVMC of 0.87. This child used constructions like That’s + NP and What’s that? frequently in the language sample but did not produce a variety of tense and agreement morphemes in other grammatical contexts. The possibility that the FVMC could overestimate children’s tense development may also account for the lack of correlation between TP scores and the FVMC. For instance, the five children with the lowest TP2 scores all had an FVMC at or above 70%. One of these five children, SF, even had a TP2 score of 1 but had an FVMC of 88%. Researchers of future studies in which the FVMC is used to document children’s developmental level of tense may want to take the memorized forms into account and explore whether the diagnostic capabilities of the FVMC would increase without these forms.

Conclusion

In the present study, we tested how subject types affected the production accuracy of auxiliary is in children with different developmental levels. We observed clear changes in which subject types were facilitative across different levels of TP. The developmental change of the effect of subject types was not found when the predictors were MLUm or the FVMC. These findings have theoretical and methodological implications. Theoretically, these results support the GML hypothesis such that the development of the tense system is gradual and children start to produce tense through grammatical encoding before age 3. Methodologically, the present findings suggest that the use of precise measures of developmental levels, in addition to age, are of importance not only when researchers investigate the effect of frequency on the production of tense and agreement morphemes but also when they undertake other studies of the processes/factors that influence language acquisition.

References


The Role of Developmental Levels in Examining the Effect of Subject Types on the Production of Auxiliary Is in Young English-Speaking Children

Ling-Yu Guo, Amanda J. Owen Van Horne, and J. Bruce Tomblin

*J Speech Lang Hear Res 2011;54;1658-1666*

DOI: 10.1044/1092-4388(2011/10-0140)

The references for this article include 12 HighWire-hosted articles which you can access for free at: [http://jslhr.asha.org/cgi/content/full/54/6/1658#BIBL](http://jslhr.asha.org/cgi/content/full/54/6/1658#BIBL)

This information is current as of December 16, 2011

This article, along with updated information and services, is located on the World Wide Web at: [http://jslhr.asha.org/cgi/content/full/54/6/1658](http://jslhr.asha.org/cgi/content/full/54/6/1658)