SQUIBS AND REPLIES

Copula Omission Is a Grammatical Reflex

Misha Becker
University of North Carolina

1. INTRODUCTION

A well-known characteristic of children’s language is that, in the early stages of language production, children frequently omit grammatical constituents from their speech, where such omissions are not grammatical in the target (adult) grammar. Omitted constituents may be subjects, objects, determiners, auxiliary verbs, verbal inflections, prepositions, and in some cases main verbs (L. Bloom (1970), Brown (1973)). One of the important questions in language acquisition research is whether these omissions arise as a consequence of grammatical constraints (i.e., the child represents a slightly different grammar than the adult does) or as a consequence of processing limitations (the child has the same grammar as the adult but is not able to produce sentences that are as long as adult sentences, and so some constituents are omitted). In this squib, I address this question with respect to children’s omission of the copula be.

Previous research on this topic has focused on explaining children’s omission of subjects in non-pro-drop languages (L. Bloom (1970), P. Bloom (1990), Clahsen (1986), Hyams and Wexler (1993), Valian (1991), among others). P. Bloom (1990) argued that children represent the same grammar as adults but are limited in the length of the sentences they are able to produce, because they have a more limited processing ability than adults. He drew support for his approach from the fact that children’s subjectless sentences contain verb phrases (VPs) that are significantly longer (in number of words) than the VPs in sentences with an overt subject. His claim is that because children’s processing ability does not
match that of adults, children must shorten their utterances. Thus, in sentences
with longer VPs, the subject is sometimes omitted.

Hyams and Wexler (1993) took a different perspective, arguing for a grammar-
based account of children’s Null Subjects. Their argument has both a theoretical
and an empirical component. Speaking to the empirical correlation between VP
length and the absence of subjects, Hyams and Wexler noted that the pattern of
finding longer VPs in Null Subject sentences (compared to shorter VPs in overt
subject sentences) is attested in adult Italian just as in child English. This fact led
them to argue that this pattern does not arise from processing difficulty but rather
from a property of pro-drop languages.

The theoretical component of Hyams and Wexler’s (1993) argument draws on
the observation that cross-linguistically “there is a systematic association between
the child’s use of Null Subjects and various properties of the early inflectional
system” (p. 424). For instance, the production of overt subjects in child German
correlates strongly with the production of overt verbal inflection, and conversely,
the absence of an overt subject correlates with the absence of finiteness marking
on verbs (Clahsen (1986)). Additional evidence from Sano and Hyams (1994)
shows that overtess of subjects in child English correlates with the overtess of
(uncharted) auxiliary be.

In this squib, I evaluate both a grammar-based and a processing-based expla-
nation for children’s omission of the copula be. If children’s omission of the cop-
ula stems from properties of the child’s grammar, one would expect to find some
syntactic contingency on the distribution of children’s copula omissions. For in-
stance, the copula would be overt in clauses that have a particular syntactic prop-
erty but would be omitted in clauses that lack or differ in this syntactic property. If
children’s omission of the copula is due to processing difficulty (thus essentially a
performance error), one would instead expect to find a pattern in which longer ut-
terances tend to lack a copula, whereas shorter utterances contain an overt copula.
That is, one would expect that if children leave out the copula due to high process-
ing load, they would do so predominantly in longer utterances. I argue that in line
with Hyams and Wexler’s (1993) account of Null Subjects, a grammar-based ap-
proach to children’s omission of the copula provides a better account on both the-
oretical and empirical grounds than a processing-based account.

2. GRAMMAR-BASED ACCOUNT

Around the age of 2 years, children acquiring English sometimes produce an overt,
inflected copula, as in (1a,b), and they sometimes omit the copula, as in (1c,d).

(1) a. Patsy’s a girl (Peter, 2;3, from L. Bloom (1970))
b. She’s a crocodile (Naomi, 2;3, from Sachs (1983))
c. I in the kitchen (Nina, 2;1, from Suppes (1974))
d. he way up dere [there] (Adam, 3;0, from Brown (1973))
Based on the data of four children (Adam, Nina, Naomi, and Peter) from the CHILDES database (MacWhinney and Snow (1985)), I argued in Becker (2002) that children’s omission of the copula is a result of grammatical constraints. I observe that the copula tends to be overt in utterances with a nominal predicate (NP), as in (1a,b), and tends to be omitted in utterances with a locative predicate (PP), as in (1c,d). The data are given in Table 1. (In this and subsequent tables, \( N \) represents the total number of [overt- and null-copula] utterances out of which the percentage was calculated, i.e., the denominator.)

For each child, the proportion of nominal predicate utterances with overt be is significantly greater than the proportion of locative utterances with overt be. For Nina, Pearson’s \( \chi^2(1, N = 258) = 95.495, p \leq .001 \); for Peter, \( \chi^2(1, N = 488) = 173.940, p \leq .001 \); for Naomi, \( \chi^2(1, N = 152) = 46.791, p \leq .001 \); and for Adam, \( \chi^2(1, N = 328) = 18.805, p \leq .001 \).

1 The data are based on the following files:

<table>
<thead>
<tr>
<th>Child</th>
<th>Files</th>
<th>Age Range</th>
<th>Mean MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nina</td>
<td>7–13</td>
<td>2;0–2;2</td>
<td>2.98</td>
</tr>
<tr>
<td>Peter</td>
<td>6–11</td>
<td>2;0–2;3</td>
<td>2.84</td>
</tr>
<tr>
<td>Naomi</td>
<td>35–68</td>
<td>2;0–2;7</td>
<td>3.09</td>
</tr>
<tr>
<td>Adam</td>
<td>10, 15, 18, 20, 25, 27, 28</td>
<td>2;7–3;4</td>
<td>3.38</td>
</tr>
</tbody>
</table>

Note. MLU = mean length of utterance.

The first file selected for each child was the first file in which the copula was used overtly in all relevant construction types; the asymmetry found between nominal and locative predicates was consistent throughout the period examined.

2 The percentages and total sample sizes reported here differ slightly from those reported in Becker (2002). There are two reasons for the difference. One is that upon reexamination of the data, minor counting errors were found and corrected; second, the mean was calculated here by dividing the total number of copula occurrences for each child by the total number of possible environments for that child rather than by calculating the rate of copula use in each file and then averaging across the files. The resulting difference is sufficiently small as to not affect the overall pattern or the analysis.
The analysis of this strong asymmetry given in Becker (2002) hinges on a semantic difference between nominal and locative predicates: Nominal predicates typically denote permanent properties (cf. Mommy is a girl), whereas locative predicates typically denote temporary properties (cf. my pen down there). This permanent versus temporary distinction closely parallels the individual-level (IL) versus stage-level (SL) distinction (Carlson (1977)). Some researchers, such as Schmitt (1992) and Kratzer (1995), have argued that the semantic SL–IL distinction corresponds to a syntactic distinction. Following Schmitt, I analyze IL predicates as nonaspectual, whereas SL predicates are analyzed as aspectual. My analysis of the child data is that the copula is a grammatical reflex of the temporal anchoring of a clause containing a nonaspectual predicate (i.e., IL predicates). The temporal anchoring of a clause containing an aspectual predicate (i.e., SL predicates) does not result in an overt copula.

Temporal anchoring is the formal binding relation between a tense operator (T_{op}) in the C-domain (Enç (1987), Guérón and Hoekstra (1995)) and a syntactic tense or aspect node in the main clause. This binding relation serves to anchor the utterance to the discourse and was argued to be obligatory for main clauses in adult grammar (Becker (2002)). In adult English main clauses, the temporal anchoring requirement must be satisfied through T_{op} binding Tns (not Asp), as all main clauses in adult English are tensed. My claim is that in child English, SL predicates, by virtue of having an aspectual projection (AspP), may be temporally anchored by T_{op} binding Asp. A bound Asp node does not, however, yield a tensed clause, and so a copular sentence in which T_{op} binds Asp will be untensed, that is, lack a tensed copula. In other words, SL predicates in child English satisfy the temporal anchoring requirement but are not tensed. IL predicates, on the other hand, have no AspP projection and therefore require T_{op} to bind Tns, yielding a tensed (finite) clause. Copular sentences in which T_{op} binds Tns will contain a tensed copula.

Among adjectival predicates, some denote permanent/IL properties (e.g., tall), and others denote temporary/SL properties (e.g., tired). Thus, it is of interest to know whether the IL–SL distinction made by children between nominal and locative predicates also extends to the adjectival domain. In this data set, adjectives were classified on the basis of the following criteria. Adjectives denoting color, size, aesthetic properties (pretty, ugly); physical properties having to do with hardness, softness, or texture; and other inherent properties were classified as IL. Adjectives denoting temperature, physical sensations (hungry, sick), emotions, and other accidental or temporary properties were classified as SL. Exceptions were made where the context clearly indicated a different meaning (e.g., Naomi’s wait till it’s yellow was classified as SL because there was obviously a color change), but this contextually determined classification was rare (on average, only 5.5% of the children’s adjectival predicatives were classified on a contextual basis, less than 8% for any one child). Children’s rate of producing the copula in SL and IL adjectival predicative utterances is given in Table 2. A paral-
lel trend to the nominal/locative pattern is found in these utterances: For three of the four children, the copula is overt more frequently with IL adjectives than with SL adjectives. (The fourth child, Adam, showed a slight trend in the other direction.) Although only Naomi’s difference in the percentage of overt be with IL versus SL adjectives was significant, \( \chi^2(1, N = 94) = 4.579, p < .05 \), both Nina and Peter showed a trend in the same direction.

The grammar-based account given in Becker (2002) provides an account of the asymmetry in children’s production/omission of the copula in English: The copula is overt and inflected when the predicate is nonaspectual, as temporal anchoring in this case yields a tensed clause; the copula is null when the predicate is aspectual, as temporal anchoring of the clause by binding Asp yields an untensed clause in the grammar of child English. The reader will have noticed that the asymmetry in copula production with aspectual versus nonaspectual predicates is not perfect: The copula is overt less than 100% of the time with nonaspectual predicates, and it is overt more than 0% of the time with aspectual predicates. Such counterexamples would appear to violate the child English grammar. IL predicates may, however, occur without tense marking in other languages (e.g., adult Chinese, adult Russian in present-tense constructions) just as SL predicates may occur with tense marking (e.g., adult English). Thus, there may be other grammatical mechanisms that interact with the dominant constraints in the grammar of child English. For instance, if one thinks of the learning procedure as involving the resolution of a competition between multiple possible grammars (cf. Yang (2002)), 2-year-old English speakers may be in a stage in which they mostly produce utterances generated by the child English grammar proposed here (IL tensed, SL untensed) but sometimes produce utterances conforming to a Chinese-type grammar (null copula even with IL predicates) and sometimes to an English-type grammar (overt, tensed copula even with SL predicates).

There is a further way in which a grammar-based account of children’s copula omission receives support. Some versions of the grammar-based approach hold that, although the child’s grammar is not identical to the adult grammar of the target language, it may contain syntactic properties of an adult grammar of a differ-

<table>
<thead>
<tr>
<th>Child</th>
<th>IL Adjectives</th>
<th>SL Adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>N</td>
</tr>
<tr>
<td>Nina</td>
<td>62.5</td>
<td>24</td>
</tr>
<tr>
<td>Peter</td>
<td>57.1</td>
<td>28</td>
</tr>
<tr>
<td>Naomi</td>
<td>93.5</td>
<td>29</td>
</tr>
<tr>
<td>Adam</td>
<td>37.1</td>
<td>35</td>
</tr>
<tr>
<td>M</td>
<td>62.6</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* IL = individual level; SL = stage level.
ent language (e.g., Hyams (1986), Hyams and Wexler (1993)). Thus, in Hyams’s (1986) view of child English, subjects are omitted because their grammar allows pro-drop, as in adult Italian. In the case of the copula, we find a similar parallel to certain adult languages. Distinguishing IL from SL predicates by using either a different copula (cf. Spanish, Portuguese) or an overt versus null copula (cf. Hebrew) is found cross-linguistically. Where the distinction is one of overt versus null copula, as in Hebrew, the pattern is strikingly similar to that found in child English: When the predicate denotes a permanent or inherent property, the copula is overt; when the predicate denotes a temporary or noninherent property, the copula is (or can be) null (Greenberg (1994)). This fact provides further support for the view that children’s output is constrained by grammatical principles.

3. PROCESSING-BASED ACCOUNT

The grammar-based account provides an adequate account of the data. However, there appears to be a processing-based explanation that is also consistent with the data: Given that PPs generally have one word more than NPs (in the house vs. the house), perhaps children omit the copula in sentences with PP predicates because those utterances are longer than utterances with NP predicates.

The mean length of copular utterances (in words, counting only noncopula words) is given in Table 3. Utterances are separated by predicate type. A log-linear model revealed that there were on average 1.17 times as many (noncopula) words in children’s locatives as in their nominal predicative utterances, and this difference was significant, Wald $\chi^2(1, N = 1,123) = 44.73, p \leq .0001$. There was, however, no main effect of copula overtness on utterance length whether taking predicate type into account, Wald $\chi^2(1) < 1, p = .35$, or not taking it into account, Wald $\chi^2(1) = 1.36, p = .24$. Additionally, there was no significant interaction effect between predicate type and copula overtness, Wald $\chi^2(1) < 1, p = .32$. That is, although locatives contain more noncopula words on average than nominal predicatives, those utterances lacking a copula are not significantly

<table>
<thead>
<tr>
<th>Child</th>
<th>Overt Copula</th>
<th>No Copula</th>
<th>Overt Copula</th>
<th>No Copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nina</td>
<td>3.2</td>
<td>0.9</td>
<td>3.1</td>
<td>0.7</td>
</tr>
<tr>
<td>Peter</td>
<td>3.2</td>
<td>1.0</td>
<td>3.2</td>
<td>0.9</td>
</tr>
<tr>
<td>Naomi</td>
<td>3.3</td>
<td>1.0</td>
<td>2.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Adam</td>
<td>3.5</td>
<td>1.2</td>
<td>3.4</td>
<td>1.3</td>
</tr>
<tr>
<td>M</td>
<td>3.3</td>
<td>3.1</td>
<td>3.7</td>
<td>3.8</td>
</tr>
</tbody>
</table>
longer than those utterances containing an overt copula. The method for determining significance was appropriate for count data with multiple observations within each participant.

Another way to look at this issue is to compare the rate of overt *be* in nominal predicates to the rate of overt *be* in locative predicates holding sentence length constant. If children omit the copula due to processing load (as determined by sentence length) and not because of grammatical factors, then among utterances of equal length (e.g., three words), there should be no difference between nominal and locative predicates in the proportion of overt *be*. Moreover, as sentence length increases (e.g., to four words, five words, and so on), the proportion of overt *be* should decrease in both predicate types. However, as I show in Table 4, omission rates within each sentence length differ significantly as a function of predicate type. Also, in looking across different sentence lengths, the proportion of overt *be* does not decrease at all in nominal predicates, and in locative predicates, it does not decrease as sentences increased from three to five words. (As in Tables 1 and 2, *N* represents the total number of null- and overt-copula utterances of each predicate type, i.e., the denominator.)

For sentences that are two to five words long, the difference between the proportion of overt *be* with nominal versus locative predicates was significant by a logistic regression: two words, \( \chi^2(1, N = 234) = 11.25, p < .01 \); three words, \( \chi^2(1, N = 530) = 112.12, p < .001 \); four words, \( \chi^2(1, N = 323) = 96.14, p < .001 \); five words, \( \chi^2(1, N = 89) = 30.37, p < .001 \). In summary, given a nominal predicate utterance and a locative predicate utterance of equal length, the nominal utterance is very likely to occur with an overt copula, and the locative utterance is very likely to occur without a copula. Examples of children’s utterances are given in (2–9):³

³Children’s use of the contracted copula (e.g., *that’s*) was counted as an overt, tensed copula, equivalent to an uncontracted main clause copula. The justification for this approach is that all four of the children used and omitted the copula in both contractible and uncontractible environments (*that’s, that, this is, this*), and all children contracted the copula on both pronominal/demonstrative subjects (*he’s, that’s*) and on full NP subjects (Georgie’s, Mommy’s). Thus, contracted forms such as *that’s, he’s* are not taken to be unanalyzed forms.

### Table 4

Mean Proportion of Overt *be* With Nominal and Locative Predicates by Sentence Length

<table>
<thead>
<tr>
<th>Sentence Length (in Noncopula Words)</th>
<th>Nominal Predicate</th>
<th>Locative Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>M</em></td>
<td><em>N</em></td>
</tr>
<tr>
<td>Two words</td>
<td>73.8</td>
<td>206</td>
</tr>
<tr>
<td>Three words</td>
<td>75.3</td>
<td>457</td>
</tr>
<tr>
<td>Four words</td>
<td>73.6</td>
<td>216</td>
</tr>
<tr>
<td>Five words</td>
<td>74.1</td>
<td>54</td>
</tr>
<tr>
<td><em>M</em></td>
<td>74.2</td>
<td>21.4</td>
</tr>
</tbody>
</table>

Downloaded By: [Becker, Misha] At: 21:56 3 June 2009
(2) Nominal: Two words
   a. This is lady (Naomi, 2;0)
   b. Dat me (Adam, 3;0)

(3) Locative: Two words
   a. It’s here (Peter, 2;1)
   b. Stove there (Nina, 2;1)

(4) Nominal: Three words
   a. One is a mommy (Nina, 2;1)
   b. That cuckoo fish (Naomi, 2;5)

(5) Locative: Three words
   a. Doggie’s in bag (Peter, 2;0)
   b. Lady on that (Nina, 2;0)

(6) Nominal: Four words
   a. Dose are Donald-Duck safety rules (Adam, 2;11)
   b. That a tiny circle (Peter, 2;2)

(7) Locative: Four words
   a. It’s on my slipper (Naomi, 2;3)
   b. Foot in the water (Nina, 2;0)

(8) Nominal: Five words
   a. But that’s a funny fish (Naomi, 2;5)
   b. Dat my sixty new rambler (Adam, 3;2)

(9) Locative: Five words
   a. One house is way up air (Adam, 3;0)
   b. More wheel on a truck (Peter, 2;1)

The means given in Table 4 and the mean proportion of overt be for nominal sentences that are six words in length are shown in the graph in Figure 1.\(^4\) The data in Table 4 and Figure 1 show that there was a slight decrease in the proportion of overt be in locatives from two- to three-word sentences, suggesting that sentence length may play a partial role in determining copula overtness. A series of logistic regressions modeling the proportion of overt be as a function of length and predicate type (adjusting for multiple observations within each participant) revealed a mild effect of sentence length and a much stronger effect of predicate type than sentence length in predicting copula overtness. These models yielded a

\(^4\)Sentences longer than six words in length for nominal predicates or longer than five words in length for locatives were too few in number for their means to be meaningful.
Z statistic and odds ratio for each of three parameters. Modeling sentence length alone (disregarding predicate type), the odds ratio was 1.29 (Z = 2.52, p = .01). Thus, length alone is a significant factor. Modeling both sentence length and predicate type together, the odds ratio was 1.10 (Z = 2.11, p = .03). Thus, length was still significant but less so, and the effect was very small (the odds ratio was very close to 1.0, which represents the null hypothesis). Modeling predicate type alone, the odds ratio was 13.61 (Z = 7.01, p < .0001). Thus, predicate type alone was highly significant and the effect was very large.\(^5\)

In the case of adjectival predicates, there is typically only one word in the predicate (she is tall), but adverbs and other modifiers may be added. The processing-based account would predict that the copula is omitted in sentences containing more words as compared to sentences with an overt copula. However, this prediction was not borne out. On average, sentences containing an overt copula had about the same number of noncopula words as sentences without a copula, as shown in Table 5.

A log-linear model reveals that none of the effects (main or interaction) were significant. There was no main effect of predicate type (IL vs. SL predicate), Wald $\chi^2(1) < 1, p = .58$; there was no main effect of copula overtness, either taking predicate type into account, Wald $\chi^2(1, N = 417) = 1.47, p = .23$, or not taking it into account, Wald $\chi^2(1) = 2.79, p = .09$; there was no Copula $\times$ Predicate interaction, Wald $\chi^2(1) < 1, p = .94$. The prediction of the processing-based approach,

---

\(^5\)The fact that a small effect of sentence length was found in these models of the data is not inconsistent with the lack of a sentence length effect reported under Table 3. In Table 3, I modeled sentence length as a function of copula overtness (or predicate type; dependent variable was sentence length, independent variables were copula overtness and predicate type). Here, I am modeling copula overtness as a function of sentence length (dependent variable was copula overtness, independent variable was length). Thus, the difference in effect results from switching the dependent and independent variables and using different models.
4. CONCLUSION

The purpose of this squib was to compare the predictions of a processing-based and a grammar-based account of children’s omissions of the copula. The grammar-based account was supported by the data. In particular, a correlation was found between omission of the copula and a semantic feature of the predicate. This sort of correlation mirrors the correlation found cross-linguistically between omission of the subject and other grammatical aspects of the clause (such as the marking of verbal inflection). A further advantage of the grammar-based approach over the processing-based approach is the parallel between the child English grammar and adult grammars of languages other than English.

The processing-based account, on the other hand, was not supported. It was shown that although the copula tends to be omitted in utterances with locative predicates as opposed to nominal predicates and that locative predicates tend to be longer (in number of noncopula words) than nominal predicates, sentence length alone is not a good predictor of copula omission. Predicate type (nominal vs. locative) was shown to have a much stronger effect on copula overtneas than sentence length, making it doubtful that copula omission is a performance error resulting from a processing overload.

ACKNOWLEDGMENTS

I thank Nina Hyams, Carson Schütze, Chris Wiesen, and two anonymous reviewers for helpful comments and discussion. I claim responsibility for all remaining shortcomings.
REFERENCES


