The Incidence of Error in Young Children's *Wh*-Questions

Caroline F. Rowland
University of Liverpool,
Liverpool, United Kingdom

Julian M. Pine
University of Nottingham,
Nottingham, United Kingdom

Elena V. M. Lieven
Max Planck Institute
for Evolutionary Anthropology,
Leipzig, Germany

Anna L. Theakston
University of Manchester,
Manchester, United Kingdom

Many current generativist theorists suggest that young children possess the grammatical principles of inversion required for question formation but make errors because they find it difficult to learn language-specific rules about how inversion applies. The present study analyzed longitudinal spontaneous sampled data from twelve 2–3-year-old English speaking children and the intensive diary data of 1 child (age 2;7 [years;months] to 2;11) in order to test some of these theories. The results indicated significantly different rates of error use across different auxiliaries. In particular, error rates differed across 2 forms of the same auxiliary subtype (e.g., auxiliary *is* vs. *are*), and auxiliary *DO* and modal auxiliaries attracted significantly higher rates of errors of inversion than other auxiliaries. The authors concluded that current generativist theories might have problems explaining the patterning of errors seen in children's questions, which might be more consistent with a constructivist account of development. However, constructivists need to devise more precise predictions in order to fully explain the acquisition of questions.

KEY WORDS: errors, questions, grammar acquisition

The patterning of errors in children's speech has done much to shape both generativist and constructivist theories of the acquisition process. The finding that children frequently omit certain grammatical structures has required researchers in the generativist tradition to look closely at how children's knowledge differs from that of adults. Conversely, the fact that children make mistakes that indicate understanding of the grammatical rules of their language (e.g., past tense overgeneralization errors) is one of the strongest pieces of evidence we have against simple learning accounts based on behaviorist principles.

An area in which error rates have been intensively studied is English *wh*-question acquisition, particularly the acquisition of object and adjunct *wh*-questions. Research on English children's questions seems to suggest surprisingly high error rates in acquisition (e.g., Bellugi, 1965). This finding is inconsistent with the idea that children very quickly map the rules governing question formation in English onto the possibilities in universal grammar (UG) and thus produce very few errors of commission in their early speech. Particular attention has been given to the errors that indicate a problem applying the rules that govern the positioning of tense and agreement (and therefore the auxiliary and copula) in English questions (see, e.g., Bellugi, 1965, 1971; Brown, 1968; Erreich, 1984; Hurford, 1975; Labov & Labov, 1978; Maratsos & Kuczaj, 1978). These studies indicate that children may fail to raise tense and agreement out of the verb phrase (VP) or inflectional phrase (IP) (e.g., *what he does do?, what he does?*), they may omit obligatory tense and agreement markers altogether (e.g., *what he do?*), or they may fail to acknowledge that tense

and agreement should only be marked once (e.g., *what does he does?, what does he does do?*).

In response, many generativist theorists have incorporated a role for high error rates in question acquisition. Some theorists (e.g., Radford, 1990; Vainikka, 1994) have argued that errors will only occur before certain types of grammatical knowledge mature, but these accounts have problems explaining why correct questions co-occur with errors. More successful are full competence accounts that suggest that, although the possibility of inversion is present in UG, children still have to learn language-specific rules of question formation (DeVilliers, 1991; Santelmann, Berk, Austin, Somashekar, & Lust, 2002; Stromswold, 1990; Valian, Lasser, & Mandelbaum, 1992). However, the extent to which these theories can account for the patterning of individual auxiliaries and the copula in children's *wh*-questions is unclear. The aim of the present study was to test these theories against the naturalistic data of thirteen 2–3-year-old children learning English.

Current Theories

The central theme of many current accounts is that the possibility of inversion is present in UG. There are a number of specific theoretical representations, but most are based on the idea that inversion (or movement in some approaches) is a general principle of UG. As such, it is available very early on, which means that children quickly learn that in most object and adjunct *wh*-questions, tense and agreement are marked on the copula or auxiliary, which is placed in the presubject (inverted) position. Children are, thus, capable of producing correct questions from the very beginning of the multiword speech stage, a prediction that has been borne out by the data (e.g., Bellugi, 1965, 1971).

On this view, when children make errors it is not because their grammar lacks inversion but because they have to coordinate innate knowledge with the task of learning when and how inversion applies in their particular language. The nature of the difficulty differs from theory to theory. DeVilliers (1991) and Valian et al. (1992) have suggested that problems arise from the identity of the *wh*-word. DeVilliers argued that children initially analyze *wh*-questions as being in topic position of the inflectional phrase, which means early questions do not involve movement. The production of correct *wh*-questions only becomes possible once the *wh*-word is reanalyzed as being in the specifier position of the complementizer phrase. Inversion comes in "piecemeal" (DeVilliers, p. 171), *wh*-word by *wh*-word, with adjunct *wh*-words (*why* and *how*) occurring with errors for longer than argument *wh*-words (*what, who*). Valian et al. argued that children are applying an optional inversion rule to *wh*-questions, which allows for the production of both inverted and uninverted questions. They suggest that children then learn that inversion is obligatory *wh*-word by *wh*-word. Thus, they too predict that some *wh*-words may attract higher rates of error than others.

Stromswold (1990) and Santelmann et al. (2002) conceptualize the child's problems with questions differently. They argue that although a question-producing child knows that inversion is obligatory, she or he may have difficulties with specific auxiliaries and the copula. In particular, they argue that questions requiring copula *BE*[1] and auxiliary *DO* will attract high rates of errors because the rules governing their positioning in questions are peculiar. The copula undergoes inversion, unlike all other main verbs in English, and, they argue, children struggle to integrate this fact with their knowledge that "main verbs do not raise in their grammar for their language" (Santelmann et al., 2002, p. 837). Forming a question with *DO*-support requires the child to manipulate the inflectional features of the language as well as apply inversion. As Stromswold (1990) stated, *DO*-support is the "jerry-rigged result" (p. 246) of the property of English that requires the raising of tense and agreement but prohibits main verbs from raising. This peculiarity of English is also considered to be difficult to master.

To distinguish between these theories, it is important to analyze the data according to auxiliary subtype as well as *wh*-word. To an extent, these two types of theory predict different things about auxiliary and copula use in *wh*-questions. According to DeVilliers (1991) and Valian et al. (1992), errors pattern according to the identity of the *wh*-word and not the auxiliary, which means their theories predict that all auxiliaries and the copula should attract similar levels of error (see Rowland & Pine, 2000). On the other hand, Stromswold (1990) and Santelmann et al. (2002) predict that copula *BE* and auxiliary *DO* will attract higher rates of error than other auxiliaries because of their peculiar English-specific properties. However, what all of these theories share is the assumption that children are applying English inversion rules, albeit inaccurately, at some level above that of the lexical item, and that errors will pattern accordingly. Even if, within the theory, the child has not yet mastered the rules of inversion for each auxiliary subtype, she or he must know, at the very least, the relationship between two forms of the same auxiliary (e.g., auxiliary *is* and *are*), which means that she or he should know that if inversion applies to one it should apply to the other. There is no scope within these current generativist theories for the lexical form of the auxiliary to determine the structure of the question.

[1]Throughout the article, capital letters are used to refer to the auxiliary subtype (e.g., *DO* refers to all incidences of *do, does, did,* etc.).

This assumption is explicitly stated in Stromswold (1990): "Once they [children] hear a particular auxiliary, they generalize across tense, number, and person within the *BE*, *DO* and *HAVE* subtypes" (p. 20). However, it is also implicit in the other theories. For example, Santelmann et al. (2002) made generalizations about the behavior of a particular auxiliary subtype (e.g., auxiliary *BE*) on the basis of children's performance with only one member of the subtype (e.g., *is*); this idea carries the assumption that all members of the subtype will behave similarly. In addition, one of the central tenets of the full competence approach is that the abstract elements TENSE and AGREEMENT underlie children's use of inflected forms from very early on. This, it is argued, explains why children make "essentially no inflectional errors with auxiliaries" (Stromswold, 1990, p. 53) and why, when children do use inflectional material, they almost always do so correctly (Brown, 1973). If these abstract elements underlie auxiliary use, the implication is that the child has analyzed person and number and that, once she or he starts using two forms of an auxiliary, especially two forms that mark the same tense (e.g., *is* and *are*), she or he will be aware of the relationship between them. Given these facts, there seems to be little scope within these formulations for there to be different levels of correct use across different forms of the same auxiliary subtype (see Wilson, 2003, for similar arguments).

There is some evidence that this assumption may be unfounded. For example, Kuczaj (1986) found that different forms of copula and auxiliary *BE* showed different patterns of acquisition, and Theakston, Lieven, Pine, and Rowland (2005) have demonstrated that some forms of auxiliary *BE* and *HAVE* may be omitted less often than others in children's utterances. If we find in *wh*-question acquisition that errors pattern according to the lexical auxiliary, with some forms of an auxiliary (e.g., *is*) occurring more often correctly than others (e.g., *are*), it is difficult to see how the theories described above can explain the data.

The first aim of the present study was to investigate the patterning of correct use and errors in *wh*-question acquisition. First, we tested the prediction of DeVilliers (1991) and Valian et al. (1992) that there will be similar levels of correct use across all auxiliaries and the copula and across different lexical forms of the same subtype. Second, we tested the predictions of Stromswold (1990) and Santelmann et al. (2002) that, although copula *BE* and auxiliary *DO* should attract higher error rates than other auxiliaries, there should be similar levels of correct use across different lexical forms of the same auxiliary or copula subtype.

If levels of correct use vary between different forms of the same subtype, it is possible that the generativist account could explain the data if the proportions of correct use differed according to *wh*-word. Thus, for example, if auxiliary *is* was used more often with the *wh*-word *what* (in argument questions), and auxiliary *are* most often with the *wh*-word *why* (in adjunct questions), the discrepancy in correct use between these auxiliaries could be explained in terms of differences in use of *wh*-words. The third analysis, therefore, tested whether there were *wh*-word specific differences in the use of particular auxiliary and copula forms.

The analyses proposed so far are designed to compare predictions about the amount of correct use in children's *wh*-questions. However, it could be argued that strong versions of these theories do not make predictions about correct use overall because they allow the possibility of other explanations for auxiliary omission errors in which tense and agreement are unmarked. For example, Valian et al. (1992) specified that it is possible that children lack knowledge that tense must be lexicalized, Santelmann et al. (2002) stated that there may be a productive null auxiliary in early child language, and Stromswold (1990) excluded auxiliary-less questions from some analyses on the basis that they may involve production errors. All of these ideas would provide independent explanations for auxiliary omission.

These constraints should, however, apply equally to all forms of all auxiliaries. Thus, it is unlikely that the inclusion of these additional explanations allows theories to predict different rates of correct use across auxiliary forms. One could perhaps argue that some forms of the auxiliary (e.g., the plural marker) may impose greater constraints on the production mechanism than others (e.g., singular marker) but there is currently little evidence that such production constraints can explain the patterning of auxiliary omission in the data. In particular, Theakston et al. (2005) demonstrated that auxiliary omission cannot be explained in terms of a production constraint on utterance length. In fact, Stromswold dismissed a production constraint explanation of auxiliary omission on the basis that it fails to account for the presence of auxiliary-less questions with tensed main verbs (e.g., *where he goes?*).

However, in order to take account of these potential criticisms, the analyses were repeated only on those *wh*-questions that contained a present auxiliary or copula form and were either correct or contained an error that could be attributed to problems with inversion (double marking errors, such as *what does she does?* and subject-auxiliary/copula inversion errors, such as *what she does do?*). These errors unambiguously reveal how children are applying the rules that govern question formation in English.

Corpora

A major problem associated with analyzing error rates in spontaneous speech is that sample size constraints affect the reliability of results. As traditional sampling techniques result in a sample of approximately 1% to 2% of the child's speech, the chances of producing meaningful rates of errors when analyzing less frequently produced parts of the *wh*-question system are low. In addition, the prevalence of errors that occur for only a short period of time is likely to be seriously underestimated in longitudinal sampled data. To take a hypothetical example, if a child is recorded for 1 hr a week over a 6-month period and an error is made once an hour but only for a 4-week period, only four errors will be sampled. However, this error may have been produced as many as 336 times during that 4-week period, assuming that the child is awake for 12 hr a day. Sampled data will lead us to the erroneous conclusion that a relatively frequently produced error is extremely rare (see Tomasello & Stahl, 2004, for a more detailed consideration of these issues).

On the other hand, rich data sets tend to be restricted to only a few children, and there is the danger of making generalizations to the language acquisition process from characteristics that are merely individual quirks. It is important then to be able to analyze error rates on large numbers of children and on rich data sets together.

An alternative to naturalistic data analysis is the use of experimental data. However, it is very difficult to design successful experimental methods to elicit *wh*-questions from children and equally difficult to gain information about children's acceptance of grammatical errors in comprehension studies (see, e.g., Rowland & Fletcher, 2003). Such studies benefit enormously from information provided by naturalistic data as to the types of utterance that children produce in spontaneous speech. The aim of the present study was to provide such information by comparing the sampled data from 12 children with the data from an intensive diary study of 1 child.

To summarize, the present study recorded the incidence and patterning of errors in *wh*-question acquisition in order to assess the nature of errors in children's early *wh*-questions and test the predictions of some current theories of acquisition. To achieve this, the following questions were addressed:

1. What is the pattern of correct use and error in English children's early *wh*-questions?
2. Does the rate of correct use differ according to auxiliary/copula subtype and lexical form?
3. Does the rate of inversion error differ according to auxiliary/copula subtype and lexical form?

Method

Manchester Corpus

Participants

The participants were 12 children who took part in a longitudinal study of development. Six were from Nottingham, England, and 6 were from Manchester, England. The children were recruited through local nurseries, doctors' surgeries, and newspaper advertisements. Children were only included if they were deemed to be typically developing British English language learners. Three criteria were used in this decision. First, the language level of potential participants was assessed through the MacArthur Communicative Development Inventory (CDI, Toddlers; Fenson et al., 1993) and a screening audiotape recording of 15 min taken during an initial screening visit. Only children with a vocabulary of approximately 100–300 words, as measured by the CDI, and a mean length of utterance (MLU) of between 1 and 2 morphemes, according to the screening tape, were included as participants. Second, children were only recruited if they had no history of language or cognitive difficulty and were not born prematurely. Third, only children between approximately 1;8 (years; months) and 2;2 were recruited. Table 1 demonstrates the vocabulary scores calculated from the CDI and the MLU calculated from the screening tape for the 12 participants.

One participant, Ruth, did not fit all the criteria. Ruth's vocabulary as measured by the CDI was only 44 words, which was below the 10th percentile reported in the CDI norming study (Fenson et al., 1993). However, it was felt that the score might not accurately reflect Ruth's language level. Her MLU as measured by the screening tape was well within our required range (1.43) and she had no history of language or cognitive difficulty. We concluded that the score might be a result of underreporting of vocabulary by Ruth's mother. This decision was later confirmed by the fact that Ruth seemed to be developing typically compared to the other Manchester corpus children in terms of her MLU range over the study (1.41–3.35).

All the children were monolingual, English-speaking, firstborn children whose mothers were the primary caregivers. No formal information about socioeconomic status was recorded. Ages ranged from 1;8.22 to 2;0.25 at the start and 2;8.15 to 3;0.10 at the end of the study (see Table 1). The MLU of the first and last transcript for each child was calculated using the MLU function of the CLAN program (MacWhinney, 2000). MLU was calculated on all utterances produced in a transcript. Bound morphemes were marked on the main line of the transcripts to ensure that the MLU program counted morphemes rather than whole words

Table 1. Participant information.

Child	CDI score	MLU from screening tape	Age range	MLU range	Total no. *wh*-questions
Anne	180	1.47	1;10.7–2;9.10	1.61–3.46	619
Aran	153	1.47	1;11.12–2;10.28	1.41–3.84	395
Becky	138	1.24	2;0.7–2;11.15	1.46–3.24	1,040
Carl	187	2.50	1;8.22–2;8.15	2.17–3.93	770
Dominic	153	1.25	1;10.24–2;10.16	1.20–2.85	203
Gail	262	1.48	1;11.27–2;11.12	1.76–3.42	495
Joel	122	1.13	1;11.1–2;10.11	1.33–3.32	351
John	191	2.12	1;11.15–2;10.24	2.22–2.93	177
Liz	359	Recording failed	1;11.9–2;10.18	1.35–4.12	447
Nicole	102	1.14	2;0.25–3;0.10	1.06–3.26	304
Ruth	44	1.43	1;11.15–2;11.21	1.41–3.35	201
Warren	124	1.62	1;10.06–2;9.20	2.01–4.12	316
M	167.92	1.53	—	1.58–3.49	443.17
Lara	—	—	2;7.21–2;11.14	MLU at start = 3.39	3,062

Note. CDI = MacArthur Communicative Development Inventory; MLU = mean length of utterance.

(see below for a description of which morphemes were marked and included in MLU counts).[2] Imitated and repeated utterances, utterance fragments, and routines were coded on the main line and excluded from the MLU count. The MLU of each sample ranged from 1.06 to 2.22 at the beginning and 2.85 to 4.12 at the end of the study (see Table 1). The corpus is available on the CHILDES database (http://childes.psy.cmu.edu; MacWhinney, 2000) and is referred to as the Manchester corpus (Theakston, Lieven, Pine, & Rowland, 2001).

Procedure

The 12 children were audio-recorded by an investigator in their homes for 2 separate hours every 3 weeks for a year. Taping started as soon as possible after the screening tests were completed. A Marantz CP430 audio recorder with an external microphone (Marantz stereo microphone EM-8) was used for the recording.

During recording, the children engaged in everyday play activities with their mothers. Each hour-long recording session was divided into two separate sessions of 30 min in which mother and child interacted. The first 30-min session consisted of free play in which mother and child engaged in normal play activities. The second 30-min session took place after a break and consisted of structured play activities in which mothers were asked to play with a set of toys provided by the investigator. The production of new toys was aimed at stimulating the children to play for longer. Children were not restricted to the toys provided but were encouraged to play with them.

During all recording sessions, mothers were asked to turn televisions and radios off. For some of the sessions, younger siblings were present. However, these children were all preverbal infants who had little effect on the dyadic nature of the interaction. During all sessions, the investigator attempted to remain in the background as far as possible to enable contextual notes to be taken.

The primary two investigators were Caroline Rowland, who collected the data from the 6 children based in Nottingham, and Anna Theakston, who collected the data from the 6 children from Manchester. The same investigator was present during all recordings with each child except for one 2-month period during which the sessions with the Nottingham children (Anne, Becky, Dominic, Gail, Joel, and Nicole) were conducted by a third investigator. The third investigator was a graduate research assistant who was trained in data collection and transcription by Caroline Rowland.

Transcription

The data were orthographically transcribed using the CHILDES system. The investigator who had been present at the recording conducted all transcription. The Nottingham data that had been collected by a third investigator were transcribed by this investigator and checked by Caroline Rowland. Only child speech and adult child-directed speech were transcribed unless a child utterance was produced in response to adult-directed speech.

To ensure transcription accuracy, the four authors agreed on a set of transcription and coding guidelines before the start of the study and all three investigators

[2] At the time of writing, main line morphemization and morphological coding has been removed from the Manchester corpus available on CHILDES in order to make it compatible with the new Unicode format. The analyses presented here were conducted on the original transcripts that were transcribed and coded by the authors.

were given training and practice in transcription. On the main line, regular forms of plurals, possessives, progressive *ing,* perfective *have,* past tense *ed,* and third person singular main verbs were marked for morphemization according to the CHAT convention. Contracted forms of auxiliaries, the copula, and negation markers were also morphemized on the main line (e.g., *I do-n't like it, I-'ve got it*) and some were transcribed in full in order to distinguish between phonologically identical forms (e.g., *he-'is/he-'has*). Postcodes were used on the main line to mark utterances that were incomplete, routines, imitations, or repetitions. Utterances were considered repetitions or imitations if they were partial or complete repetitions or imitations of an utterance that had occurred five or fewer speaker utterances earlier, unless that utterance had been over 10 s removed in time. A morphological coding dependent tier was also added for each utterance, which provided information about the syntactic category of each morpheme produced. The tier was generated by the MOR program, and disambiguated and checked for accuracy by the transcribers. More detail can be found in the programs and database manuals on the CHILDES Web page.

Initially, four transcripts were independently transcribed by each of the two primary investigators and compared to check for accuracy. In transcription reliability, we checked not only for the accuracy of word identification but also for the consistency of the application of CHAT conventions (e.g., transcription of errors and bound morphology, application of postcodes and other CHAT symbols, coding on the morphological coding dependent tier). The level of agreement was 85%. Subsequently, we checked approximately 1 in 20 transcripts to ensure continuing levels of agreement between the transcribers. In total, 28 transcripts were checked. The level of agreement remained constant at approximately 85%.

Diary Data—Lara

Participant

Lara was the firstborn monolingual English daughter of two White university graduates and was born and brought up in Nottinghamshire, England. Her age ranged from 2;7.21 at the beginning to 2;11.14 at the end of the study.

Procedure

The data used in this article are part of that collected for a larger study on Lara's language acquisition between the ages of 1;9 and 3;3. The data consist of a diary record of the *wh*-questions that Lara produced during the 4 months between 2;7.21 and 2;11.14. The diary began when caregivers informally reported that Lara was starting to produce a variety of *wh*-questions with different auxiliaries and ended when approximately 90% of her *wh*-questions were correct. The diary was filled in by her caregivers (parents and grandparents), who were provided with notebooks to record all *wh*-questions produced both within and outside the home. The diary keepers were trained to record the exact speech of the child (e.g., to omit auxiliaries when not pronounced, to indicate contractions) and to recognize the different types of *wh*-questions. Training started with the provision of verbal and written information about the types of *wh*-questions produced by English learning children and the difference between *wh*-questions and yes–no questions. The information included examples of children's questions taken from the Manchester corpus. The caregivers were also trained to recognize the types of error made by young children and to make careful note if the utterance was an error. Training in error marking consisted of diary keepers and Caroline Rowland together coding 100 different errors extracted from the transcripts of the Manchester corpus. Caroline Rowland then monitored the diary keepers' accuracy during their first day of interaction with the child, explaining and correcting any errors in record keeping. Diary keepers were asked to mark on the diary whether they were unclear about the exact form of the question. All such questions were excluded from the analysis. As no notes were made when the child was at nursery (for parts of 2 days a week), it is estimated that the diary contains approximately 80% of the *wh*-questions that were produced by Lara during this period.

Transcription

The data from the diary were then retranscribed into CHILDES format onto a computer by Caroline Rowland. The transcription conventions were identical to those used for the Manchester corpus children. Because the diary comprised a written, rather than oral, record, no reliability check was possible.

Speech Corpora

All spontaneous, complete, matrix object and adjunct *wh*-questions were extracted from all the children's data. We excluded partially intelligible or incomplete utterances, utterances with parts marked as unclear, quoted utterances, and routines (e.g., counting, nursery rhymes and songs). Full or partial repetitions or imitations were also excluded. Subject *wh*-questions (questions that do not require inversion), embedded *wh*-questions, and fragments (e.g., *what cow?* as a response to *look at that cow*) were excluded.

The decision was made to include questions with contracted auxiliaries. A number of researchers, including Stromswold (1990), have argued that utterances with contracted auxiliaries should be removed from analyses because they are potentially rote learned forms (see also Pinker, 1984). However, these questions are extremely frequent in children's questions (see Rowland, 2000) and an explanation of their presence and the pattern of their acquisition needs to be included in theories of question acquisition, even if they are considered rote learned forms within the theory. We return to this issue in the discussion.

Manchester Corpus

Up to 34 one-hour transcripts were available for each child. There were some missing transcripts—Aran's 14th transcript, Carl's 14th transcript (1/2 hr only available) and 24th transcript, John's 15th and 16th transcript, Ruth's 4th transcript, and Warren's 3rd transcript (1/2 hr only available). MLU was calculated per 1-hr taping session from all complete spontaneous utterances produced by the child in that session using the MLU function of the CLAN program. The data were then divided into stages according to Brown's (1973) criteria. At Stage I, MLU ranged from 1.00 to 1.99; at Stage II, MLU ranged from 2.00 to 2.49; and at Stage III, MLU ranged from 2.50 to 2.99. Transcripts for which the MLU was 3.00 or above were placed in Stage IV. A child was regarded as moving to the next stage of development when three consecutive transcripts had MLUs over the MLU boundary, in which case all three transcripts would be placed in the upper MLU stage.

Lara

The data from Lara could not be matched to the Manchester corpus in terms of traditional MLU measures because transcripts of Lara's speech from the data collection period were not available. Instead, the mean MLU of the *wh*-questions produced by the Manchester children in each stage was calculated and compared to a *wh*-question MLU calculated for Lara's data. Although the overall Stage IV MLU of Lara's questions was higher than the mean MLU of the Manchester corpus children's Stage IV data (Lara = 4.48, Manchester mean = 4.09), it was within the MLU range (3.57–4.54). A separate MLU count taken from one 30-min naturalistic data sample recorded at the beginning of the data collection period at age 2;7.25 was 3.39, confirming that Lara's data were from Stage IV.

The corpus was divided into eight data points. Each data point roughly corresponded to just over a 2-week period.

Error Coding

All *wh*-questions produced by all 13 children were then coded by Caroline Rowland as explained below.

Correct Questions

For questions with auxiliaries, the choice and placement of *wh*-word, auxiliary, main verb, and subject had to be correct. For copula questions, the choice and placement of *wh*-word, copula, and subject had to be correct. Questions with omissions and errors not pertinent to the grammatical rules that apply specifically to questions (e.g., determiner omission) were included.

Omission

1. *Auxiliary/copula omission*: Errors where the auxiliary/copula was omitted and tense was not overtly marked on the main verb (e.g., *where he going?, where he go?, where that?*).
2. *Subject omission*: Errors with omitted subjects (e.g., *where's going?*).
3. *Subject + auxiliary/copula omission*: Questions with auxiliary/copula and subject omitted (e.g., *where going?*).

Errors of Inversion

1. *Double marking errors*: These errors included doubling of the auxiliary/copula (e.g., *where does he does go?*), errors in which tense and agreement were correct but were marked on both auxiliary and main verb (e.g., *where does he goes?*), and errors in which an auxiliary was present but tense and agreement were marked only on the main verb (e.g., *where do he goes?*).
2. *Raising errors*: Errors in which the auxiliary was omitted and tense and/or agreement remained on the main verb (e.g., *where he goes?*). These errors were coded as inversion errors as they indicate that the child has failed to raise tense and agreement.
3. *Noninversion errors*: Subject auxiliary/copula inversion error (e.g., *where he does go?*).

Other Errors of Commission

1. *Agreement errors*: Errors in which an auxiliary/copula was present but did not agree with the subject (e.g., *where does you go?, where do he go?*).
2. *Case errors*: Errors in which the subject had incorrect non-nominative case (e.g., *where's her going?*).
3. *Unclassifiable*: Errors in which it was impossible to determine what mistake had been made. For example, the question *why is the doctor make your tummy better?* would be coded as unclassifiable as it

is unclear whether the target is a progressive (*is making*) or present tense (*does make*) form.

In cases of multiple errors, the following decisions were made:

1. Double marking errors with two different auxiliaries were included under both auxiliaries and counted in each auxiliary's error count (e.g., *what does he is doing?*). There were a very small number of such errors.
2. Questions with both an error of commission and an error of omission (e.g., *where me go?*) were coded as commission errors because the error could not be attributed solely to the omission of an element.

One of the diary keepers was trained in error coding by Caroline Rowland and independently coded all the *wh*-questions produced by 1 of the children. The level of agreement between coders was 97.5%.

Results

Overall Error Rates

Figures 1 and 2 demonstrate the percentage of *wh*-questions that were correct, errors of omission, and errors of commission in the Manchester corpus and Lara's data. Figure 1 shows the mean percentage of correct use and error across the four sampled stages (I–IV) of the Manchester corpus. The data revealed a clear trend toward increasing correct use over time and demonstrate that the most frequent type of error was auxiliary or copula omission. By the end of the sample at Stage IV, correct questions accounted for 68% of questions, with omission errors accounting for 24% of the questions produced. The high proportion of correct questions at Stage I may be due to the fact that a small number of questions were produced by the children at this stage compared with the later stages (Stage I mean number of questions = 41.11, range = 4–109), which meant that a small number of frequently repeated forms such as *what's that?* contributed disproportionately to the calculation.

There was a slight increase in the number of errors of commission produced over the four stages but these (inversion + other commission errors) never accounted for more than a mean of 7.16% of the children's questions at any stage, and errors of inversion never accounted for more than 1.38% of the questions produced at any stage.

Figure 2 demonstrates the percentage of *wh*-questions overall that were correct questions and errors in Lara's data (Stage IV data, divided into eight substages of approximately 2 weeks duration). The results were consistent with the Stage IV data from the Manchester corpus. Overall, correct questions accounted for 67% of questions, mirroring the 68% correct at Stage IV for the Manchester corpus. There was also a trend toward greater correct use through the stage; mirroring the trend across Stages I–IV for the Manchester corpus data. Errors of omission accounted for 24% of the data, corresponding to the mean for the Manchester corpus of 24% for Stage IV, and showed the same trend of decline. Over Stage IV, errors of commission accounted for 7.64% of the questions produced and errors of inversion only accounted for 2.35%. Thus, the more complete data from Lara confirmed the pattern demonstrated by the sampled Manchester corpus data.

Auxiliary-Specific Error Rates

The second set of analyses investigated the level at which children produced correct questions with different forms of the same auxiliary or copula form. The percentage of correct questions for different forms of copula *BE*, progressive auxiliary *BE*, and perfective auxiliary *HAVE* was calculated. In cases of questions with an omitted auxiliary/copula but a present subject, the identity of the missing form was clear from the context (e.g., *where you going* has an omitted *are*). In cases of subject + auxiliary/copula omission, the question was excluded from the analysis because the identity of the omitted auxiliary/copula form could not be ascertained. Only copula and auxiliary *is* and *are* and auxiliary *has* and *have* were included, because other forms (e.g., *am, was*) occurred only rarely. The mean percentage of correct questions produced for each auxiliary/copula form across the 12 Manchester corpus children and the percentages for Lara's data are presented in Figures 3 and 4, respectively. In each case, the denominator for the percentage calculation for each child was the number of *wh*-questions produced that required the particular auxiliary under consideration (i.e., total number of obligatory contexts). For example, the percentage correct use of copula *is* was calculated by dividing the number of correct *wh*-questions with copula *is* by the number of *wh*-questions produced that required copula *is*.

For both corpora, correct questions accounted for a substantial proportion of questions requiring copula *is* from the start of data collection (see Figures 3 and 4). Questions with copula *are* were not produced correctly in large numbers until Stage IV in the Manchester corpus data and the final substages of Stage IV in Lara's data. Even by Stage IV, correct questions only accounted for a mean of 50% of the questions that required copula *are* in the Manchester corpus data. In Lara's data, they accounted for less than 50% of the questions that required copula *are* until Substage 5.

Figure 1. Manchester corpus: mean percentage of *wh*-questions that were correct and errors.

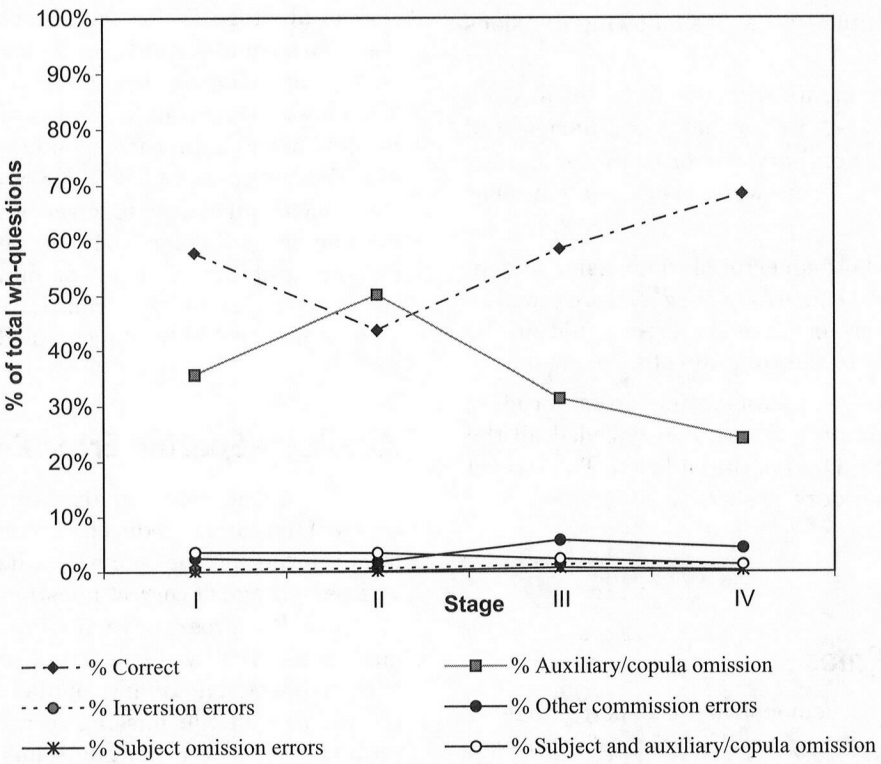


Figure 2. Lara's data: percentage of *wh*-questions that were correct and errors.

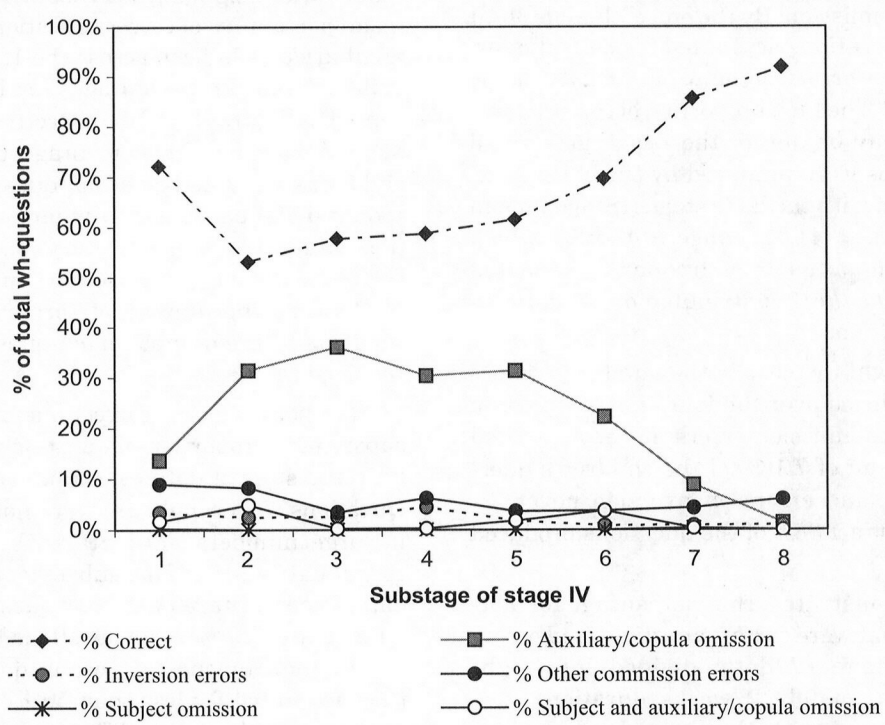


Figure 3. Manchester corpus: mean percentage correct use for copula *is* and *are*, auxiliary *is* and *are*, and auxiliary *has* and *have*.

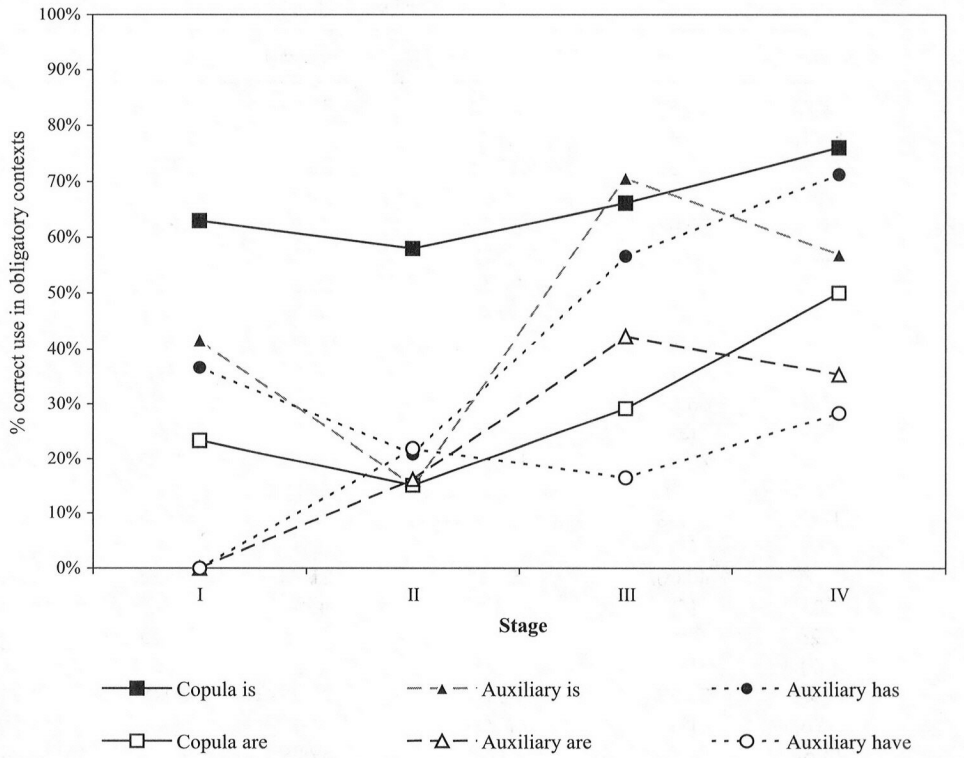


Figure 4. Lara's data: percentage correct use for copula *is* and *are*, auxiliary *is* and *are*, and auxiliary *has* and *have*.

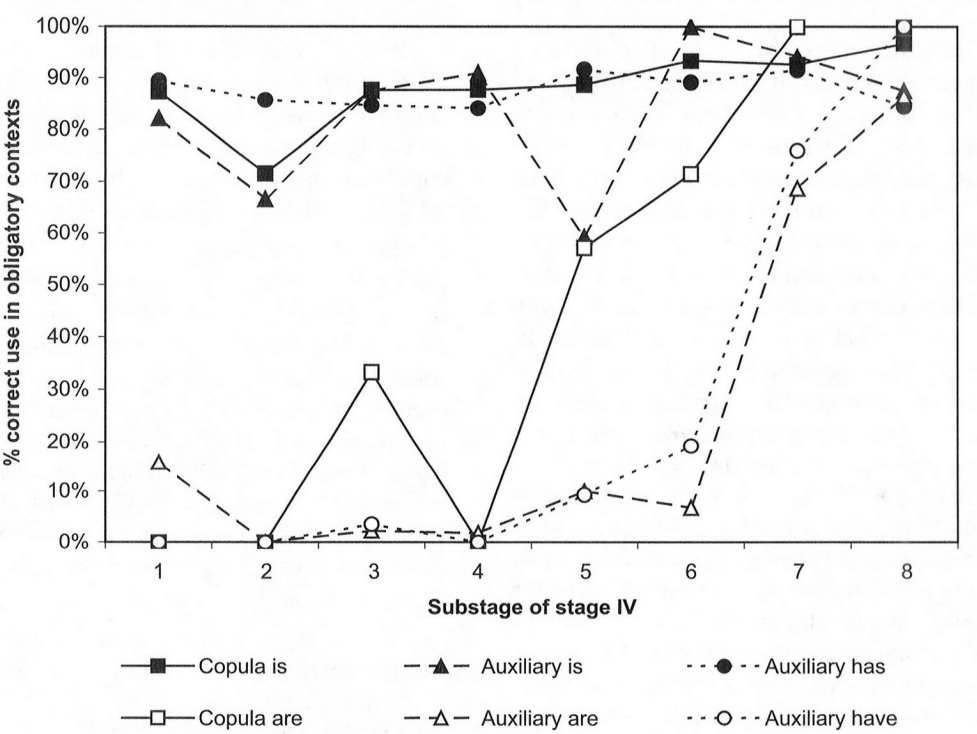


Table 2. Number of questions requiring each auxiliary (Aux.) and copula (Cop.) form (*N*) and percentage correct use: Stages III and IV combined—Manchester corpus.

	% correct use (*N*)					
Child	**Cop. *is***	**Cop. *are***	**Aux. *is***	**Aux. *are***	**Aux. *has***	**Aux. *have***
Anne	74.73	65.52	62.50	72.73	73.33	60.00
	(281)	(29)	(16)	(11)	(45)	(5)
Aran	73.50	28.00	55.56	16.67	59.26	20.00
	(200)	(25)	(18)	(6)	(27)	(5)
Becky	85.26	60.87	87.50	70.83	76.09	60.87
	(441)	(23)	(32)	(72)	(46)	(23)
Carl	87.77	25.00	84.67	30.00	75.76	25.00
	(188)	(4)	(13)	(10)	(33)	(4)
Dominic	87.80	0.00	No aux. *is* contexts	100.00	85.71	0.00
	(41)	(6)		(2)	(7)	(1)
Gail	85.89	66.67	72.22	50.00	86.27	0.00
	(241)	(3)	(18)	(6)	(51)	(3)
Joel	86.50	70.83	100.00	54.55	79.17	37.50
	(200)	(24)	(7)	(11)	(24)	(8)
John	81.08	33.33	No aux. *is* contexts	0.00	100.00	0.00
	(37)	(3)		(1)	(6)	(1)
Liz	81.77	40.00	62.07	21.74	95.65	33.33
	(203)	(5)	(29)	(23)	(46)	(6)
Ruth	4.26	0.00	4.17	0.00	0.00	0.00
	(47)	(1)	(24)	(10)	(4)	(2)
Warren	72.45	33.33	75.00	22.22	50.00	25.00
	(196)	(9)	(16)	(9)	(40)	(4)
M	74.64	38.50	67.08	39.89	71.02	23.79
	(189)	(12)	(19)	(15)	(30)	(6)

Note. *N* = the number of *wh*-question contexts requiring the auxiliary/copula form; this was used as the denominator for the calculation of percentage correct use.

For the Manchester corpus data, too few children produced sufficient questions at each stage to look at whether the developmental differences were statistically significant (i.e., if there was an interaction between auxiliary and stage). Only 3 children produced both questions that required copula *is* and those that required copula *are* in Stage I, and only 8 children produced both types of question in Stage II. However, it was possible to conduct statistical tests to investigate the differences in correct use across auxiliaries in Stages III and IV. Because of very small numbers of copula *are* contexts in Stage III (number of children producing questions requiring copula *are* = 10, mean number of questions requiring copula *are* = 6.90, *SD* = 9.01), data from Stages III and IV were combined to increase the reliability of the analysis. Stage I and II data were not combined with Stage III and IV data in order to minimize any differences in auxiliary use that could be attributed to development. Table 2 demonstrates the total number of *wh*-questions produced that required copula *is* or *are* and the percentage of these questions that were correct. One child, Nicole, was excluded because she produced no Stage III or IV data.

For all analyses, differences were considered significant if $p < .05$. There was a clear distinction between the two forms, with questions with copula *is* produced correctly significantly more often than questions with copula *are* across the 11 children, $F(1, 10) = 23.07$, $p = .001$, $\eta_p^2 = .70$ (see Table 2 for means).

For the purposes of statistical comparison, Lara's data were split into four monthly periods, each incorporating two of the substages noted above (see Table 3). For the first 3 months there was a clear division between the forms. Questions requiring copula *is* were produced correctly significantly more often than questions requiring copula *are* in the first 3 months (for all comparisons, Fisher's exact $p < .0008$; odds ratios (ORs): Month 1 = not calculated[3]; Month 2 = 49.98, Month 3 = 6.03). This division remained significant until the final month, when there was a rapid increase in the rate of correct use of copula *are*.

The data on auxiliary *is* and *are* demonstrated similar trends to those for copula *is* and *are* (see Figure 3).

[3]OR was not calculated because the number of correct questions with copula *are* = 0.

Table 3. Number of questions requiring each auxiliary (Aux.) and copula (Cop.) form (*N*) and percentage correct use—Lara corpus.

Month (substage)	% correct use (*N*)					
	Cop. *is*	Cop. *are*	Aux. *is*	Aux. *are*	Aux. *has*	Aux. *have*
Month 1	81.78	0.00	77.50	7.89	88.12	0.00
(1–2)	(269)	(10)	(40)	(38)	(101)	(24)
Month 2	87.71	12.50	89.80	2.03	84.29	1.33
(3–4)	(301)	(8)	(49)	(197)	(261)	(75)
Month 3	90.75	61.90	79.25	8.94	90.85	13.04
(5–6)	(335)	(21)	(53)	(179)	(153)	(69)
Month 4	93.38	100.00	92.00	75.00	88.04	85.00
(7–8)	(151)	(12)	(50)	(104)	(92)	(40)

Note. *N* = the number of *wh*-question contexts produced that required the copula/auxiliary form; this was used as the denominator of the percentage calculation.

In the Manchester corpus data, the majority of questions requiring auxiliary *is* and *are* were errors at Stages I and II. However, the proportion of correct questions with auxiliary *is* climbed steeply at Stages III and IV; 71% and 57% of questions requiring auxiliary *is* at Stages III and IV, respectively, were correct questions. There were much lower levels of correct use across the board with auxiliary *are*. Caution must be taken in interpreting these figures, though, as only 1 child produced both auxiliary *is* and *are* contexts at Stage I, and only 5 children produced both contexts at Stage II.

Lara's data demonstrated a similar division between auxiliary *is* and *are* in the proportion of correct use (see Figure 4). Correct questions always accounted for more than 50% of questions requiring auxiliary *is*. Correct questions with auxiliary *are* never accounted for more than 16% of questions requiring that auxiliary until the final two substages of Stage IV.

Once again, there were too few Stage I, II, and III data to investigate stage and auxiliary interactions in the Manchester corpus (1 child produced both contexts at Stage I, 5 children at Stage II, and 7 at Stage III). Statistical comparisons on the Stage III and IV Manchester corpus data (see Table 2) revealed that the two forms of auxiliary *BE* demonstrated significantly different rates of correct use across Stages III and IV combined, $F(1, 8) = 15.32$, $p = .004$, $\eta_p^2 = .66$. For Lara's data the differences in correct use were significant for all 4 months (Fisher's exact $p < .0001$ for first 3 months, $p = .02$ for Month 4; ORs: Month 1 = 40.19, Month 2 = 424.60, Month 3 = 38.90, Month 4 = 3.83; see Table 3).

There was also a difference between auxiliary *has* and *have* (see Figures 3 and 4). A pattern of increasing correct use with age emerged with auxiliary *has,* with the Manchester corpus children starting to produce *has* questions correctly more than 50% of the time at Stage III, and Lara producing a substantial number of correct questions all the way through Stage IV. Auxiliary *have* was omitted substantially more often; there was very little increase in the percentage of correct use in the Manchester corpus data and the first six substages of Lara's data. Again, caution must be applied when interpreting data from the early stages, as only 2 children from the Manchester corpus produced contexts requiring both *has* and *have* at Stage I, 5 children at Stage II, and 6 at Stage III. The difference in the percentage of correct use at Stages III and IV combined was significant in the Manchester corpus data (see Table 2 for means), $F(1, 10) = 22.26$, $p = .001$, $\eta_p^2 = .69$. For Lara's data, the difference in the percentage of correct use was significant for the first 3 months of Stage IV (see Table 3; all Fisher's exact $p < .0001$; ORs: Month 1 = not calculated, Month 2 = 397.07, Month 3 = 66.19) but not for Month 4.

To summarize, there were significant differences across the two forms of the auxiliary subtypes copula *BE*, auxiliary *BE*, and auxiliary *HAVE*. Copula *is,* auxiliary *is,* and auxiliary *has* occurred correctly placed in *wh*-questions significantly more often than copula *are,* auxiliary *are,* and auxiliary *have*. Overall, the pattern of data seems inconsistent with the predictions of the theories discussed. Nevertheless, it is possible that the data could be explained by the generativist accounts if the level of correct use differed according to *wh*-word. For example, if auxiliary *is* was used most often with the *wh*-word *what* and auxiliary *are* most often with the *wh*-word *why,* the differences between these auxiliaries could be explained in terms of differences in use of *wh*-words.

Overall, there were not enough data to compare correct use across *wh*-words. For example, only 3 children produced *wh*-questions that required copula *is* and *are* with adjunct words (*how* and *why*), and only 2 produced *wh*-questions that require auxiliary *is/are*

Table 4. Percentage correct use with each *wh*-word and copula (Cop.)/auxiliary (Aux.) form—Manchester corpus Stages III and IV combined.

	% correct use with *what*						% correct use with *where*					
Child	Cop. *is*	Cop *are*	Aux. *is*	Aux. *are*	Aux. *has*	Aux. *have*	Cop. *is*	Cop *are*	Aux. *is*	Aux. *are*	Aux. *has*	Aux. *have*
Anne	91.07	50.00	70.00	75.00	100.00	100.00	69.30	57.14	50.00	50.00	75.00	50.00
Aran	71.43	18.18	66.67	0.00	63.64	0.00	72.00	20.00	33.33		46.15	0.00
Becky	92.00	100.00	88.89	75.00	57.14	42.86	79.79	44.44	100.00	83.33	77.78	66.67
Carl	96.39		66.67	50.00	50.00	100.00	81.55	25.00	100.00	0.00	77.42	0.00
Dominic	93.33			100.00	100.00		70.00	0.00			83.33	
Gail	90.63		70.59	60.00	71.43	0.00	82.98	66.67	100.00		88.64	
Joel	91.25	62.50	100.00	55.56	40.00	28.57	84.31	0.00	100.00	0.00	89.47	100.00
John	100.00			0.00			72.00	33.33			100.00	0.00
Liz	90.22	100.00	69.57	27.78	100.00	0.00	74.31	33.33	33.33	0.00	95.45	50.00
Ruth	5.26		4.35	0.00	0.00		3.85	0.00	0.00			0.00
Warren	90.74	75.00	78.57	12.50	50.00	0.00	63.43	0.00	66.67	100.00	50.00	
M	82.94	67.61	68.37	41.44	63.22	33.93	68.50	25.45	64.81	38.89	78.32	33.33
Mean no. obligatory contexts (*N*)	73.64	6.17	14.89	10.73	4.40	4.00	107.09	5.36	3.67	3.00	27.70	1.89

Note. Empty cells indicate that no questions were produced.

and *has/have* with adjunct words. It was possible, however, to compare the use of the auxiliaries within the *wh*-words *what* and *where*. If the *wh*-word specific explanation is correct, then the differences between the auxiliaries reported above should disappear when analyses are performed within particular *wh*-words.

Table 4 demonstrates the percentage of correct use in obligatory contexts for copula *is/are,* auxiliary *is/are,* and auxiliary *has/have,* broken down by *wh*-word (*what* and *where*) for the Manchester children in Stages III and IV combined. These results must be considered cautiously as only a few contexts were produced for some of the auxiliary forms (e.g., mean number of obligatory contexts requiring *what* and copula *are* = 6.17) and many children did not produce any obligatory contexts for some forms (e.g., only 6 children produced *what* + copula *are* contexts). All the trends were in the direction predicted by the analyses above, with copula *is* produced correctly more often than copula *are,* auxiliary *is* produced correctly more often than auxiliary *are,* and auxiliary *has* correctly more often than auxiliary *have* for both *wh*-words. However, not all differences reached significance. For *what*, auxiliary *is* was produced correctly significantly more often then auxiliary *are*, $F(1, 8) = 10.54$, $p = .01$ $\eta_p^2 = .57$, but no other difference reached significance. For *where,* copula *is* was produced correctly significantly more often than copula *are,* $F(1, 10) = 31.85$, $p < .001$, $\eta_p^2 = .76$, and auxiliary *has* was produced correctly significantly more often than auxiliary *have,* $F(1, 6) = 8.60$, $p = .03$, $\eta_p^2 = .59$. The difference between auxiliary *is* and *are* failed to reach significance.

The analyses were repeated on the more substantial Lara data from the period during which the biggest differences between auxiliary forms were found (Months 1 and 2, Substages 1–4). Table 5 shows the percentage of correct questions produced for each *wh*-word and auxiliary and the results from Fisher's exact tests. Once again, there were very few contexts produced for many of the auxiliaries. However, despite this, for questions with both *what* and *where,* copula *is,* auxiliary *is,* and auxiliary *has* occurred correctly in *wh*-questions significantly more often than copula *are,* auxiliary *are,* and auxiliary *have*. There is some evidence that significant differences between the correct provision of copula and auxiliary *is* and *are* and auxiliary *has* and *have* remain even within *wh*-word.

Analyses by Auxiliary Subtype

In the third set of analyses, we investigated the suggestion that copula *BE* will attract higher rates of error than other auxiliaries. Table 6 demonstrates the mean percentage correct use for copula *BE* (*is* + *are*), auxiliary *BE* (*is* + *are*), and auxiliary *HAVE* (*has* + *have*) for the Manchester corpus data, and the percentage correct use for Lara's data across Stage IV.

Contrary to the prediction, copula *BE* did not attract higher rates of error than auxiliary *BE* and auxiliary *HAVE* in either corpus. In fact, copula *BE* tended to attract lower rates of error across the board than the other auxiliary subtypes. Statistical tests on the Stage III and IV data combined revealed significantly higher

Table 5. Percentage correct use with each *wh*-word and copula (Cop.)/auxiliary (Aux.) form: Months 1 and 2 of Stage IV—Lara.

Wh-word	Aux.	% correct (*N*)	*p* value (Fisher's exact) and odds ratio (OR)
What	Cop. *is*	95.76 (283)	Cop. *is* × *are*
	Cop. *are*	0 (5)	$p < .0001$
	Aux. *is*	90.74 (54)	Aux. *is* × *are*
	Aux. *are*	2.51 (199)	$p < .0001$
			(OR = 380.24)
	Aux. *has*	60.00 (15)	Aux. *has* × *have*
	Aux. *have*	0 (79)	$p < .0001$
Where	Cop. *is*	82.01 (239)	Cop. *is* × *are*
	Cop. *are*	0 (3)	$p = .0065$
	Aux. *is*	91.67 (24)	Aux. *is* × *are*
	Aux. *are*	0 (17)	$p < .0001$
	Aux. *has*	89.12 (331)	Aux. *has* × *have*
	Aux. *have*	0 (8)	$p < .0001$

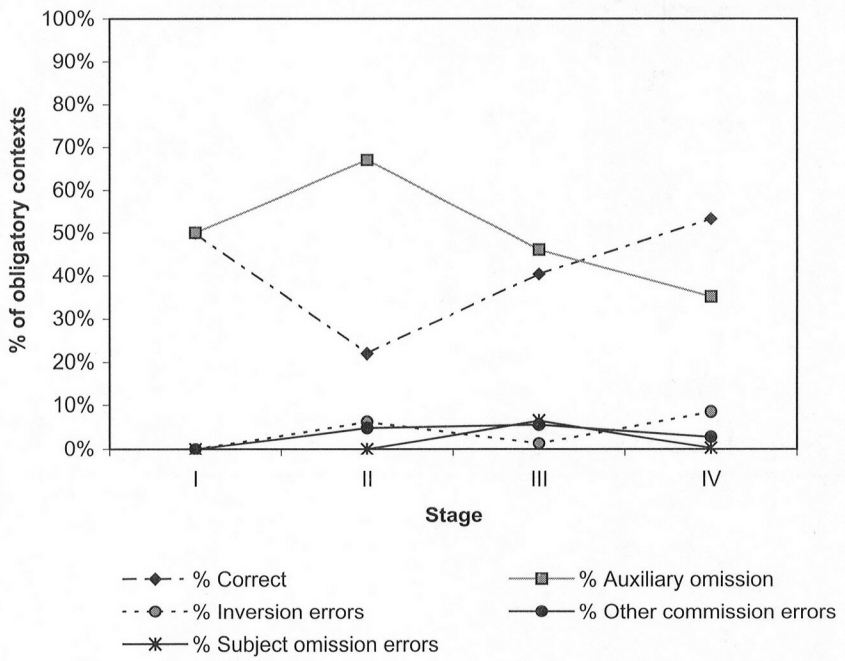
Note. OR was not calculated when one cell = 0. *N* = the number of obligatory contexts.

Table 6. Percentage correct use of copula (Cop.) *BE*, auxiliary (Aux.) *BE*, Aux. *HAVE*, and Aux. *DO*/modals.

	Mean % correct use (mean *N*)			
Manchester corpus	**Cop. *BE***	**Aux. *BE***	**Aux. *HAVE***	***DO*/modal auxiliaries**
Stage I	61.97	41.67	33.85	50.00
	(29)	(2)	(3)	(2)
Stage II	54.10	8.32	26.15	21.87
	(79)	(9)	(13)	(4)
Stage III	63.17	58.67	51.49	40.36
	(105)	(12)	(19)	(16)
Stage IV	74.33	46.51	64.54	53.32
	(115)	(22)	(21)	(36)
Stage III & IV	72.10	53.76	64.87	49.87
	(201)	(34)	(36)	(45)
	% correct use (*N*)			
Lara (Stage IV)	**Copula *BE***	**Auxiliary *BE***	**Auxiliary *HAVE***	***DO*/modal auxiliaries**
Month 1	78.85	43.59	71.20	40.91
	(279)	(78)	(125)	(44)
Month 2	85.76	19.51	65.77	48.15
	(309)	(246)	(336)	(81)
Month 3	89.04	25.00	66.67	77.14
	(356)	(232)	(222)	(70)
Month 4	93.87	80.52	87.12	93.48
	(163)	(154)	(132)	(92)

Note. *N* = the number of *wh*-question obligatory contexts used as the denominator of the percentage calculation.

Figure 5. Manchester corpus: mean percentage of *wh*-questions requiring auxiliary *DO* and modal auxiliaries that were correct and errors.



rates of correct use for copula *BE* than auxiliary *BE*, $F(1, 10) = 5.94$, $p = .04$, $\eta_p^2 = .37$, and auxiliary *HAVE*, $F(1, 9) = 5.40$, $p = .05$, $\eta_p^2 = .38$. In Lara's data, copula *BE* was produced correctly in obligatory contexts significantly more often than auxiliary *BE* during the whole of Stage IV (Month 1: Fisher's exact $p < .0001$, OR = 4.83; Month 2: $p < .0001$, OR = 24.84; Month 3: $p < .0001$, OR = 24.19; Month 4: $p = .0003$, OR = 3.70) and copula *BE* was correct significantly more often than auxiliary *HAVE* in Months 2 and 3 (Month 2: $p < .0001$, OR = 3.14; Month 3: $p < .0001$, OR = 4.06).

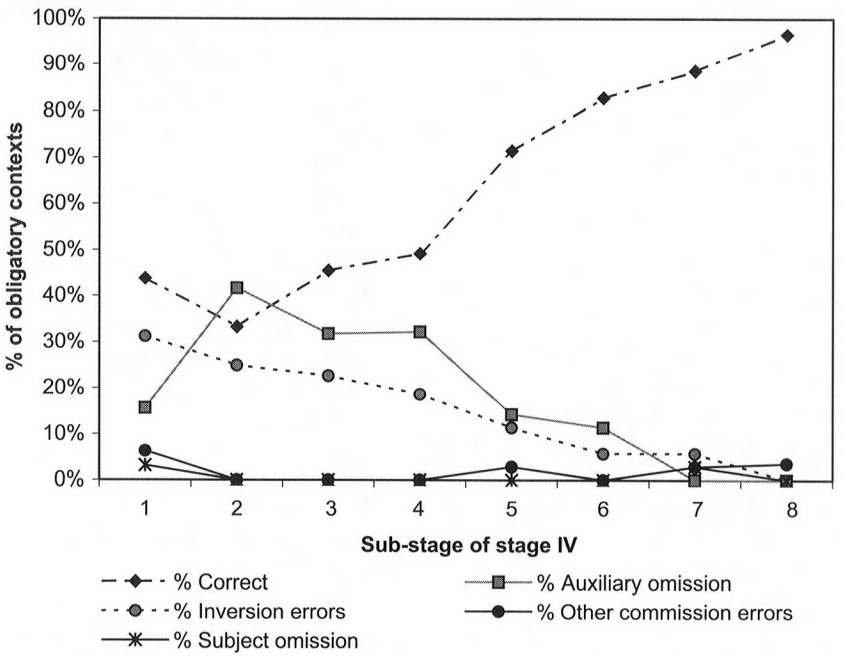
The fourth set of analyses tested the prediction that auxiliary *DO* will attract higher rates of error than auxiliary *BE* and *HAVE*. For questions with omitted auxiliaries, it was difficult to determine the identity of the intended auxiliary because the target could be a form of *DO* or one of the modals (e.g., *can, would, will*, etc.). As a result, we investigated the rate of correct use in questions that require either a form of *DO* (e.g., *do, does, did, didn't, doesn't*) or a modal auxiliary (e.g., *can, could, should, will*). Questions with omitted auxiliaries and subjects (e.g., *where go?*) were excluded to ensure consistency with the analyses performed on copula *BE*, auxiliary *BE*, and auxiliary *HAVE*. Figures 5 and 6 illustrate the percentage of correct use and errors for questions that required *DO* or modal auxiliaries in the Manchester and Lara data. Substantial numbers of errors were produced during all stages of the Manchester corpus data and during the first 4 substages in Lara's data.

Statistical comparison of rates of error use in Stage III and IV Manchester data combined (see Table 6 for means) demonstrated that *DO*/modal auxiliaries occurred correctly in obligatory contexts less often than auxiliary *BE* and *HAVE*. However, these differences failed to reach significance: auxiliary *BE* × *DO*/modals, $F(1, 10) = 0.45$, *ns*; auxiliary *HAVE* × *DO*/modals, $F(1, 9) = 3.38$, *ns*.

In Lara's data, there were differences between *DO*/modals and auxiliary *HAVE*, with *DO*/modals significantly less likely to be used correctly in obligatory contexts than *HAVE* in Months 1 and 2 (see Table 6, Month 1: Fisher's exact $p = .0005$, OR = 3.57; Month 2, $p < .0001$, OR = 2.07). However, questions with *DO*/modals were not less likely to be used correctly than questions with auxiliary *BE*. In fact, for Months 2, 3, and 4, questions requiring *DO*/modals were more likely to be correct than questions requiring auxiliary *BE* (Months 2 and 3, Fisher's exact $p < .0001$, ORs = 3.83 and 10.13, respectively; Month 4, $p = .005$, OR = 3.47). In summary, there is some support in Lara's data for the prediction that auxiliary *DO* may attract higher rates of error than auxiliary *HAVE*, but, contrary to the prediction, it seemed to attract significantly fewer errors than auxiliary *BE*.

Unexpectedly, Lara produced an extremely high number of errors of inversion in the early substages of Stage IV. These errors accounted for 31% of Lara's questions that required *DO*/modals at Substage 1,

Figure 6. Lara's data: percentage of *wh*-questions requiring auxiliary *DO* and modal auxiliaries that were correct and errors.



25% at Substage 2, 23% at Substage 3, and 19% at Substage 4. These error rates are much higher than those reported for the Manchester corpus. If reliable, the data would indicate that although children may not produce more errors overall with *DO* or the modals than with other auxiliaries, they may produce substantially more errors that derive specifically from problems applying the rules governing inversion in English.

As Lara's corpus is more comprehensive than that of the Manchester corpus children, it could be that it is simply more reliable at capturing a period of high error use. However, it is also possible that Lara is simply an outlier—a child who produces an atypically large number of errors compared to the majority of English-learning children. To investigate these two possibilities, we compared Lara's Stage IV data with the mean Stage IV data from the Manchester corpus. Table 7 demonstrates the percentage of correct questions and errors produced overall in Lara's data and the mean percentages, standard deviations, and 95% confidence intervals for the Manchester corpus. From the table it is clear that Lara is extremely similar to the Manchester corpus children in terms of the proportions of correct questions and errors produced overall.

We also considered the possibility that Lara was an outlier only in her *DO*/modal production, so we calculated the mean percentages, standard deviations, and 95% confidence intervals for the *DO*/modal auxiliaries for Stage IV for the Manchester children to compare to

Table 7. Correct use and errors at Stage IV.

Question type	Mean % of total questions—Manchester corpus Stage IV	*SD*—Manchester corpus Stage IV	Confidence interval (95%) Manchester corpus Stage IV	% of total questions Lara's data Stage IV
All *wh*-questions				
Correct	68.43	25.73	49–88	67.02
Omission errors	24.13	28.45	4–48	23.84
Inversion errors	1.39	1.84	0–3	2.35
Other commission errors	4.40	3.45	2–7	5.29
DO/modal questions				
Correct	53.32	34.39	27–80	68.64
Omission errors	35.30	38.22	6–65	16.38
Inversion errors	8.51	9.42	1–16	12.89
Other commission errors	2.87	4.44	0–6	2.09

the overall Stage IV percentages in Lara's data (see Table 7). Lara produced more correct questions and fewer questions with omitted auxiliaries than the Manchester corpus children on average, but her scores were well within 1 *SD* of the Manchester corpus mean and the 95% confidence interval. More importantly, her overall rate of inversion errors across the whole of Stage IV was very similar to that of the Manchester corpus children, only 4% higher than the Manchester mean.

Thus, the high levels of *DO*/modal auxiliary errors reported for Lara during the early substages of Stage IV cannot be attributed to her falling outside the range of normal production as characterized by the Manchester corpus. Instead, the results suggest that commission errors may be high for all children for a short period of time, but that the Manchester corpus is not dense enough to capture the period in sufficient depth.

Errors of Inversion in Questions With Present Auxiliaries

The final set of analyses investigated whether some auxiliaries attracted more errors of inversion than others. Because it is very difficult to distinguish between questions that require forms of auxiliary *DO* and those that require a modal when the auxiliary is omitted, the analysis included only those questions that had an auxiliary or copula form present. The analysis was, thus, restricted to correct questions and inversion errors with a present auxiliary.

Removing other types of errors, particularly auxiliary omission errors, substantially reduces the numbers of utterances available for analysis. The Manchester corpus was not dense enough to perform reliable analyses on these errors—on average the Manchester corpus children produced only 6.36 inversion errors in Stages 3 and 4—so only Lara's data were used. In addition, as we are only interested in the patterning of errors during the period when errors occur, the analysis was conducted only on the first 2 months of Lara's Stage IV data (Substages 1–4) when she produced 25 of these errors. Table 8 demonstrates the results.

There were too few data to compare error rates across different forms of the same auxiliary or copula subtype. However, enough data were available to conduct significant tests across different auxiliary subtypes. The results suggested that questions requiring auxiliary *DO* occurred with errors significantly more frequently than questions requiring auxiliary *BE* and auxiliary *HAVE* (auxiliary *DO* × auxiliary *BE*, Fisher's exact $p = .0007$, OR = 7.45; *DO* × auxiliary *HAVE*, $p < .0001$, OR = 11.27). Questions requiring modal auxiliaries also occurred with significantly higher rates of error than questions requiring auxiliary *BE* and *HAVE* (modals × auxiliary *BE*, Fisher's exact $p = .007$, OR = 12.62; modals × auxiliary *HAVE*, $p = .0005$, OR = 19.08).

Table 8. Correct questions and errors of inversion for Substages 1–4 of Stage IV—Lara's data.

Auxiliary	No. of correct questions	No. of inversion errors	% inversion errors
Copula *is*	484	5	1.02
Copula *are*	1	1	50.00
Copula *BE*	485	6	1.22
Auxiliary *is*	75	2	2.60
Auxiliary *are*	7	0	0.00
Auxiliary *BE*	82	2	2.38
Auxiliary *Has*	309	3	0.96
Auxiliary *Have*	1	2	66.67
Auxiliary *HAVE*	310	5	1.59
Auxiliary *DO*	44	8	15.39
Modal auxiliaries	13	4	23.53

Note. The table includes only questions with an auxiliary present.

Questions requiring copula *BE* did not occur with inversion errors significantly more often than questions requiring auxiliary *BE* or *HAVE* and, in fact, occurred with inversion errors significantly less often than questions requiring modal auxiliaries, Fisher's exact $p = .0002$, OR = 24.87. Overall, the results suggest that questions with auxiliary *DO* and modal auxiliaries, but not those with copula *BE*, attract higher rates of inversion errors than questions with auxiliary *BE* and *HAVE*.

Discussion

The aim of this study was to investigate the pattern of correct use and errors in young children's *wh*-questions in order to test some of the predictions of current theories of acquisition. The results must be interpreted with caution because there were only small numbers of utterances available for many of the analyses. Bearing this in mind, it is, nevertheless, clear that analyses on overall error rates may hide significant differences in the amount of correct use across different auxiliaries. Our results indicate that *wh*-questions with copula and auxiliary *is* and auxiliary *has* may attract higher rates of correct use than questions with copula and auxiliary *are* and auxiliary *have*. For the more substantial Lara data, this was the case during the early substages of Stage IV even when we restricted the analysis to questions beginning with *what* and *where*, although it must be noted that there were only small numbers of utterances included. If replicated, these

findings seem inconsistent with the predictions of the four generativist theories we have considered (DeVilliers, 1991; Santelmann et al., 2002; Stromswold, 1990; Valian et al., 1992).

The differences cannot be attributed to the child having failed to learn the lexical forms. All the Manchester corpus children except Ruth produced a number of examples of copula *are,* auxiliary *are,* and auxiliary *have* in their speech by Stage III and were capable of producing correct utterances with these forms in other structures at the same time as they produced substantial numbers of *wh*-question errors (Ruth did not produce any correct utterances with auxiliary *are* but produced copula *are* forms by Stage II and auxiliary *have* forms by Stage III). Lara produced examples of all three auxiliary forms during the sample transcript recorded at the beginning of the study (age 2;7.25). The differences cannot be reliably attributed to an absence of the relevant lexical forms in the children's vocabulary.

The analyses by auxiliary subtype demonstrated that copula *BE* occurred with significantly lower rates of error than auxiliary *BE* and auxiliary *HAVE*. There was some evidence that questions requiring *DO*/modals might attract higher rates of error than questions requiring auxiliary *HAVE* in Lara's data but very little evidence that they attracted higher rates of error than auxiliary *BE* in either corpus.

Nevertheless, Lara's data showed that, for a short period of time, errors of inversion accounted for a significant proportion of the questions produced with *DO*/modals. It was demonstrated that the failure to find this pattern in the Manchester corpus could be due to a lack of data as a result of the sampling technique. If these errors could be attributable to problems with *DO*-support, the finding would support the prediction of Stromswold (1990) and Santelmann et al. (2002), who suggested that children have problems learning the language-specific rules governing the presence of auxiliary *DO* in *wh*-questions.

The analyses conducted only on questions with a present auxiliary and which focused on errors of inversion demonstrated that, although questions with *DO*-support attracted high levels of error, the patterning of errors on other auxiliaries did not support the predictions. Stromswold (1990) and Santelmann et al. (2002) predicted a distinction between auxiliary *DO* and copula *BE* (which should attract high rates of error) and the modals, auxiliary *BE*, and auxiliary *HAVE* (which should attract lower rates of error). Our results suggest, instead, a distinction between auxiliary *DO* and the modals, which attract high rates of inversion error, and copula *BE*, auxiliary *BE*, and auxiliary *HAVE*, which attract lower rates of inversion error. Thus, it would seem that the auxiliaries that cause problems in the sense that children fail to produce correctly inverted *wh*-questions with them are *DO* and the modals, but not forms of copula *BE*. Generativist theories do not currently predict this pattern of error use.

In summary, although some of the results fit some of the predictions, it is clear that overall the data do not provide unqualified support for the generativist theories under discussion. Nevertheless, there may be potential generativist explanations. For example, there is no reason why theorists could not posit a greater role for the learning of individual auxiliaries. One potential solution is to posit a large role for the rote learning of contracted auxiliaries. The idea that utterances with contracted auxiliaries could be rote learned forms is well established in the literature (see, e.g., Pinker, 1984) and there is some evidence from our data that levels of correct use may differ substantially for contracted and uncontracted forms. For example, removing contracted forms of copula *is* from the data leads to a substantial drop in the mean percentage of correct use over Stages III and IV in the Manchester corpus (from 75% to 34%), suggesting that most of the correct questions produced with copula *is* contained the contracted form.

We suggest that is it probable that contracted and noncontracted forms may show different patterns of error use and that a generativist theory that clearly distinguishes between the acquisition of contracted and uncontracted forms may well be able to explain our data. However, for such a theory to be empirically testable the concept of rote learning would have to be clearly operationally defined and productivity criteria would need to be carefully applied. Taking a broad definition of productivity, we could argue that there is little evidence for rote learning in the data. For example, although all nine *wh*-questions produced with copula *is* in the first of Anne's transcripts are of the form *what's this,* Anne also produced the utterance *oh it's there* in the same transcript, which could provide evidence for the productive use of contracted copula *is*. However, if we applied a much more conservative productivity criterion such as a measure of contrastive use within *wh*-questions, we would conclude that these nine questions were rote learned (see Aguado-Orea & Pine, 2002, for a further discussion of productivity criteria). These issues will need to be considered carefully if a substantial role for rote learning is to be proposed.

Additionally, the present study has demonstrated very different levels of correct use across questions even quite late in the acquisition process (Stage IV). An explanation of these differences in terms of the rote learning of contracted forms implies a much greater role for rote learning for a longer period of time than has

previously been claimed. Even more importantly, theories incorporating rote learning as an explanation of our results would have to include a discussion of why, even in Stage IV, children are still failing to generalize their knowledge about the correct positioning of contracted forms not only to other forms of the same auxiliary subtype but to noncontracted forms of the same auxiliary.

It seems to us that in order to explain the pattern of results we present here, generativists will need to specify a much greater role for learning than is currently suggested. In such formulations, the learning mechanism would be responsible for identifying both the form and function of question words, for distinguishing questions from other utterances with different functions, and for learning language-specific rules about the application of inversion on an auxiliary-by-auxiliary basis, with little scope for generalization across different auxiliary forms or even contracted and uncontracted forms of the same auxiliary. With such an emphasis on the learning mechanism, it will be important that generativist theorists ensure that these theories are distinguishable empirically from those that posit little or no innate knowledge.

The alternative is to consider these results in the light of a constructivist usage-based theory, which contends that the basis of children's early utterances is a learned knowledge of utterance–meaning pairings based on lexically specific constructions (see Pine, Lieven, & Rowland, 1998; Tomasello, 2000, 2003). On this account, the child's early knowledge of grammar is tied to individual lexical items or lexical frames (e.g., *it's a* or *where's the*). From this, more abstract linguistic schema develop as the child starts to generalize across items and frames that share common distributional properties. Children's ability to acquire new constructions and generalize across them is seen as influenced by a combination of cognitive and linguistics factors such as semantic complexity, input frequency, phonological salience, and the child's own social and cognitive interests.

In *wh*-question acquisition, this theory makes the prediction that for an extended period of time, children will produce the majority of their correct questions with rote learned semiformulaic frames such as *what's X doing?, where's X gone?,* where *X* stands for a variety of noun phrases (see Dąbrowska, 2000; Rowland & Pine, 2000). Questions that can be constructed using these frames will tend to be used correctly, which would explain the high level of correct use with copula *is,* auxiliary *is,* and auxiliary *has* in our data. If, as seems likely, copula *are,* auxiliary *are,* and auxiliary *have* are of low phonological salience in the speech that children hear, the theory would predict that these forms will be omitted until later on, perhaps until the child starts to coordinate his or her disparate knowledge of the behavior of auxiliaries in different constructions. Other errors will occur when children have no relevant *wh*-question frame with which to construct their intended question but have started to develop the ability to create questions by generalizing across different frames. For example, the predominance of errors with *DO* and modal auxiliaries is explicable in terms of the child combining *wh*-word or *wh*-word + auxiliary frames with novel but inappropriate declarative fragments (e.g., *what did + he can do (it) = what did he can do, what + he did (it) = what he did?*).

The idea that some of children's mistakes may be examples of conjoining errors in which the child joins a *wh*-word + auxiliary pivot with an inappropriate variable is not new. Kuczaj (1976) suggested a similar explanation for double tensing errors such as *what's that is?* (Kuczaj, 1976, p. 424). He argued that children fail to segment out the *wh*-word and contracted auxiliary as separate constituents, treating them as unanalyzed entities and combining them inappropriately with a further auxiliary lexeme. The explanation presented by the constructivist theory would take this idea one step further by suggesting that many of children's *wh*-word + auxiliary frames, perhaps even those with noncontracted auxiliaries, may be unanalyzed or at most partially analyzed, with the child failing to comprehend that the preposed auxiliary element should be considered as a replacement for the base-generated auxiliary or the tensed main verb.

However, although the constructivist usage-based theory is compatible with the pattern of error presented here, it needs to make more detailed predictions about errors. In particular, it must be able to explain why some errors are relatively frequent in children's questions (e.g., subject auxiliary inversion errors) whereas others (e.g., subject auxiliary agreement errors) seem extremely rare (see Stromswold, 1990). At present, it is often difficult to ascertain how the influences on acquisition suggested by the theory will interact, and what pattern of correct use and errors we would expect to see emerge from this interaction (though see Lieven, Behrens, Speares, & Tomasello, 2003; Theakston et al., 2005; Tomasello, 2000, 2003, for clear ideas about how children generalize across lexical formulae, ideas which are relevant to predictions about errors). If constructivism is to present a strong challenge to nativism, it must begin to incorporate more detailed proposals about the possible mechanism behind language acquisition. These proposals will allow us to make predictions about when errors should occur, what parts of the system should be affected, and when we would expect errors to disappear.

To conclude, the present study has demonstrated that *wh*-questions requiring different auxiliaries can show very different rates and patterns of error use. More importantly, it has demonstrated that we need to be cautious about arguments based on low error rates in sampled data. Errors can occur with high frequency in some, usually rarely produced, parts of the grammatical system and can occur for only short periods of time in high numbers. This finding is perhaps not surprising, as we would expect children to make errors on the parts of the system on which they have had little chance to practice and improve. However, it does emphasize that error rates that are calculated on the basis of sampled data or that do not take into account potential differences across different lexical items cannot be used as the basis of an effective argument that children have early adult-like competence in language production.

Acknowledgments

This study was supported by the Economic and Social Research Council, R000236393 and RES000220241. We would like to thank all the families who took part in the research reported here, Ewa Dąbrowska, Jeremy Miles, John and Mildred Hadwin for their comments on a draft, and Mildred and John Hadwin, George and Elsie Rowland, and Rachel Edden for their role in data collection. Part of this article was presented at the Child Language Seminar, Newcastle-Upon-Tyne, United Kingdom, 2003, and the 6th Annual Gregynog Conference on Child Language, Nant Gwrtheyrn, United Kingdom, 2004.

References

Aguado-Orea, J., & Pine, J. M. (2002, July). *Assessing the productivity of verb morphology in early child Spanish.* Paper presented at the IX International Congress for the Study of Child Language, Madison, WI.

Bellugi, U. (1965). The development of interrogative structures in children's speech. In K. Riegel (Ed.), *The development of language functions* (pp. 103–138) [University of Michigan Language Development Program, Report No. 8]. Ann Arbor: University of Michigan.

Bellugi, U. (1971). Simplification in children's language. In R. Huxley, & E. Ingram (Eds.), *Language acquisition: Models and methods* (pp. 95–119). London: Academic Press.

Brown, R. (1968). The development of Wh questions in child speech. *Journal of Verbal Learning and Verbal Behavior, 7,* 279–290.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Dąbrowska, E. (2000). From formula to schema: The acquisition of English questions. *Cognitive Linguistics, 11,* 83–102.

DeVilliers, J. (1991). Why questions? In T. Maxfield, & B. Plunkett (Eds.), *Papers in the acquisition of WH* (pp. 155–173). Amherst: University of Massachusetts Occasional Papers.

Erreich, A. (1984). Learning how to ask: Patterns of inversion in yes-no and wh-questions. *Journal of Child Language, 11,* 579–602.

Fenson, L., Dale, P. S., Bates, E., Reznick, J. S., Thal, D., & Pethick, S. J. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual.* San Diego, CA: Singular.

Hurford, J. (1975). A child and the English question formation rule. *Journal of Child Language, 2,* 299–301.

Kuczaj, S. A. (1976). Arguments against Hurford's 'Aux Copying Rule.' *Journal of Child Language, 3,* 423–427.

Kuczaj, S. A. (1986). General developmental patterns and individual differences in the acquisition of copula and auxiliary be forms. *First Language, 6,* 111–117.

Labov, W., & Labov, T. (1978). Learning the syntax of questions. In R. Campbell & P. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches* (pp. 1–44). New York: Plenum Press.

Lieven, E. V. M., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30,* 333–370.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* Mahwah, NJ: Erlbaum.

Maratsos, M., & Kuczaj, S. A. (1978). Against the transformationalist account: A simpler analysis of auxiliary overmarkings. *Journal of Child Language, 5,* 337–345.

Pine, J. M., Lieven, E. V. M., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics, 36,* 807–830.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: MIT Press.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax: The nature of early child grammars in English.* Oxford, England: Basil Blackwell.

Rowland, C. F. (2000). *The acquisition of wh-questions in early English multi-word speech.* Unpublished doctoral dissertation, University of Nottingham, England.

Rowland, C. F., & Fletcher, S. (2003, July). *Why modeling does matter: The use of elicitation to study wh-question production in children.* Poster session presented at the Child Language Seminar, Newcastle-upon-Tyne, England.

Rowland, C. F., & Pine, J. M. (2000). Subject-auxiliary inversion errors and wh-question acquisition: What children do know? *Journal of Child Language, 27,* 157–181.

Santelmann, L., Berk, S., Austin, J., Somashekar, S., & Lust, B. (2002). Continuity and development in the acquisition of inversion in yes/no questions: Dissociating movement and inflection. *Journal of Child Language, 29,* 813–842.

Stromswold, K. (1990). *Learnability and the acquisition of auxiliaries.* Cambridge, MA: MIT Working Papers in Linguistics.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of argument structure: An alternative account. *Journal of Child Language, 28,* 127–152.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary syntax: *BE* and *HAVE*. *Cognitive Linguistics, 16,* 247–277.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74,* 209–253.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31,* 101–121.

Vainikka, A. (1994). Case in the development of English syntax. *Language Acquisition, 3,* 257–325.

Valian, V., Lasser, I., & Mandelbaum, D. (1992). *Children's early questions*. Unpublished manuscript, Hunter College, New York.

Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language, 30,* 75–115.

Received October 2, 2003

Revision received April 1, 2004

Accepted July 26, 2004

DOI: 10.1044/1092-4388(2005/027)

Contact author: Caroline Rowland, School of Psychology, University of Liverpool, Bedford Street South, Liverpool, L69 7ZA, United Kingdom. E-mail: crowland@liverpool.ac.uk

Julian M. Pine is now at the University of Liverpool, Liverpool, United Kingdom.